Doctoral Dissertation
Doctoral Program in Civil and Environmental Engineering (36^{th}cycle)

Fish collective behaviour in flowing waters

By
Gloria Mozzi

******

Supervisors:
Prof. C. Comoglio, Supervisor
Prof. C. Manes, Co-Supervisor

Doctoral Examination Committee:
Dr. Luiz G. M. Silva, Referee, ETH Zurich
Dr. Ana T. Silva, Referee, Norwegian Institute for Nature Research

Politecnico di Torino
2024

Declaration

This thesis is licensed under a Creative Commons Licence, Attribution - Noncommercial - NoDerivative Works 4.0 International: see ww.creativecommons.org. The text may be reproduced for non-commercial purposes, provided credit is given to the original author.

I hereby declare that, the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Gloria Mozzi
2024

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

A mia mamma,
mio babbo,
e mia nonna.

"Non è l'acqua di fonte che battezza i figli dei marinai.
Nascere sul mare, e non importa se poi il carro ti conduce altrove,
ovunque te ne andrai saprai sempre di sale."

Egisto Malfatti, 1995

Acknowledgements

I miei primi ringraziamenti vanno ai miei supervisori, che mi hanno costantemente supportato e guidato in questo lungo percorso offrendomi uno spazio di confronto sano, con una costante attenzione alle mie necessità - scientifiche e non. Sono ben consapevole che molti ambienti lavorativi non hanno queste qualità; non posso che sentirmi fortunata nell'essere capitata con voi ed estremamente grata per avermi scelto per questa avventura. A Costantino, per essere stato sempre presente, per il suo indomito ottimismo e per la sua continua fiducia nelle mie capacità. Fin dalla prima mail - ancor prima di applicare - mi hai ispirato grande fiducia e sintonia grazie all'entusiasmo e alla curiosità che più ti caratterizzano. A Claudio, che mi ha mostrato come una propensione ad un animo buono e gentile possa essere trasformata in una vocazione di lavoro. Sapere che ci sono persone come te in questo settore mi rende più ottimista verso un mondo più sostenibile. A Daniel, che è sempre stato il primo a rispondere nei momenti di necessità, offrendo costantemente consigli e supporto.

Thanks to my RIBES ESR colleagues. It has been an honour getting into this project with such a talented cohort. Above all, I want to thank Usama for his endless patience and company during those long hours of fieldwork. You showed me how patience and serenity are such an important part of research – and life - and I am extremely grateful to have had you by my side. Tou are possibly the only person in the world who could have stood me for so much time! I want to thank Sophia for her extreme dedication into our work and for the constant cheerfulness during those long-lasting hours. With your enthusiasm and laughter, you managed to turn our weeks of work in blissful moments of friendship.

Ai miei colleghi di laboratorio, Andrea, Roberto(2), Elia, Elisabetta e Davide. Non smetterò mai di stupirmi dell'atmosfera di altruismo, supporto ed amicizia che il buon Bidone ha creato. Siete delle persone meravigliose, oltre che estremamente capaci. Ogni qual volta avevo bisogno di un supporto - tecnico e non - ci eravate. Siete stati la luce in tanti momenti di buio, e senza di voi questo giorno non sarebbe semplicamente arrivato.

A tutte e tutti i compagni di dottorato e alle colleghe del DIATI2. Una menzione speciale a Luca, compagno di grandi improperi, che appariva ogni volta che sentivo di non avere nessuno. Grazie a Marta e a Carla, per la sorellanza tra ingegnere e per essere la testimonianza del fatto che avere successo e sollevare gli altri non sono aspetti in contrasto, ma parti fondamentali di uno sviluppo professionale sano. Grazie a tutto il gruppo di Cene Impossibili, che pian piano mi ha accolto ed adottato facendomi sentire parte di un gruppo in questa curiosa città.

Alla mia famiglia di Torino, Giulia, Anna e Keb. Mi avete accolto nella vostra vita con semplicità e affetto, siete stati la compagnia di avventure più inclusiva possibile, ed anche se adesso le nostre vite ci hanno un po' separato, sarò sempre grata del tempo passato insieme e vi considerò sempre famiglia.

Alle mie coinquiline e coinquilino di Casa Cernaia, Sofia, Martina (e Fra!), Paula, Giulio e Carlotta. Trovare voi di ritorno a casa è stato il più dolce aiuto dopo le giornate più brutte. Grazie per avermi ascoltato, accolto, e distratto nei momenti più duri.

Alle mie amiche di Viareggio, Sara, Francesca, Elena, Ilaria, Giulia, Giada, Anne, e a Luca. La nostra amicizia storica mi ha reso quella che sono, e per questo, ve ne sarò eternamente grata. Anche se il carro mi ha portato altrove, voi mi ricordate che so sempre di sale.

A Roberto, che nonostante (o grazie?) la sua testaccia dura è riuscito a diventare la persona per me più importante al mondo. Grazie per accogliermi ed accettarmi nelle mie perfezioni ed insicurezze, per starmi accanto nei momenti più bui, con la pazienza e l'amore che mi stupiscono ogni volta. La luce che vedo nei tuoi occhi è una forza costante che con il tempo mi è diventata necessaria. Non soltanto per l'amore profondo che trasmette, ma perché ci vedo riflessa la stessa stima che provo per te. Sei la persona più brillante, tenace, e capace che conosca, e mi sento incredibilmente fortunata nell'essere la tua compagna.

Alla mia famiglia, che dagli inizi del mio percorso ha sempre creduto in me. A mia nonna, per gli infiniti pranzi, per il sorriso, e per quella saggezza spensierata che sempre mi porterò sempre dietro. A mia mamma, la persona più buona che conosca, e a mio babbo, che è riuscito a trasmettermi la consapevolezza della gratitudine. Nel crescere e vedervi come esseri umani imperfetti, non posso che meravigliarmi della purezza del vostro affetto. Mi rendo conto del privilegio che ho avuto nel nascere e crescere in Piazza Shelley, e di quanto vi siate impegnati per darmi tutto quello che pensevate avessi bisogno, con un'attenzione particolare a quello che vi è mancato. Posso solo sperare, se mai diventerò madre, di essere in grado di fare lo stesso.

Infine, voglio dedicare un ultimo grazie a me stessa. Non per i successi o i grandi traguardi raggiunti, ma per essere riuscita ad uscire dal letto quando sembrava così difficile. Per aver continuato a respirare nonostante i momenti in cui mi sentivo affogare. Per aver avuto il coraggio di piangere. Per quei giorni in cui soltanto essere, andare avanti, sembrava un traguardo. Se ce l'ho fatta è solo grazie al mio nucleo intimo, alla mia Donna Selvaggia - *la que sabe* – che è riuscita a darmi quel filo di forza, nonostante gli anni del dottorato l'abbiano sopita. Spero solo di riuscire a diventare sempre più capace ad accudire, a nutrire, a coltivare quell'intuito che mi portato fin qui. Mi piace immaginarmi tra diversi anni, riaprire queste pagine, e finalmente capace di correre coi lupi. Libera.

Summary

Freshwater fish populations face a significant threat due to the fragmentation of river systems, leading to a decline in their numbers. Despite the construction of fish passages aimed at mitigating these impacts, their relatively low efficiency remains a challenge. Existing research primarily focuses on understanding fish behaviour individually, often overlooking collective behaviour dynamics, particularly in relation to flow dynamics. This thesis addresses this gap by analysing fish collective behaviour across varying flow velocities. An experimental campaign conducted from May to June 2021 with wild *Telestes muticellus* provided the groundwork for the study. An artificial intelligence-based approach is developed to extract trajectory data of individual fish from videos taken during the experimental campaign. The study investigates the social facilitation effect across different group sizes and flow velocities while also examining pairwise interactions using linear approaches. Furthermore, a novel transfer entropy-based method is introduced to quantify information transfer, considering non-linear dynamics. This research contributes to a deeper understanding of fish collective behaviour and its interaction with flow dynamics, offering insights for enhancing fish passage design and conservation efforts in freshwater ecosystems.

Table of Contents

List of Figures

Appendix Figures

List of Tables

List of Symbols

Symbols (in order of appearance)

TP	True Positive
FP	False Positive
FN	False Negative
mAP	mean Average Precision value
AP	Average Precision
IoU	Intersection over Union
F1	Standard measure of positive performance
MD	Missing Detections
$N_{CNN\ det}$	Number of detections from the convolutional neural network
N_{fish}	Group size
N_{frames}	Total number of frames in the video trial
x	Position coordinate on the horizontal axis longitudinal to the flow
y	Position coordinate on the horizontal axis transversal to the flow
z	Position coordinate on the vertical axis
$\dot{x}, \dot{y}$	Velocity on the horizontal coordinates
t	Current time

Δt	Time interval between consecutive states
$\vec{x}_t$	State vector
x^p	Predicted state
x^e	Estimated state
$\vec{m}_t$	Actual measurement at the time step
$x_{CNN,t}$, $y_{CNN,t}$	Position detected by the convolutional neural network
x_0	Initial state
m_0	Initial measurement
V	Covariance matrix
V_0	Initial covariance matrix
$var(\mathrm{x})$	Variance on the position x
$covar(x, y)$	Covariance on the positions x and y
A	Dynamic matrix
Q	Noise covariance matrix
σ_a	Standard deviation on acceleration
G	Conversion noise covariance matrix
K	Kalman gain
H	Measuring matrix
R	Noise covariance matrix
I	Identity matrix
D	Dumper matrix
d_v	Velocity dumper coefficient
NaN	Non-available number
$b_{i,t}$	Sum of the norms of the acceleration vectors for the configuration i at the time-step t
$q_{i,t}$	Probability associated to the configuration i at the time-step t
M_1	Matrix to extract position vector

M_2	Matrix to extract velocity vector
$\vec{a}_{k,t}$	Acceleration vector of the fish k at the time-step t
$\vec{v}_{k,t}$	Acceleration vector of the fish k at the time-step t
EA	Explored area
TL	Trajectory length
$P_H(x,y)$	Horizontal probability density
$P_V(z)$	Vertical probability density
$\widetilde{P_H}(x)$	Medians of probability densitiy
TL*	Adjusted trajectory length for bottom-filtered analysis
N_{bottom}	Number of frames in which no fish is detected above 5 cm from the flume bottom
χ^2	Goodness of fit test
N_{bottom}	Number of frames in which no fish is detected above 5 cm from the flume bottom
N	Number of experiments (20 for single fish, 10 for two fish, and 10 for six fish)
U_{ms}	Maximum sustained velocity
U_B	Bulk flow velocity
U_{B10}	Flow velocity of 10 cm/s
U_{B20}	Flow velocity of 20 cm/s
U_{B35}	Flow velocity of 35 cm/s
SR	Swimming regions
t_{SR}	Time spent in SR
t_{SR}^*	Time fish spent in SR normalised for the area extension of SR
u	Fish velocity on the longitudinal (x) component
v	Fish velocity on the lateral (x) component
C_u, C_v	Mean correlation on longitudinal and lateral components
$\vec{r}$	Relative position vector
τ	Time lag

d	Relative distance between fish
θ	Relative orientation between fish (θ=0° corresponds to a side-by-side configuration, while θ=90° stands for an in-line arrangement)
$u_{L,n}$	Longitudinal velocity component of the Lead fish of the nth experiment
$u_{R,n}$	Longitudinal velocity component of the Rear fish of the nth experiment
$\overline{u_{L,n}}$	Mean longitudinal velocity of the Lead fish in the nth experiment
$c_{u,n}$	Correlation on the longitudinal velocoity components on the nth experiment
$t_{r,u}$	Average response time on the longitudinal component
$H(X)$	Entropy of a random variable X
$P(x)$	Probability distribution of the variable X
$H(X,Y)$	Joint entropy
$H(X\|Y)$	Conditional entropy
$I(X;Y)$	Mutual information
$TE_{Y\to X}$	Transfer entropy from the source variable Y to the destination variable X
$ETE_{Y\to X}$	Effective transfer entropy from the source variable Y to the destination variable X
$netTE_{Y\to X}$	Net transfer entropy from the source variable Y to the destination variable X
$totTE_{Y\leftrightarrow X}$	Total transfer entropy from the source variable Y to the destination variable X
$netETE_{Y\to X}$	Net effective transfer entropy from the source variable Y to the destination variable X
$totETE_{Y\leftrightarrow X}$	Total effective transfer entropy from the source variable Y to the destination variable X
u_{th}, v_{th}	Threshold velocities for state definition in the longitudinal and lateral components, respectively
$netETE_u$	Net effective transfer entropy on the longitudinal component
$ETE_{u_L\to u_R}$	Effective transfer entropy from the Lead fish to the Rear one on the longitudinal component
$TE_{synth\to uR,n}$	Synthetic transfer entropy with destination variable the longitudinal state of the Rear fish of the nth experiment
N_s	Number of synthehic simulations

Abbreviations

°C	Celsius degrees
2D	Two-dimensional
3D	Three-dimensional
ABM	Agent-based Models
AI	Artificial Intelligence
BL	Body Lengths
cm	centimeters
CNN	Convolutional Neural Network
GUI	Graphical User Interface
LED	Light Emitting Diode
m	meters
PIT	Passive Integrated Transponders
R-CNN	Region-based Convolutional Neural Network
s	Seconds
SD	Standard Deviation
SSD	Single Shot multibox Detectors
SVM	Support Vector Machines
YOLO	You Only Look Once

Introduction

Background

In recent decades, there has been a significant decline in migratory freshwater fish populations, largely attributed to the proliferation of engineering interventions such as dams, weirs, and culverts [1]. These structures pose substantial threats to these species, as they represent barriers to their migration routes, often resulting in blockage, injury or delay [2]. Proposed solutions to alleviate this pressure involve the construction of fish passages, which vary from more nature-like to hard-concrete structures [3,4]. Despite their widespread implementation, our understanding of their efficacy is limited, with some studies revealing the ineffectiveness of such passages, particularly for non-salmonid species [5,6]. Consequently, substantial research efforts have been directed towards improving fish passage efficiency [7].

Addressing this low passage efficiency requires a comprehensive understanding of fish behaviour and swimming performance. In this regards, many knowledge gaps persist, notably in the domain of collective behaviour among gregarious species in flowing waters. It is well-established that aggregation in fish can influence swimming performance, stress, and behaviour [8]. Nevertheless, collective behaviour is frequently overlooked when it comes to fish passage design and assessment. Specifically, there is a paucity of information concerning the impact of flow hydrodynamics on the collective behaviour of riverine fish. Given the diverse flow conditions characterising the natural habitats of these species, these knowledge gaps significantly impede our comprehensive understanding of collective behaviour of riverine species. Bridging these gaps is essential to developing more effective fish passage designs and, consequently, ensuring the conservation and restoration of migratory freshwater fish populations.

Objectives

The aim of this thesis is to explore the interplay between collective behaviour and bulk flow velocity in riverine fish.

To accomplish this goal, three specific objectives were set, namely:

i. investigate social facilitation in fish at different flow velocities, focusing on exploratory behaviour and swimming activity;
ii. quantify mutual fish interactions at different flow velocities;
iii. disentangle linear and non-linear mutual interactions and quantify information transfer among fish pairs.

In order to achieve these objectives, the first research action undertaken was to conduct and present a comprehensive literature review, delving into the current research landscape regarding collective behaviour in riverine fish, with a special emphasis on applications for fish passage. Building upon this review, which identified relevance and literature gaps, an experimental campaign was conducted in May-June 2021 to investigate the collective behaviour of Italian riffle dace (*Telestes muticellus*) under varying flow velocities. A specialised movable flume, conceived and constructed at Politecnico di Torino, was employed to conduct the experiments. The apparatus comprised an open channel flume with transparent walls, allowing control over flow rate and water depth. The pre-existing functionality of the flume at the commencement of this study expedited its immediate utilisation. The author was responsible for conceiving and implementing the sensory system while also contributing to the development of some components to optimise flow stability. After completing the experimental campaign, the author devised a methodology tailored to the equipment used in this study for extracting trajectory data from videos. This methodology employed a two-step approach, which included fish detection using artificial intelligence (deep learning) and a custom-made tracking algorithm.

Thesis structure

Chapter 1 establishes a comprehensive framework within the existing literature concerning the intricate relationship between hydrodynamics and collective behaviour in freshwater fish, specifically emphasising its applications in the context of fish passage. This Chapter delineates the significance of investigating the interplay between flow velocity and collective behaviour. These foundational insights serve as the underpinning for the subsequent Chapters, providing a solid groundwork for the entire study.

Chapter 2 delineates the developed methodology for extracting fish trajectories from videos captured during the experimental campaign. This methodology consists of two integral components: Detection and Tracking. In the Detection phase, objects are identified within the image frame. This phase involved the implementation of a customized Deep Learning approach, wherein two Convolutional Neural Networks were trained to detect fish heads in videos recorded both from the bottom and the side of the flume. The Tracking phase involved the reconstruction of fish trajectories over time using the detection data obtained. In order to accomplish this, an algorithm rooted in a Kalman filter with integrated identity recognition was devised and applied.

The subsequent three Chapters are dedicated to addressing the three stated objectives by utilising trajectory data acquired during the experimental campaign to investigate fish collective behaviour under distinct water velocities. Chapter 3 delves into a detailed examination of the interplay between flow velocity and social facilitation as per objective i. This analysis specifically scrutinised fish exploration and swimming activities across varying group sizes (i.e., single fish, two fish, six fish). This section characterizes the spatial utilisation by fish in both the horizontal and vertical planes, providing a comprehensive analysis and discussion of the observed behaviours.

Chapter 4 addresses objective ii by delving into the interaction dynamics among fish pairs. Here, linear cross-correlation analysis is employed to scrutinise these interactions, extracting key variables, including decorrelation distance, response time, and spatial arrangement patterns between fish. This section systematically introduces, presents, and discusses the influence of flow velocity and boundary effects on the observed interactions.

In the fifth and final Chapter, we introduce a pioneering methodology, named state-based transfer entropy, to explore fish non-linear information transfer, aligning with objective iii. This innovative approach, rooted in information theory, addresses certain limitations of transfer entropy, making it more accessible to audiences beyond the fields of physics and mathematics. In contrast with traditional linear cross-correlation analysis, this methodology enables the precise quantification of direct information transfer

between fish pairs. The resulting outcomes are compared with those presented in Chapter 4, providing a robust cross-validation for the identified key parameters.

Each Chapter in this thesis is crafted as an independent article, intended for submission to peer-reviewed journals. While adapted for the thesis structure, the content retains its scholarly integrity. Chapter 1 has been published in Advances in Hydraulic Research of the GeoPlanet: Earth and Planetary Sciences book series [9], while Chapter 3 has been published in Scientific Reports, with Chapter 2 included as Supplementary Material [10]. A manuscript for Chapter 4 has been finalised and is ready for submission to Animal Behaviour, and a manuscript for Chapter 5 is currently being prepared for submission.

The concluding section brings closure to this work by providing a concise overview of the accomplished objectives and summarizing the key findings. Additionally, it offers reflections on potential avenues for future research.

Chapter 1. Aggregation in riverine fish from a fish passage perspective

Abstract [1]

According to the most recent Living Planet Report, freshwater fish species are among the most threatened species on Earth, with many of them showing a decline in population due to altered river connectivity caused by barriers. Fish passages are conservation measures aimed at mitigating the impact of such obstructions for migratory freshwater fish, providing corridors that should not harm, kill, stress, or excessively delay fish movement. Despite many species being gregarious, research aiming at assessing and improving passage efficiency has focused almost solely on behaviour and swimming performance of individual fish. Collective behaviour can, in fact, affect the way fish approach, enter, and pass a fishway. The mechanisms for which group behaviour affects fish movement in hydrodynamically complex environments, such as those occurring within fish passages, are multiple and not limited to: reduced energy expenditure, better navigation, reduced stress levels, increased exploratory behaviour, and change of predation dynamics. In this work, we review current research to illustrate how collective behaviour can be relevant for fish passage research. Our aim is to provide an overview of how collective behaviour might affect fish passage efficiency and how future research could improve fish passage design.

[1] This chapter has been accepted for publication in in Advances in Hydraulic Research of the GeoPlanet: Earth and Planetary Sciences book series.

Citation: Mozzi, G., Manes, C., Nyqvist, D., Domenici, P., & Comoglio, C. (2023). Aggregation in Riverine Fish: A Review from a Fish Passage Perspective. *International School of Hydraulics*, 265-280.

1.1 Introduction

The loss of biodiversity caused by climate change and other anthropogenic activities is among the most concerning environmental issues of our time. The estimated rate at which species go extinct is so fast that the scientific community warns that a sixth mass extinction is ongoing [11]. Freshwater ecosystems are vital for global biodiversity, hosting approximately 10% of the total known species and one-third of all vertebrate species, despite occupying less than 1% of Earth's surface [12]. They also constitute vulnerable ecosystems, as freshwater species are those showing the most dramatic population decline. Migratory freshwater fishes represent an emblematic example of such loss, with a reported global decline of about 76% in abundance from 1970 to 2016 [1].

Migratory fish regularly move along rivers for their survival, requiring different environments and resources for reproduction, foraging, hiding, and resting [13]. As a result, river fragmentation caused by dams, weirs, and culverts represents a tremendous challenge for migratory fish, and the increasing number of obstructions has been recognised as a critical driver for population decline. A survey conducted for 290 species found that habitat loss and degradation, particularly barriers to migration routes, represented the major threat [14].

Dams are constructed for various purposes, such as flood control, water supply, recreation, navigation, and electricity generation [15], and over the last century, the number of barriers in rivers has increased. A study by Grill et al. (2019) shows that only 37% of the global rivers longer than 1 000 km remain free-flowing over their entire length [16]. Artificial impoundments larger than 100 m^2 worldwide are estimated to be around 16.7 million, and this number is expected to increase [17]. Although in some cases these structures are obsolete and removable [18–20], many will likely have to remain in place to play a role in both climate change mitigation and adaptation [21]. Dams can, in fact, store water during times of drought [22], regulate water flow to prevent flooding [23], and generate renewable energy through hydropower, which can help reduce reliance on fossil fuels and mitigate climate change impacts [24]. As a result, migratory fish need to be supplied with a route for passing these barriers.

Fish passes - also referred to as *fishways* or *fish passages*- are artificial structures that provide a corridor for fish to safely overcome an anthropogenic obstacle along the river. They are built by opening an alternative waterway in the river and can be either built as concrete structures or nature-like [3,4]. Overall, the goal of every fish pass is to promote *safe and timely* fish movement across a barrier. A fishway should not substantially alter a fish's typical migration and life cycle: it should not cause excessive delay, stress, injuries and - obviously – mortality [25].

Even though much work on fish passes has focused on design principles [3,4,7,26–31], information regarding their effectiveness, once they are in place, is not as exhaustive. In fact, in contrast with the increasing number of fish passages constructed worldwide, a relatively small proportion has an integrated monitoring system to assess passage efficiency [5]. At many sites where assessments have been conducted, efficiency is relatively poor, especially for non-salmonid species (~60% vs ~25%) [5,6].

The variable functionality of fishways highlights the need for new research-based solutions and approaches to mitigate the effects of barriers on migration. This is no trivial work: fishways must attract fish to their entrance, induce them to access and pass through and provide them with all conditions to do so without harm or delay [32]. This implies that fishway design should involve different aspects related to behaviour and locomotion. Among these, collective behaviour has been almost entirely neglected in investigations concerning fish passage, while it could represent a relevant aspect to consider when targeting gregarious species.

In this work, we illustrate why and how aggregation can be relevant in relation to fish passage. Additionally, we advance some reflections on how this aspect can be integrated into future research.

1.2 Collective behaviour in hydrodynamically complex environments

Living and moving in a group can bring multiple benefits, such as reduced energy cost of movement, reduced stress, better navigation, foraging benefits, information transfer, and reduced predation risk [8]. Over the years, many researchers have focused on understanding collective behaviour, its evolution, functional complexity and factors affecting it [33–36]. Collective behaviour in swimming fish can manifest itself through two main modes: schooling and shoaling [37]. Schooling involves a group of fish swimming together and interacting in an organized and coordinated way, while shoaling involves groups that stay close together without movement synchronisation or polarization between individuals.

In the aquatic environment, hydrodynamic forces play an important role in shaping collective behaviour because they influence both the speed at which information propagates through a group as well as the structure of the group itself [38–40]. For instance, high turbulence levels can disguise the presence of other fish that would usually be detected through their sensory line, thus reducing their ability to detect and transfer information to other group members, as well as stimuli response of the whole group [39]. Additionally, hydrodynamics can affect school structure by influencing where individuals position themselves within a school due to favourable, energy-saving flow conditions [40]. Here, we illustrate key aspects of collective behaviour that could be relevant from the perspective of fish passage research and design.

1.2.1 Energy expenditure

The reduction in locomotion costs associated with schooling is confirmed in the literature by multiple experimental studies. By analysing tail beat frequency and oxygen consumption, it was observed that energy savings associated with schooling might exceed 20% in golden shiners (*Notemigonus crysofeucas* [41]), sea bass (*Dicentrarchus labrax* [42]), and roach (*Rutilus rutilus* [43]). These reduced locomotion costs were found to be related to the position of fish in the school, with trailing individuals saving more energy compared to the fish in front of the school. Additionally, Marras et al. (2015) found that grey mullet (*Liza aurata*) swimming in a group would have reduced swimming costs regardless of their position in the school [44]. From tail beat frequency analysis, they observed that even individuals swimming in front of their neighbours would gain energetic benefits compared to fish swimming in isolation at the same speed, although less than companions swimming at the rear of the school.

Despite this experimental evidence, mechanisms for which fish save energy while schooling remain not fully understood, and the research community still does not

converge on one theoretical framework. The idea that synchronized movement could bring energetic benefits due to the exploitation of hydrodynamics dates back to the theory of Weihs (1973), who proposed a 2-dimensional model of schooling fish improving swimming performances by assuming a diamond-shaped configuration [45]. Even if a diamond pattern could justify the decreased energy costs of moving in water, no evidence of a consistent occurrence of this configuration has been observed in nature. The limitations of such theory lay in the strictly 2D approach, as well as the assumption of fish keeping a constant position relative to the group and not adapting to flow gradients.

Different studies have then proposed models coupling behavioural rules with hydrodynamics, trying to depict the complexity of interactions among individuals at different flow conditions. Gazzola et al. (2016) first proposed a model using vortex dipoles to represent swimmers, showing that – in contrast with classical agent-based model rules - a reinforcement learning algorithm could explain the adaptive response of individual swimmers and, therefore, schooling in complex hydrodynamic environments [46]. Verma et al. (2018) also used reinforcement learning to find positions where numerically simulated swimmers could best exploit the shed vortices created by neighbouring fish [47]. They found that when fish place themselves following off-centre a steady leader, they could gain energy benefits by synchronizing the head movement with the wakes even in unsteady flows. Filella et al. (2018) used numerical simulations to demonstrate the advantage of cohesive schooling in a large group of swimmers emerging from an in-line configuration [48].

To identify the rules governing individual interactions in schools, many studies involving laboratory experiments have been conducted. Among these, the study by Katz et al. (2011) stands out [33]. These authors studied interactions between golden shiners (*Notemigonous crysoleucas*) swimming in groups of 2, 3, 10 and 30. The analysis of trajectories showed that individual motions do not result from linearly averaging the input from each separate neighbour (as it was commonly assumed in previous research on collective behaviour), hence highlighting the complexity of group dynamics and multi-information processing. Their work led to a successive study on collective states (i.e. swarm, milling, and polarized groups) that analysed the emerging coherent patterns of fish groups [36]. Even though these studies reveal exceptional insights into collective behaviour, no information can be derived regarding response to hydrodynamics, as they were conducted in shallow aquariums with quiescent water.

In uniform flows at different mean velocities, Ashraf et al. (2016) studied pairs of red nose tetra fish (*Hemigrammus bleheri*), revealing an increase in synchronisation of the caudal fin between pairs with increasing speed, possibly related to energy savings [49]. In addition, they observed that the distance between nearest neighbours stayed constant, arguing that it was maintained by visual contact. In contrast, Li et al. (2020) showed that

pairs of goldfish (*Carassius auratus*), regardless of whether their vision or lateral line sensing were impaired, synchronized their tail-beat movement depending on their front-back distance (a strategy referred to as 'vortex phase matching' [50]). De Bie et al. (2020) observed that the preferred configuration of pairs of Eurasian minnows (*Phoxinus* phoxinus) depended on the mean flow velocity [51]. Fish showed a higher tendency to swim closely in a side-by-side configuration at high velocities, whereas, in the absence of flow, they tended to minimize visual impediments by employing a 'leader-follower' strategy. In another study by Ashraf et al. (2017), the 3D positions of red nose tetra fish were recorded at different flow velocities in shallow water, finding that fish tended to swim side-by-side at high velocities [52]. Although position was recorded in the vertical coordinate, the water depth in these experiments was very shallow - only 2.2 cm, resulting in a depth-to-Body-Length ratio of 0.6. This shallow depth is not fully representative of behavior in a truly three-dimensional flow domain.

Overall, these studies consistently demonstrate that schooling is very much affected by mean flow velocities in ways that seem to promote energy-saving in predominantly shallow flows where swimming is essentially constrained over two dimensions. Very little is known about energy-saving strategies adopted by individuals swimming in three-dimensional flow domains. Moreover, the effects of other hydrodynamic metrics on schooling, such as turbulence, remain to be studied.

Understanding these dynamics is crucial for the design of fish passages, which aim to recreate suitable hydrodynamic conditions for migrating fish. Currently, principles of fish pass design aiming at recreating suitable hydrodynamic conditions are commonly based on swimming performance curves of fish swimming alone. As far as energy expenditure is concerned, this probably represents a conservative (hence effective) design approach. However, it does not account for the complex energy-saving behaviors seen in schooling fish in natural environments. By gaining a deeper understanding of how fish in schools conserve energy while swimming in hydrodynamically complex environments, we can improve the design of fishways to better support the natural behavior and energy efficiency of migrating fish, thereby increasing passage efficiency and reducing passage timing.

1.2.2 Navigation

Migration is often undertaken in large groups for many taxa, including fish [53,54]. Experimental studies showed that fish can better follow migration routes with other conspecifics and that homing accuracy usually increases with group size [55–57]. Over time, different explanations for this phenomenon have been suggested.

According to the '*many wrongs*' principle, groups can average out information errors individuals make, resulting in higher accuracy [58–60]. The *leadership* theory, instead, considers navigation led by informed individuals that guide the rest towards the correct destination [57]. Another explanation is that information passes progressively from informed individuals to uninformed ones over time, what is called *social learning* [61–63]. *Collective learning* consists instead of the generation of navigation information of the group over time through social interactions [64]. Finally, *emergent sensing* consists of the ability of the group as a whole to sense and follow gradients through social interactions, like a distributed sensor [55].

These mechanisms are not mutually exclusive, and navigation can result from a complex interplay between them, which might depend on species, life stage, environment, and navigational task [65]. In fact, navigation is not used exclusively for migration but also for other tasks such as finding food, shelter, or the entrance of a fish pass [65–67]. As a result, collective navigation may result in an easier and faster discovery of a fish pass entrance and a lower passage time [68].

1.2.3 Stress

Since schooling provides antipredator benefits thanks to dilution and confusion effects [69–72], it may reduce the need for vigilance and thereby reduce stress. In salmonid species, social hierarchies can cause social stress [73–76], but for most gregarious animals, the presence of conspecifics results in physiological benefits by significantly lowering metabolic rates. This phenomenon - commonly referred to as the 'calming effect' - has been observed in different gregarious fish in both standing waters and flowing conditions [77–80], regardless of whether or not they swim in schools [81]. Moreover, social interactions can result in a faster recovery from stress, as solitaire fish take longer to return to a normal physiological state after an acute stress event (e.g., a simulated predator attach or brief air exposure) than fish in groups [79,82,83].

Results from the studies on stress response presented here highlight the importance of collective behaviour in fish passes from a physiological perspective. In fact, fish passes must not only facilitate the movement across a barrier but also ensure that the process does not cause excessive stress on the fish. This is because the crossing of a fish pass has the potential to cause acute stress in fish, which can temporarily impact various physiological aspects, such as metabolism [84], memory [85], and locomotion [86]. From this point of view, it seems imperative to design fishway hydrodynamics to avoid as much as possible group fragmentation not only to improve navigation but also to minimise stress levels during migration. Lower stress levels in groups can also affect behaviour and potentially increase the passage performance of migratory fish [87].

1.2.4 Exploration

The trade-off between searching for resources and avoiding predators affects the degree of exploratory behaviour of fish. Exploration is usually linked with an individual's inclination to take risks (also referred to as *boldness*) and results from an interplay between personality traits, physiological needs and external conditions [88]. "Animal personalities" can be described as consistent and stable differences in behaviour within a population that are consistent across various contexts [89]. Individuals with different behavioural types or "personality traits" exhibit suites of correlated traits, which are also known as "syndromes" [89]. In groups, there will be fish that tend to assume bolder or dominant behaviour compared to other companions [90], changing pathways of information transfer among group members [91]. As a result, group personality composition can influence individuals' interactions with their environment and other group mates, thus affecting foraging performances, collective decision-making and predation avoidance [92]. The willingness of a fish to take the risk to explore unfamiliar or unsheltered environments will also depend on its physiological needs (e.g. foraging or reproduction) and external conditions, such as light, predation, or hydraulic stimuli [93–98].

Social facilitation is the mechanism that refers to an increased level of boldness in animals in the presence of other conspecifics. Different studies proved that this phenomenon occurs in some freshwater species [99–101]. Magnhagen and Bunnefeld (2009) observed that fish that alone were the shiest displayed the most changes in behaviour, becoming bolder when in a group, whereas fish that exhibited a bolder personality trait kept similar behaviour in the presence of conspecifics [101]. Cote et al. (2011) also showed that fish dispersal was affected by group mean boldness score, with individuals being more likely to disperse with bold members regardless of their own personality type [102].

From a fish passage perspective, a fish pass represents a novel environment. For some species, social facilitation might affect the willingness of a fish to approach and enter a fish pass [103]. A study evaluating fish guidance structures in a downstream laboratory facility found that solitary barbels (*Barbus barbus*) tended to seek shelter in the flume and were reluctant to move, while small groups of barbels displayed higher motivation to swim through the flume [103].

1.2.5 Predation

The attraction of predators at the entrance of fish passes is a phenomenon observed at many sites across the globe. Fish shoals concentrating in these confined areas can induce intensified predation from other larger fish, birds, mammals or reptiles [104–107]. A

combination of high population densities and long resident time at the entrance of a fish pass can result in increased predation risks [108].

From this perspective, shoaling provides anti-predator benefits due to an interplay of lower predation probability ('dilution effect'), lower predator success ('confusion effect'), and higher vigilance of the group as a whole [69,109,110]. For instance, Landeau & Terborgh (1986) observed that largemouth bass (*Micropterus salmoides*) performed better in capturing solitary silvery minnows (*Hybognathus nuchalis*) than when they were in a group, taking much longer to prey as school size was increased, as a result of the 'confusion effect' [109]. The 'dilution effect' instead, indicates the lower chance for an individual to be captured when in the presence of other potential prey and has been observed in different experimental studies [69,71,72]. Additionally, individuals can better discern the presence of a predator by olfactory and visual cues provided by informed group mates [111,112]. Fish that have perceived the presence of a predator may change their behaviour by displaying a fright response, which other group mates can observe [110,113].

The presence of predators stalling at fish pass sites is mainly determined by the rich abundance of fish found at the entrance or the exit of such structures [114]. However, this concentration of prey may result from low fish pass efficiency and the consequent crowding at the fishway entrance, rather than from collective behavior itse. These findings highlight the importance of designing fishway entrances that are highly attractive and accessible to fish, to reduce predation risks and improve passage efficiency.

1.3 Implications for fish passage design and future research

The concepts presented here highlight the need to take group behaviour into account in fish passage science. We suggest taking collective behaviour into account throughout different phases of river management, including evidence-based design, implementation and monitoring. Certainly, engineers in charge of designing a fish pass could benefit from research into group behaviour of freshwater species.

For all the aspects previously discussed, individuals passing a fishway alone, separated from the rest of their group, may face problems continuing migration, particularly along a highly engineered river with multiple consecutive fish passes. Group fragmentation caused by a challenging and delayed movement through a fish pass may inhibit the ability of fish to complete their migration successfully. Although scientific evidence strongly support the role of collective behaviour during migration [56,67,115], little is known about the effect of fishway hydrodynamics on school fragmentation.

For instance, laboratory investigations could provide experimental-based knowledge on schooling mechanisms and preferred flow conditions by groups, providing vital information concerning fish pass attractivity. Schools could also be investigated to find velocity and turbulence thresholds at which groups are no more able to swim coordinately and transfer vital information within.

Field studies also represent a central source of information. As biotelemetry technologies such as acoustic, passive integrated transponders (PIT), and radio telemetry continue to develop rapidly, researchers are increasingly able to study fish groups in their natural habitats. Moreover, the development of user-friendly and accessible artificial intelligence techniques is converting large datasets, such as geographical position data and video recordings, to effective and time-saving sources of information on fish movement and behaviour, both in field and laboratory settings [116,117]. In the past, for instance, videos had to be manually examined by a person, but now automated object detection methods such as deep learning are becoming more available [118–120]. Thanks to all these technologies, it is now possible to comprehensively study the relationships between passage efficiency, group size and other abiotic factors, such as light, sound and hydrodynamics.

Also, installing monitoring frameworks, such as fish counters, could help to gain insight into the behaviour of fish at existing structures [121], even from a collective behaviour standpoint. The data provided by fish counters, which typically include the number of fish that pass through, the time taken to pass, and the number of attempts, can offer valuable insights into the efficacy of fish passage over extended periods of time. With regard to modelling, collective behaviour could be incorporated into existing approaches, such as Agent-Based Models (ABMs), often utilized to simulate fish movement and behaviour at strategic locations [122]. Increasing the prediction performance of such models would better inform engineers, biologists and practitioners involved in freshwater conservation management.

These new data sources, combined with the expertise of professionals and researchers from different disciplines, could result in new evidence-based guidelines for fish pass design with optimized geometry and hydraulic flow conditions for gregarious species.

Chapter 2. An AI-based method for fish tracking

Abstract [2]

This Chapter outlines the methodology developed to track fish positions in flow domains where optical access is disturbed by a changing background, e.g. a free surface of an open channel flow. At these conditions, classical computer segmentation software often proves inadequate for detecting small objects. The methodology comprises two primary components: detection and tracking. Detection involves identifying objects in an image against the background, and for this phase, a deep learning approach was employed. Two convolutional neural networks were trained to detect fish in videos showcasing an open channel flume's horizontal and lateral views, with fish swimming at various velocities. The validation of the convolutional neural networks resulted in F1 scores of 96% and 91% for the bottom and side videos, respectively. The second phase, tracking, involves assigning the identity to each detected object in an image to reconstruct its trajectory over time. An algorithm based on a Kalman filter with an embedded identity recognition approach was developed for this purpose. The tracking phase's validation included visual checks by an operator for overlapping between fish heads and tracking data. The tracking performance was assessed by quantifying the number of correct swaps between fish. The tracking approach achieved an 87% success rate in correctly capturing swaps for the bottom videos.

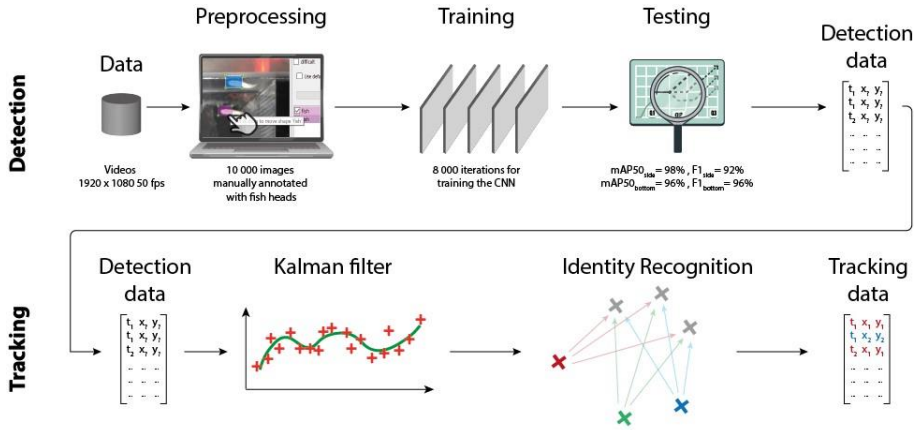

Figure 2.1: Video fish tracking pipeline: detection and tracking

[2] This chapter has been included as Supplementary Material in a work published in Scientific Reports.

Citation: Mozzi, G., Nyqvist, D., Ashraf, M. U., Comoglio, C., Domenici, P., Schumann, S., & Manes, C. (2024). The interplay of group size and flow velocity modulates fish exploratory behaviour. *Scientific Reports*, 14(1), 13186.

2.1 Introduction

In the realm of experimental research on animal behaviour, the utilisation of videos has become a pivotal tool over the past decades [123,124]. Videos enable the accurate identification and tracking of animals in controlled environments as well as in their natural habitats, offering invaluable insights into behaviour, population dynamics, and environmental changes. Manual extraction of data from videos, traditionally performed by operators, imposed limitations on the duration of experiments and the quantity of collected data [125]. In order to overcome these constraints, automated object detection has emerged as an essential advancement.

Early approaches to automated object detection in videos relied on handcrafted features and classical computer vision algorithms. Feature-based methods explored the use of distinctive features specific to animals, such as contour shapes, key body points, or texture patterns [126,127]. These approaches aimed to improve robustness and accuracy by focusing on characteristics unique to different species. However, the manual design of features limited the adaptability of these methods to diverse datasets and species [128]. Classical computer vision algorithms refer to traditional, pre-deep learning methods used in computer vision tasks. These algorithms are based on mathematical and heuristic techniques designed to process and analyse visual data. In the context of object detection, classical computer vision algorithms often include methods like edge detection [129], contour analysis [130], and color-based segmentation [131]. These methods can be applied in both air and water mediums. However, while effective in replicable laboratory conditions with still water, they face challenges in dealing with complex backgrounds and varying lighting conditions, as encountered in natural field settings or temporary setups - such as the one utilised in this thesis - which do not guarantee highly detailed replicable conditions.

The evolution of tools for animal detection in videos has transitioned over the years to more sophisticated Artificial Intelligence (AI) approaches [132]. The introduction of machine learning techniques marked a significant shift in object detection. Methods like Support Vector Machines (SVMs) and Random Forests were employed to learn discriminative features from training data [133,134]. While providing improved performance compared to handcrafted features, these techniques faced challenges in handling large datasets and achieving high precision in complex scenarios.

The advent of deep learning, particularly Convolutional Neural Networks (CNNs), revolutionised object detection in video [135]. CNNs demonstrated unparalleled success in learning hierarchical representations from raw data, automatically extracting features relevant to the task. For instance, an exemplary open-source toolbox for animal posture estimation is DeepLabCut, developed by Lauer et al. (2022) [136]. In scenarios where it is

pertinent to extract the animal's position within the picture frame exclusively, alternative techniques may be more favourable, particularly for non-computer scientists, owing to their ease of training. For example, region-based deep-learning architectures constitute valuable tools for object detection, ranging from the inception of Region-based CNNs (R-CNN) to subsequent improvements like Fast R-CNN and Faster R-CNN [137–139]. A shift away from region-based approaches is observed with architectures like Single Shot multi-box Detectors (SSD) [140], and You Only Look Once (YOLO) [141]. These approaches introduced the concept of anchor boxes, representing a departure from dynamically proposed regions in R-CNN. Anchor boxes offer a fixed set of bounding box priors for the model to adjust during training, facilitating a more streamlined and efficient real-time detection process.

You Only Look Once (YOLO) is a real-time object detection algorithm based on CNN, developed by Redmond et al. (2016) [141]. The algorithm operates by dividing the input image into a grid, directly predicting bounding boxes and class probabilities for each grid cell. In contrast to R-CNN methods, YOLO streamlines the object detection process by simultaneously handling both localisation and classification within a single neural network. This approach enhances computational efficiency, making it particularly suitable for real-time applications like video analysis and autonomous systems. Since its inception in 2016, YOLO has undergone several iterations, culminating in its latest version at the time of writing, YOLOv9 [142]. Each proposed version has consistently refined the algorithm's performance and adaptability across various object detection tasks [143]. In this study, we utilised YOLOv4 - the latest open-source version available at the time of the experimental campaign - as the primary tool for fish head detection in videos. Despite being among the earlier versions, YOLOv4 exhibited sufficient speed and accuracy, particularly since real-time processing was not required. In fact, fish detection was conducted in post-processing after the end of the experimental campaign.

After completing the detection phase in all its parts (i.e., training, testing, and implementation), the tracking phase follows. This phase transforms individual objects in each image frame into data points, forming trajectories that describe the object's position over time. A commonly employed method to accomplish this is the Kalman filter [144]. The Kalman filter, a recursive mathematical algorithm, is utilised to estimate the state of a dynamic system from a series of noisy measurements. It involves distinct phases, including prediction, update, and state estimation, making it particularly well-suited for ensuring a consistent and smooth trajectory of the object across frames, even in the presence of missing detections or uncertainties. In cases involving multiple fish, assigning identities before the filter’s estimation of their positions at a given time step is necessary. To address this requirement, we integrated a stochastic identity recognition routine within the Kalman filtering process. This routine is founded on comparing the predictions of the Kalman filter with the detections from the successive time step.

2.2 Video dataset

Videos for developing of the tracking methodology were captured during an experimental campaign conducted in May 2021. The experiments involved with wild juvenile Italian riffle dace (*Telestes muticellus*) swimming in an open-channel flume with transparent walls and took place in a hatchery near the capture location. White diffusing panels and LED lights, positioned on tripods, were used to create the highest contrast between the fish and the background. Detailed descriptions of the experimental set-up, fish capture, holding procedure, and the experimental protocol can be found in Chapter 3, section 3.2 “Materials and methods”, along with a schematic representation of the experimental apparatus (Figure 3.2).

Videos were recorded from both the bottom and the side positions of the flume using Sony FDR-AX43 camera (1920 × 1080 pixels, 50 fps). A total of 120 experiments were conducted across three different flow velocity and three group sizes (20 x 3 experiments with single fish, 10 x 3 experiments with two fish, and 10 x 3 experiments with six fish), each lasting 7 minutes. This resulted in a dataset comprising 2 520 000 frames. Examples of video frames from bottom and side camera are shown in Figure 2.2 and Figure 2.3 below.

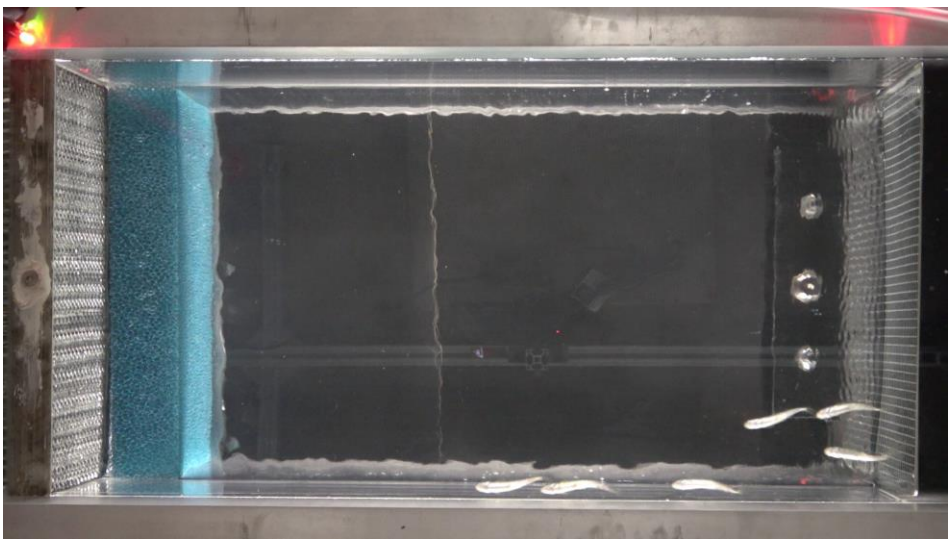

Figure 2.2: Frame sample from bottom-view
video recording six-fish experiment (low velocity treatment, 10cm/s)

Figure 2.3: Frame sample from side-view
video recording two-fish experiment (low velocity treatment, 10cm/s)

2.3 Detection

For the detection phase, two networks were trained - one for the bottom and one for the side videos, each with approximately 10 000 annotated images. The resulting F1 scores for the bottom and side Convolutional Neural Network (CNN) were 96% and 91%, respectively. Each CNN provided bounding boxes that enabled the detection of fish heads in each frame, with each detection providing the coordinates of the centroid of the bounding box, as well as its height, width, and confidence level [0,1] indicating the accuracy in representing a fish head, determined by the network.

2.3.1 Annotation

Both bottom and lateral video frames were annotated using LabelImg, a cross-platform GUI tool for image annotation, and saved in YOLO-compatible format. The annotation involved drawing a rectangle around the fish head. This approach was chosen because fish heads are more distinguishable and easier to detect compared to their entire bodies [118]. Initially, 2 000 images from single-fish experiments across three different flow velocities were manually annotated to (1 000 for bottom views and 1 000 for side views). Images from all velocity treatments were equally represented to ensure CNNs were

trained to detect fish under varying flow conditions (higher velocities resulted in more background disturbances).

Each of the preliminary CNNs was trained using these 1 000 annotated images and an additional 200 frames without fish. Subsequently, the CNNs were applied to annotate another 4 000 frames from experiments involving two-fish and six-fish groups (2,000 from bottom views and 2,000 from side views). The annotations generated were imported into LabelImg and manually adjusted as necessary. This iterative process continued until a total of 10 000 annotated images were obtained for each camera view.

2.3.2 Training

We trained two YOLOv4 CNNs using the Darknet framework [145]. The training dataset for the final CNNs consisted of approximately 10 000 images from each camera view (bottom and side). In order to optimise training efficiency, images were cropped into two parts and used for training. Additionally, images without fish were included to maximise detection performance. The parameters used to train the final CNNs are provided in Table 2.1.

Table 2.1: Training parameters for CNNs on bottom-view and side-view videos

Architecture	YOLOv4
Number of input images	10 000
Input image resolution*	1020 x 1080
Number of iterations	6000
Batch size	64
Subdivisions	64
Training\Testing ratio	85/15
Resolution after down-sizing	384x384

*images were divided in half, so the number of input images for the network doubled

2.3.3 Testing CNN performances

Various parameters can be used when evaluating the performance of neural networks for object detection. One commonly employed parameter is the F1 score [146], which represents the harmonic mean of precision and recall:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \cdot \frac{Precision * Recall}{TP + FPPrecision + Recall}$$

Where TP, FP, FN are the number of true positive, false positive and false negative, respectively. Another metric commonly used to evaluate detection performance is the mean average precision value (mAP) [141]. It is calculated by averaging each class's average precision (AP) values across different intersection-over-union (IoU) thresholds (e.g., mAP50 measures the mean average precision at an IoU threshold of 0.5). The AP represents the area under the precision-recall curve for a specific class. The IoU threshold determines the required overlap between the predicted bounding box and the ground truth bounding box for a detection to be considered correct. A higher mAP50 value indicates a more accurate network alignment with the mark around the fish's head.

Table 2.2 and Table 2.3 present an overview of how these performance indicators vary with increasing iteration numbers for the bottom and side CNN, respectively. After two training sessions with 6,000 iterations each, the indicators for the bottom CNN were as follows: precision, recall, mAP50, and F1 score were all 96%. For the side CNN, the precision was 90%, the recall was 93%, mAP50 was 98%, and the F1 score was 92%.

In order to filter the results of the detection process, a confidence level threshold of 0.3 was applied [147]. In the case of false positives - where more detections were obtained than the actual number of fish present in the flume - detections with higher confidence levels were prioritised and selected.

Table 2.2: Bottom-view CNN performances with the training iteration number. Number of detections, ground-truth, mean average precision (50), precision, recall, F1 score, True Positive (TP), False Positive (FP), False Negative (FN) and average Interception over Union (IoU).

	Iterations						
	0	1000	2000	3000	4000	5000	6000
Detections	2653347	10841	7213	7774	7189	7943	8055
Ground-truth	3574	3574	3574	3574	3574	3574	3574
mAP0.50	0	90.8	93.2	94.2	94.5	95.8	95.9
Precision0.25	0	0.93	0.95	0.96	0.95	0.96	0.96
Recall0.25	0	0.90	0.95	0.95	0.95	0.95	0.96
F1	0	0.91	0.95	0.95	0.95	0.96	0.96
TP	0	3207	3384	3383	3396	3413	3415
FP	2541117	254	174	144	174	143	141
FN	3574	367	190	191	178	161	159
Average IoU	0	67.5	73.3	73.8	74.1	74.8	75.2

Table 2.3: Side-view CNN performances with the training iteration number. Number of detections, ground-truth, mean average precision (50), precision, recall, F1 score, True Positive (TP), False Positive (FP), False Negative (FN) and average Interception over Union (IoU).

	Iterations						
	0	1000	2000	3000	4000	5000	6000
Detections	12530	12530	5890	6377	5535	5075	4882
Ground-truth	2963	2963	2963	2963	2963	2963	2963
mAP0.50	0	90.62	96.66	96.13	97.38	97.48	97.45
Precision0.25	0	0.73	0.87	0.87	0.88	0.90	0.90
Recall0.25	0	0.80	0.90	0.91	0.92	0.94	0.93
F1[1]	0	0.76	0.88	0.89	0.9	0.92	0.92
TP[2]	0	2375	2674	2696	2730	2774	2769
FP[3]	899534	899	411	389	376	314	317
FN[4]	2963	588	289	267	233	189	194
Average IoU	0	52.74	68.04	68.69	70.27	73.52	73.48

2.3.4 Detections along the vertical coordinate

In the side videos capturing six-fish groups, a high occurrence of fish swimming side-by-side at the bottom of the flume was observed. This led to numerous complete occlusions, where fish positioned closer to the camera would obstruct those swimming next to them. To assess the relevance of this issue, the number of missing detections was quantified. For each side video, the missing detections (MD) were calculated as:

$$MD_{N_{fish}} = 1 - \frac{N_{CNN\ det}}{N_{fish} \cdot N_{frames}}$$

Where $N_{CNN\ det}$ is the number of CNN detections in the trial, N_{fish} is the group size (i.e. number of fish present in the flume) and N_{frames} is the number of frames in the video trial. On average, MD_1 (i.e. MD_f with N_{fish} =1, which is the case of single fish) was approximately 16%. This value was attributed to fish swimming in low-contrast areas with the background, such as the upstream honeycomb, or when they were at rest or impinged against the downstream grid. In the case of two-fish groups, the average MD_2 was around 24%, while for six-fish groups, MD_6 escalated to 45%. The increasing number of MD_f with group size (N_{fish}) can be attributed to the numerous occlusions arising from fish swimming side-by-side in the side camera view. Although the rate of MD_1 for single fish was deemed acceptable for input into our tracking algorithm, the elevated occlusion rate for six-fish groups hindered the efficacy of the custom-made algorithm described below. Vertical space use analysis was therefore performed from CNN detections without additional filtering. It is worth emphasizing that the elevated occurrence of occlusions primarily took place in the lower section of the flume. This means that the fish's depth (i.e., distance from the flume bottom) was slightly overestimated for two-fish groups and significantly overestimated for six-fish groups. This overestimation was not considered a limitation, as it provided a conservative result aligning with the assumption made for our horizontal 2D analysis (see Chapter 3).

2.4 Tracking

For the tracking phase, a custom algorithm developed in MATLAB [148] was used to convert CNN detections (i.e., positions of the identified fish in single video frames) into tracking trajectories (i.e., the path taken by each fish over time through the video sequence). This algorithm is based on a Kalman filter approach [144], which embedded an identity recognition routine employing a stochastic approach to assign fish identity in videos featuring two or six fish. This routine systematically analyses all possible combinations of fish trajectories between two consecutive frames and selects the solution that minimises the overall group acceleration during this interval. This relatively straightforward method was developed for its compatibility with integration into the Kalman filter and its computational efficiency, particularly for our most data-intensive videos (i.e., six-fish groups). The algorithm parameters were set to maximise tracking performance during occlusions (i.e., frames in which one or more fish in a video sequence become partially or entirely obscured by other objects or the environment). The performance of the tracking algorithm was validated using a randomly selected sample of 2 minutes of videos featuring six-fish groups. By comparing the algorithm's output with a visual check conducted by an operator, it was determined that the algorithm accurately assigned identity in 87% of occlusions.

2.4.1 Kalman filter

A Kalman filter [144] was implemented to retrieve a filtered trajectory of the fish from the noisy and/or missing detections (i.e. measurements). The filter applied uses a 2D (*x,y*) model with constant velocity.

The state vector describes the state of the fish at a certain time-step *t*:

$$\vec{x}_t = \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{bmatrix}$$

Which comprises the position coordinates (x, y) and their velocity $(\dot{x}, \dot{y})$. After every time step (Δt), the model produces a predicted state x^p and an estimated state x^e based on actual measurements (z_t) of the object. Being $x_{CNN,t}$ and $y_{CNN,t}$ the detections from the CNN at the time step t, the actual measurement $(\vec{m}_t)$ can be calculated as:

$$\vec{m}_t = [x_{CNN,t}, y_{CNN,t}, \frac{x_{CNN,t} - x_{CNN,t-1}}{\Delta t}, \frac{y_{CNN,t} - y_{CNN,t-1}}{\Delta t}]^T$$

Initial state

The initial estimated state will start from the initial measure:

$$x_0^e = m_0 = \begin{bmatrix} x_{CNN,0} \\ y_{CNN,0} \\ 0 \\ 0 \end{bmatrix}$$

In order to begin the prediction of the next state, the estimate uncertainty matrix (or covariance matrix) needs to be computed. This matrix represents the covariance between the error in the estimated state and the error in the measurement of the state and provides an indication of the accuracy of the state estimate. Assuming no correlation between estimation errors on x and y, in position ($covar(x, y) = 0$) and velocity (i.e., $covar(\dot{x}, \dot{y}) = 0$), the covariance matrix becomes a diagonal matrix. The initial covariance matrix is:

$$V_0 = \begin{bmatrix} var(x) & 0 & 0 & 0 \\ 0 & var(y) & 0 & 0 \\ 0 & 0 & var(\dot{x}) & 0 \\ 0 & 0 & 0 & \mathrm{var}(\dot{y}) \end{bmatrix}$$

Prediction

The current predicted state is based on the previously estimated state:

$$x_t^p = Ax_{t-1}^e$$

Where A is the dynamic matrix (or state transition matrix):

$$A = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Also, the predicted covariance matrix is updated based on the previous step:

$$V_t = AP_{t-1}A^T + Q$$

With Q being the process noise covariance matrix, which represents the uncertainty of the process and measurement noises:

$$Q = \sigma_a^2 G \cdot G^T$$

Being σ_a the standard deviation on the acceleration and G:

$$G = \begin{bmatrix} {}^1/_2\,\Delta t^2 \\ {}^1/_2\,\Delta t^2 \\ \Delta t \\ \Delta t \end{bmatrix}$$

Estimation

After the prediction, the estimation is updated with the actual measurement z_t.

$$x_t^e = x_t^p + K(m_t - Hx_t^p)$$

Where H is the measuring matrix:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

And is K the Kalman gain:

$$K_t = \frac{V_{t-1}H^T}{HP_{t-1}H^T + R}$$

With R being the noise covariance matrix:

$$R = \begin{bmatrix} \sigma_x & 0 & 0 & 0 \\ 0 & \sigma_y & 0 & 0 \\ 0 & 0 & \sigma_{\dot{x}} & 0 \\ 0 & 0 & 0 & \sigma_{\dot{y}} \end{bmatrix}$$

The covariance matrix is also then updated:

$$V_t = (I - K_t H) * V_{t-1}$$

The Kalman gain K indicates whether the model should trust the measurement ($K \to 1$) or the prediction ($K \to 0$). With this method, at the time steps where there is a missing detection, the Kalman gain will be zero, and the state of the fish will be estimated from previous states (assuming constant velocity).

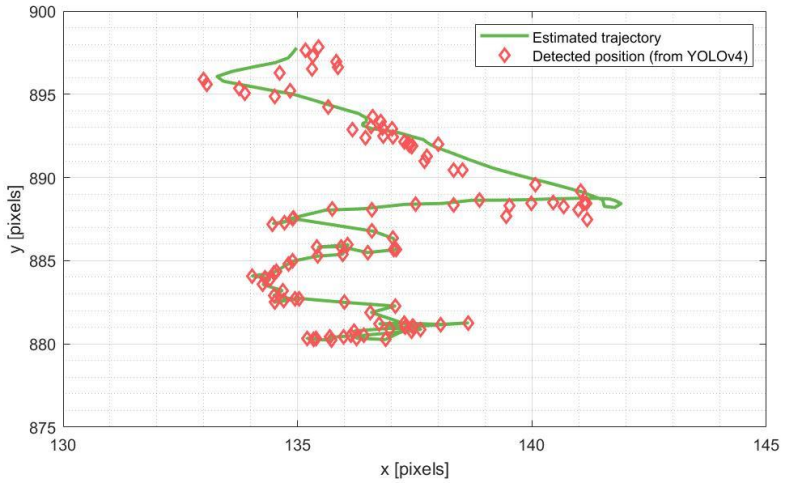

Figure 2.4: Example of observed and estimated position for single fish.
Data refer to the initial 100 frames (2 sec) of the first experiment with single fish, low velocity (10cm/s).
Note that 1 cm ≈ 30 pixels.

Velocity dumper

By using a constant velocity model, the filter estimates an undetected object by following a linear trajectory. As a result, for time periods in which the object is not detected, the estimated position keeps moving at a constant velocity, and when the non-detection interval is long, the object can also move out of the boundaries. In order to solve this issue, a condition on the prediction was added if the position was not measured:

$$if\ m_t = NaN\ \rightarrow\ x_t^p = D \cdot Ax_{t-1}^e$$

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & d_v & 0 \\ 0 & 0 & 0 & d_v \end{bmatrix}$$

Where d_v $(0 \leq d_v \leq 1)$ is a dumper that reduces the velocity from the previous time step.

2.4.2 Identity recognition

In the case of a single fish, the Kalman filter can be applied to the measurement to retrieve a filtered trajectory. For multiple fish, however, it is necessary to relate the object of a new time step to the previous one (see Figure 2.5). This was done by assigning a probability to all possible configurations of the fish identities between consecutive time

steps (permutations). Only two combinations (Np =2) are possible in the case of two fish. In the case of 6 fish, however, there are 720 possible permutations (Np =720).

For each configuration, the sum of the norms of the acceleration vectors is calculated.

$$b_{i,t} = \sum_{k}^{Nfish} \left|\vec{a}_{k,t}\right|$$

$$\vec{a}_{k,t} = \frac{\vec{v}_{k,t} - \vec{v}_{k,t-1}}{\Delta t}$$

$$\vec{v}_{k,t} = \frac{m_{k,t} - M_1 \cdot x^e_{k,t-1}}{\Delta t}$$

$$\vec{v}_{k,t-1} = M_2 \cdot x^e_{k,t-1}$$

Where M_1 and M_2 are matrixes that extract position and velocity from the state vector, respectively:

$$M_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The probability of each configuration q is then calculated as:

$$q_{i,t} = 1 - \frac{b_{i,t}}{\sum_{j=1}^{Np} b_{j,t}}$$

$$\begin{cases} if\ Nfish = 2 \ \rightarrow Np = 2 \\ if\ Nfish = 6 \ \rightarrow Np = 720 \end{cases}$$

Identity was therefore assigned by selecting the configuration with the highest probability among the *Np* possible permutations:

$$\overline{\overline{b_t}} = max_{i=1..Np}(q_{i,t})$$

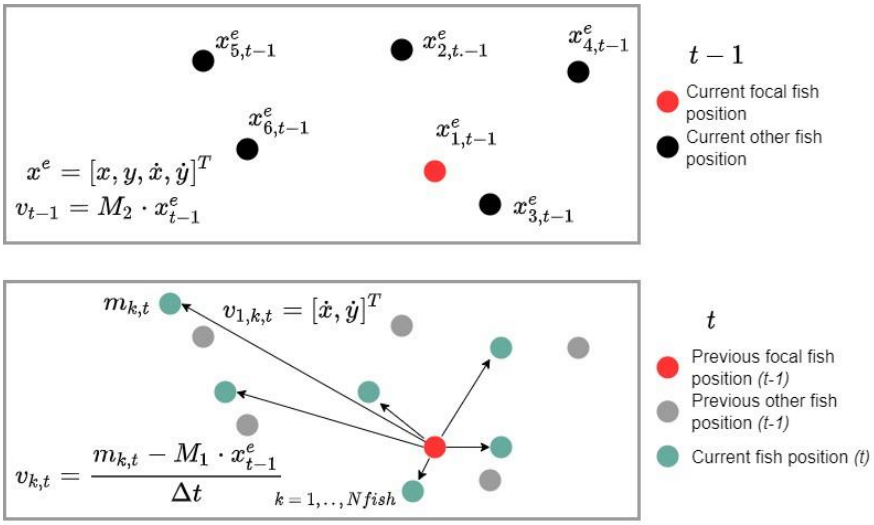

Figure 2.5: Velocity combinations of six-fish group for two consecutive time-steps

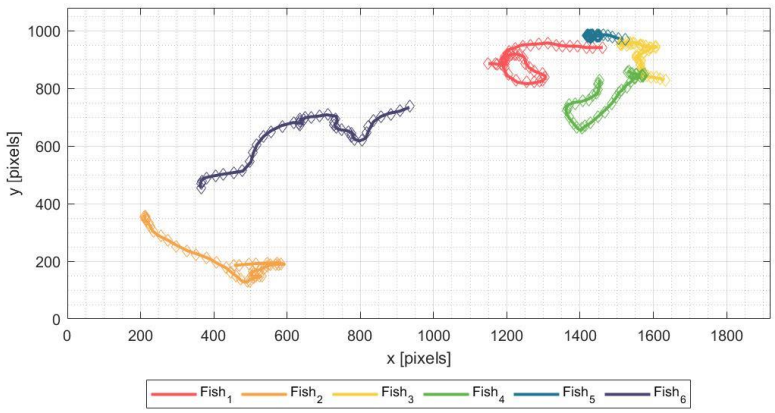

Figure 2.6: Example of observation and estimated positions for six fish.
Data refer to the initial 100 frames (2 sec) of the first experiment with six fish, low velocity (10cm/s). Note that 1 cm ≈ 30 pixels.

2.4.3 Validation

The tracking performance of videos captured by the bottom camera was assessed by evaluating the accuracy of the identity algorithm. The primary objective was to examine the assignment of identities during occlusions, which occurred when two or more fish crossed paths. The algorithm's performance was visually verified to determine if the assigned identities were correct or if there were instances of swapping (see Figure 2.8). In order to minimise the occurrence of identity swaps during occlusions, filter parameters were optimised using a trial-and-error approach. Visual confirmation was conducted on a randomly selected two-minute subset of videos from three velocity treatments, resulting in a total of 50 occlusions. The optimised filter parameters yielded an identity accuracy of 87% for the analysed occlusions.

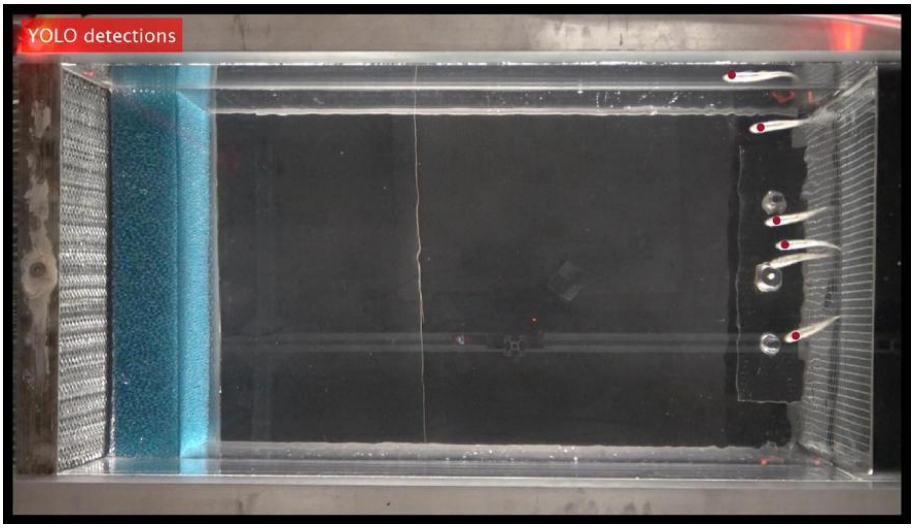

Figure 2.7: Detections from YOLOv4 over frame

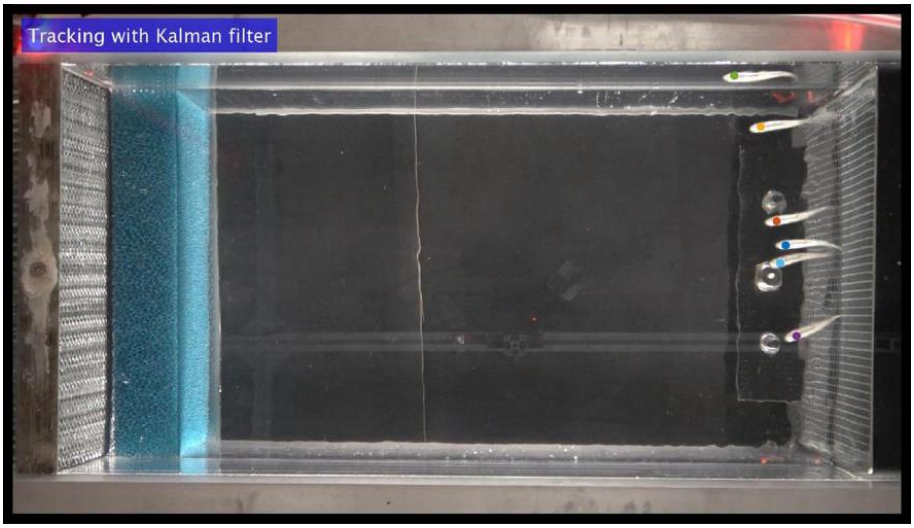

Figure 2.8: Results from tracking from Kalman filter with embedded the identity recognition algorithm

Chapter 3. Social facilitation in exploring fish under different flow velocities

Abstract [3]

Social facilitation is a well-known phenomenon where the presence of conspecifics enhances an individual's performance in a specific task. As far as fishes are concerned, most studies on social facilitation have been conducted in standing-water conditions. However, especially for riverine species, these conditions represent the exception rather than the rule, and the effects of hydrodynamics on social facilitation remain largely unknown. To bridge this knowledge gap, we designed and performed flume experiments where the behaviour of a gregarious fish species (*Telestes muticellus*) in varying group sizes and at different mean flow velocities was studied. An artificial intelligence (AI) deep learning algorithm was developed and employed to track fish positions in time and subsequently assess their exploration, swimming activity, and space use. Results indicate that energy-saving strategies seem to dictate space use in flowing waters regardless of group size. Instead, exploration and swimming activity increased by increasing group size, but the magnitude of this enhancement - which quantifies social facilitation - was modulated by flow velocity. These results have implications for how future research efforts should be designed to understand the social dynamics of riverine fish populations, which can no longer ignore the contribution of hydrodynamics.

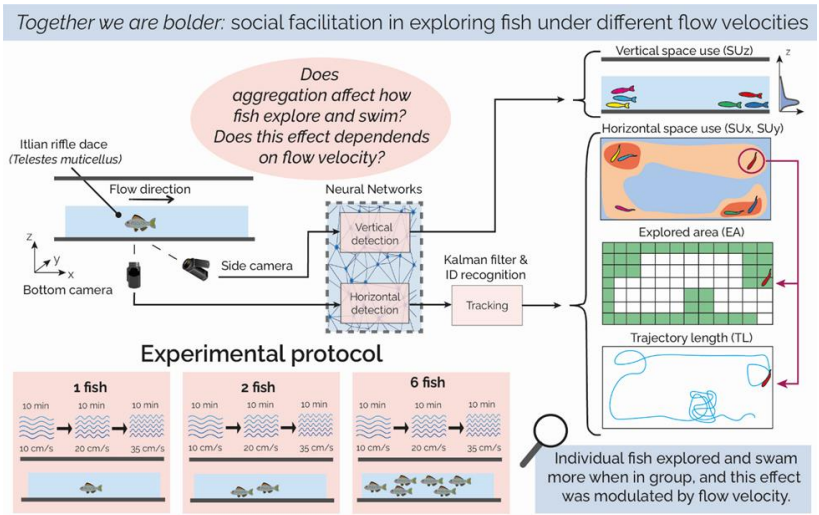

Figure 3.1: Graphical abstract of Chapter 3

[3] This chapter has been published in Scientific Reports.

Citation: Mozzi, G., Nyqvist, D., Ashraf, M. U., Comoglio, C., Domenici, P., Schumann, S., & Manes, C. (2024). The interplay of group size and flow velocity modulates fish exploratory behaviour. *Scientific Reports*, 14(1), 13186.

3.1 Introduction

Social facilitation refers to the phenomenon in which an individual's performance in a particular task is enhanced by the presence of conspecifics [149]. It has been extensively studied in diverse organisms, ranging from insects to birds and mammals [150–152], highlighting its importance in understanding social dynamics and ecological processes. In the aquatic realm, fish, with their varied social structures and diverse behaviours, offer a fascinating opportunity to explore social facilitation in different contexts. Among gregarious species, fish can display social facilitation in relation to foraging [153,154], predator avoidance [155,156], and exploration [100,101,157,158]. For instance, during foraging, groups of minnows (*Phoxinus phoxinus*) and goldfish (*Carassius auratus*) exhibited a reduction in duration and frequency in hiding behaviour as group size increased (2, 4, 6, 12, and 20 individuals) [157]. Interestingly, this group size effect was more marked on minnows than goldfish, possibly due to the higher schooling tendency of the former species. A study on three-spined sticklebacks (*Gasterosteus aculeatus*) reported that groups exhibited higher activity levels and demonstrated quicker resumption of foraging when exposed to simulated predation risk compared to a fish alone [158]. Another experiment showed that when young-of-the-year perch (*Perca fluviatilis*) were tested in single versus four-fish groups, the fish in groups spent more time exposed and approached areas close to a predator for feeding more quickly than single fish [101].

Although the energetic benefits of schooling are widely recognised [41–44], the works on social facilitations are typically carried with fish in standing waters. In riverine environments, however, these conditions represent only a limited portion of available habitats, and very little is known about how social facilitation acts on fish groups in running waters. Some hints that hydrodynamics might play an important role in the dynamics of fish groups can be gathered only from a handful of studies, which report that flow velocities can regulate the structure and polarization of fish schools [159], which, in turn, can affect information transfer among individuals [39]. Another limitation in the existing literature relates to the fact that most of the past experimental studies were performed either by forcing fish to swim in very shallow waters [52] or by neglecting their activity in the water column without thoroughly validating the exclusivity of horizontal motion [33,50,51,100]. This not only hampers the interpretation of results within the horizontal plane but also overlooks fish's spatial utilisation of the water column. Recognising how fish shoals navigate available space in flowing waters is important for understanding their social dynamics, including the exploitation of boundaries for energy saving.

This paucity of knowledge about how hydrodynamics affects social facilitation represents a bottleneck to understanding the dynamics of fish populations in riverine environments. Here, we focus on how being in a group influences the exploratory

behaviour of individual fish and, in turn, how this is affected by hydrodynamics. This is relevant because exploratory behaviour is key to acquiring vital resources such as food, shelter, and mating partners while also facilitating the acquisition of other valuable information about their surroundings, including the presence of predators and environmental changes [160]. From a practical standpoint, this holds relevance for developing efficient management strategies related to river flow regulation, restoration efforts, and the design of fish passage structures [161]. This is particularly crucial in light of the severe decline observed in migratory fish populations, often attributable to river fragmentation [14].

The aim of the present study is to investigate the role and interplay of social facilitation and mean flow velocity in dictating the exploratory and swimming behaviour of riverine fish. To accomplish this goal, we conducted flume experiments on groups of wild juvenile Italian riffle dace (*Telestes muticellus*) swimming in open channel flows. We hypothesise that *i*) group size increases the explored area (hereafter referred to as EA) and the swimming trajectory length (hereafter referred to as TL) of individual fish, meaning that social facilitation takes place and *ii*) the magnitude of this response is affected by flow velocity. Additionally, we investigated the preferred swimming areas, referred to as space use, to enrich the interpretation and discussion of results pertaining to EA and TL.

3.2 Materials and methods

3.2.1 Experimental set-up

Experiments were conducted in a hydraulic flume with water supplied by a recirculation pump connected to a downstream storage tank via a series of pipes (Figure 3.2). Flume walls were made of transparent Perspex with an aluminium frame. The test section was delimited upstream by a honeycomb aluminium flow straightener and downstream by a movable stainless-steel grid with a mesh size of 0.5cm x 0.5cm. The test section had a maximum volume of 30 cm x 150 cm x 30 cm; its length could be regulated by moving the downstream grid, while the pump frequency (controlled with an inverter) and the height of a downstream sharp-crested weir controlled the flow rate and depth. Pictures of the experimental apparatus can be found in the Appendix (Figure S 1).

A PT100 thermoresistance was used to measure the temperature of the water, and a 440 W chiller regulated the temperature. A flowmeter (AquaTransTM AT600, Baker Hughes) installed on the steel pipe at the bottom of the collection tank measured the flow rate. Water level was measured 5 cm upstream of the downstream grid with an ultrasonic sensor (BUS0025, BALLUFF). A data acquisition device (DAQ USB-6002, National Instruments) collected and transferred data to a personal computer for logging. The bulk flow velocity, calculated by multiplying the flow rate with the fixed width of the channel and the water level monitored through the ultrasonic sensor, was consistently monitored on the computer from a LabVIEW interface (National Instruments). The data acquisition device also allowed for the control of the lighting of three LED diodes installed beneath the flume within the camera view, used to synchronise video recorded from multiple cameras.

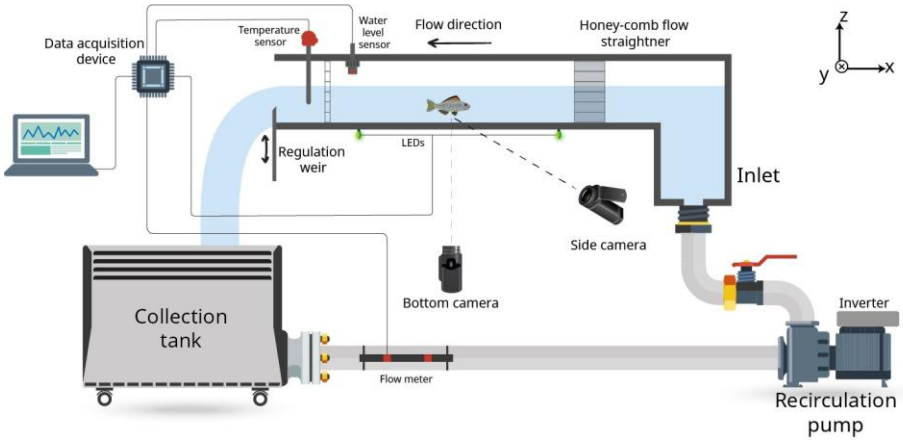

Figure 3.2: Schematic representation of the experimental set-up

3.2.2 Fish capture and holding

Italian riffle dace (*Telestes muticellus*) is a small-sized (<15 cm), omnivorous, gregarious fish native to streams and rivers on the Italian peninsula (Figure S 2a in the Appendix) [162]. About 100 juvenile Italian riffle dace of average length of 5.14 cm (SD ± 0.34 cm) were captured through electrofishing in Noce stream (44°56′18.52″N 07°23′11.24″E) in northern Italy (Figure S 2 b and c). Fish were then taken to a nearby hatchery and placed in flow-through tanks. After one week, fish were transferred to an artificial pond fed by river water with an average temperature of 15.4 °C (SD ± 0.9°C) (see Figure S 3a in the Appendix). Here, they could feed ad libitum thanks to the organic matter incoming from the stream.

The fish were relocated to three separate tanks (each of 200 x 60 x 15cm^3) supplied with spring water at 13.6 °C (SD ± 1°C) inside the hatchery 90 hours before the start of the experiments to acclimatise to the hatchery conditions. Each tank was divided into three compartments, with dimensions of 60 x 60 x 15cm^3, each hosting less than 20 fish (Figure S 3c). Opaque plastic panels were placed atop to provide a protective covering against potential disturbance caused by passing operators. Experiments were conducted in June 2021, and water temperatures in the hatchery tanks and the outside pond were continuously monitored throughout the experimental period.

3.2.3 Experimental protocol

Randomised experiments utilised three different group sizes (one, two, and six fish), exposing all fish to a sequence of mean flow velocities: low, medium, and high (10, 20, and 35 cm/s, respectively). We adopted this specific protocol with increasing velocity due to the substantial number of fish required. Testing flow velocity separately would have necessitated 300 wild fish of similar size to achieve statistically robust results. This quantity exceeded what was considered reasonable to extract from the river by the province fisheries authorities. Furthermore, the experiment timeframe was limited due to the approaching summer temperature increase. To conduct experiments with minimal temperature variation and a reasonable number of fish, we opted for an increasing velocity protocol, acknowledging the potential temporal dependence of velocity. However, we believe this choice did not impact our results; detailed discussions on this aspect are provided in Section 3.4.

The study comprised a total of 100 fish: 20 trials for single fish (i.e., 20 fish), 10 trials for two-fish groups (i.e., 20 fish), and 10 trials for six-fish groups (i.e., 60 fish), with each fish tested only once. Randomisation was obtained by employing a block sequence comprising two trials with single fish (labelled as “A”), one trial with two fish (“B”), and

one trial with six fish (“C”). Each unique combination of these trials within a block was assigned a number from 1 to 15 (e.g., 1 = “AABC”, 2 = “AACB”, 3 = “ABAC”, etc.). Subsequently, we utilised the '*rand*' command in MATLAB [148] to generate a sequence of 10 random numbers ranging from one to 15. This process yielded a random sequence of 10 blocks, each consisting of two single-fish trials, one two-fish trial, and one six-fish trial. Each group was exposed to three consecutive mean flow velocities: 10 cm/s for 15 minutes (5 minutes considered for acclimation [163] and 10 minutes for the actual trial), 20 cm/s for 10 minutes, and 35 cm/s for 10 minutes. Water depth was kept at 15 cm (i.e., about three fish body lengths), while water temperature was maintained at 14 °C (SD ± 0.3 °C) throughout the experiments.

If a fish rested at the downstream grid, it was gently tapped with a rod from behind, but no more than three times per group within the same velocity treatment. We adopted the gentle tapping method specifically because, on occasion - particularly post-release in the flume at 10 cm/s - fish tended to rest in the corners between the lateral walls and the downstream grid, aligning their bodies with the flow and with the tail resting on the grid. Observations indicated that this behaviour was not a result of fatigue: in a separate experiment focusing on fatigue, fish exhibited visible signs of hyperventilation, with their bodies flattened against the downstream grid. Consequently, we attributed the occasional tendency of fish to remain stationary more to volition or stress. To motivate swimming, we opted for gentle tapping [163], recognizing the slight interference with the experiment. In this context, we considered potential behavioural anomalies resulting from tapping to be comparable to the stress induced by the unfamiliar environment coupled with the capture-and-release process. While acknowledging the compromise introduced by tapping, we believe it is a reasonable trade-off to encourage swimming activity in the experimental setting. After each trial, the fish were anaesthetised in clove oil (Aromalabs, USA; approximately 0.05 ml clove oil/l water) and measured for weight and fork length. Some individuals underwent physiological analysis, assessing cortisol and oxidative stress levels in muscle tissue. The complete procedure and results can be found in Schumann et al. (2023) [164].

3.2.4 Data analysis

A reference system was defined with the origin at the honeycomb's bottom-right corner of the flume (upstream end of the experimental arena). The x-axis was aligned with the flow direction, the y-axis was set as transversal to the water flow in the horizontal plane, and the z-axis was defined as the vertical coordinate. Units of fish position resulting from the video analysis were converted from pixels to centimetres in the defined reference system using the coordinates of the four corners of the swimming arena for each video, as described in Nyqvist et al. [165].

Data analysis was conducted using R v4.0.5 [166]. Fish positions obtained from video-tracking were utilised to compute the following metrics: *i*) explored area (EA), *ii*) trajectory length (TL), and *iii*) space use, with the first two only analysed in the horizontal plane (x, y) – as vertical space use revealed that fish mostly stayed close to the flume bottom (see Section 3.3.2). EA was defined as the group-average ratio of the area covered by each fish to the total horizontal arena. To calculate EA, a 60 x 30 cm grid representing the experimental arena was utilised, and the cumulative count of newly covered 1 cm^2 cells by each fish along its path was determined and normalised by the total number of cells in the grid (i.e., 1800). Subsequently, the EA was calculated as the average of this count among all the fish in the investigated group. TL was determined as the horizontal distance swam by fish within a trial relative to the previously defined stationary reference system. Like EA, TL was calculated by averaging the trajectories of all fish in the group. Finally, space use was defined using two metrics: the first was $P_H(x, y)$, representing the probability density associated with a fish occupying a specific horizontal position (x, y). The second was $P_V(z)$, which represents the probability density of the vertical position of fish along the normal coordinate z. Both metrics were calculated utilising fish positions from all experiments conducted under the respective velocity and group size conditions. P_H and P_V were retrieved using a Kernel density estimation approach [167] with a fixed bin width of 1 cm for discretisation of both P_H and P_V in all spatial coordinates.

Plots of P_H and P_V were generated using the *geom_density_2d* and *ggridges* commands in the *ggplot2* v.3.4.0 [168] package in R, respectively. In the case of P_H, a logarithmic scale was employed to enhance the differences between trials. Estimating P_V posed challenges due to the numerous occlusions observed from the side camera's perspective. To derive P_V, a weighing approach was employed based on the total detections of each experiment. However, it is important to note that this approach led to an overestimation of P_V as group size increased. The overestimation was a consequence of fish swimming closely together at the bottom of the flume, resulting in overlapping within the lateral camera view. For more comprehensive details, please refer to the Chapter 2, Section 0.

The statistical analysis was conducted using the *rstatix* (v0.7.0 [169]) package in R with a significance threshold set at a p-value of 0.05. Space use medians of probability density functions $\widetilde{P_H}(x), \widetilde{P_H}(y)$ and $\widetilde{P_V}(z)$ were employed to characterise the overall positioning of fish along all three spatial coordinates ($P_H(x)$ and $P_H(y)$ are the marginal distributions of $P_H(x, y)$ and the tilde symbol stands for median of a distribution). For all analysed variables, each velocity treatment resulted in a sample size of N=[20,10,10] for [1,2,6] group sizes, respectively.

Data from EA, TL, $\widetilde{P_H}(x), \widetilde{P_H}(y)$ and $\widetilde{P_V}(z)$ did not fulfil the assumptions for parametric tests (Shapiro-Wilk test with p-value < 0.05). Consequently, non-parametric

tests were employed to examine the impact of group size and velocity on these variables. To test the hypothesis that social facilitation influenced EA and TL, we used the one-sided Jonckheere-Terpstra trend test, expecting larger effects for larger groups sizes [170,171]. Effects of group size on EA and TL was tested for each velocity separately. With regards to the effect of group size on space use (i.e., $\widetilde{\mathrm{P}_H}(x)$, $\widetilde{\mathrm{P}_H}(y)$ and $\widetilde{\mathrm{P}_V}(z)$), our focus was directed towards detecting differences rather than trends. As a result, we utilised the Kruskal-Wallis test for these variables [172]. The Friedman test was employed to examine the effect of flow velocity (within-subject factor), taking repeated measures across flow velocities into account [173]. In cases of significant effect of velocity from the Friedman test, we conducted post-hoc analyses using paired Wilcoxon signed-rank test[174]. To account for multiple comparisons, we adjusted p-values of post-hoc tests using the Bonferroni correction method [175].

In order to determine whether the results from the horizontal analysis were biased by fish movement in the water column, we conducted an additional bottom-filtered analysis. This analysis was performed by removing frames in which at least one fish in the group was detected higher than 5 cm from the bottom. Bottom-filtered explored area (EA*), trajectory length (TL*), and horizontal space use ($P_H^*(x,y)$), were computed for all the trials. As the trajectory length accumulates over the number of frames, TL* was adjusted by considering the number of non-filtered frames as follows:

$$TL^* = TL \cdot \frac{N_{frames}}{N_{bottom}}$$

Where N_{frames} represented the total number of frames in each trial, and N_{bottom} the number of frames in which no fish was detected above 5 cm.

3.3. Results

3.3.1. Explored area and trajectory length

The majority (75%) of fish displayed an EA lower than 20% and a TL lower than 13 m when alone in the flume (Figure 3.3). By contrast, fish belonging to groups of six had an average EA of 19% (SD ± 13%), 38% (SD ± 18%), and 26% (SD ± 12%) at 10, 20, and 35 cm/s, respectively. Most (75%) of individuals of the six-fish group had a TL higher than 13 m at medium and high velocities, but at low velocity, they swam on average 11 m (SD ±5 m). Fish swimming in pairs displayed an intermediate behaviour, with an average EA of 20% (SD ± 7%), 27% (SD ± 17%), and 17% (SD ± 11%) for low, medium, and high velocity, respectively. The average TL of fish in groups of two was 12 m (SD ± 4m) at 10 cm/s, 14 m (SD ± 5 m) at 20 cm/s and 12 m (SD ± 5m) at 35 cm/s (Figure 3.3). Overall, fish displayed the highest EA and TL in groups at medium and high velocities, while group size had no effect at lower flow conditions.

The Jonckheere-Terpstra test revealed a significant positive trend with group size on EA and TL only for the medium and high velocities. Please note that a positive trend indicates that EA and TL increase with group size. At a velocity of 10 cm/s, group size did not display a significant trend on EA ($z = 1.40$, p-value > 0.05) or TL ($z = 0.99$, p-value > 0.05). By contrast, at the velocity of 20 cm/s, group size significantly increased EA ($z = 3.27$, p-value < 0.001) and TL ($z = 3.42$, p-value < 0.001). At the maximum investigated velocity, 35 cm/s, trends in group size were significant for EA ($z = 3.49$, p-value < 0.001) and TL ($z = 4.33$, p-value < 0.001).

Friedman test showed that in fish belonging to groups of six, flow velocity exhibited a significant effect on both EA and TL ($\chi^2(2) = 11.4$, p-value $= 0.007$ for EA, and $\chi^2(2) = 20$, p-value < 0.001 for TL). For single fish, instead, velocity only showed significance for EA ($\chi^2(2) = 6.46$, p-value < 0.05) and not for TL (p-value > 0.05). Notably, in two-fish groups, velocity did not play a significant role in either EA or TL (p-value > 0.05). Post-hoc analysis revealed that flow velocity had a significant effect on EA for single fish between medium and high velocities (20 and 35 cm/s, respectively, Table 1). For six-fish groups instead, the effect on EA was significant in the transition between low and medium flows (10 and 20 cm/s, respectively), while the effect on TL was significant comparing all velocity conditions.

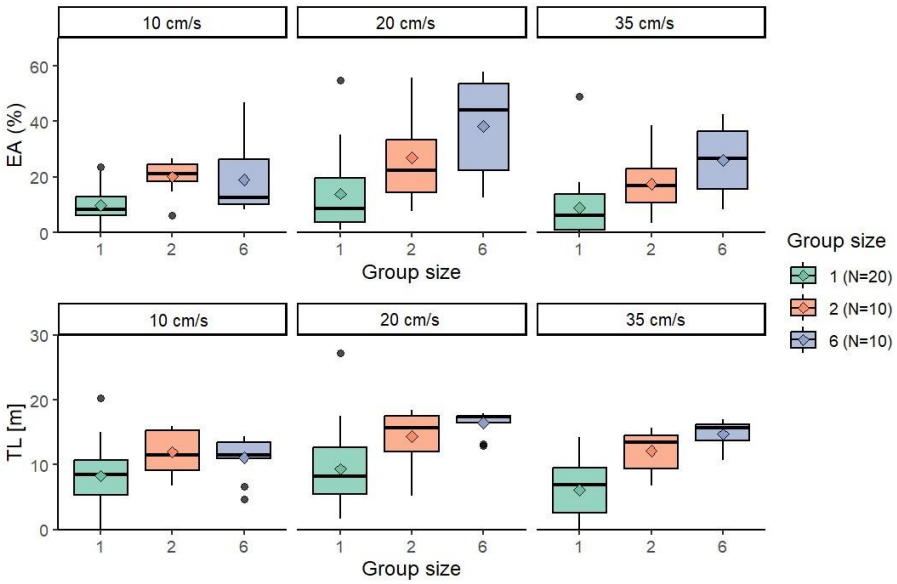

Figure 3.3: Explored area (EA, top) and trajectory length (TL, bottom) for each fish in the horizontal plane. In the case of two and six-fish groups, values represent the group average. The diamonds represent the mean values.

Table 3.1: Post-hoc analysis on the effect of flow velocity on explored area (EA) and trajectory length (TL) with paired Wilcoxon test with Bonferroni adjustment for p-values. NS stands for non-significant (p-value ≥ 0.05). For clusters in which the Friedman test did not show significance, post-hoc analysis was not computed (-).

Cluster	*Pairwise comparison*	*EA*	*TL*
Single fish	10 vs. 20 cm/s	NS	(-)
	10 vs. 35 cm/s	NS	(-)
	20 vs. 35 cm/s	0.0342	(-)
Two fish groups	10 vs. 20 cm/s	(-)	(-)
	10 vs. 35 cm/s	(-)	(-)
	20 vs. 35 cm/s	(-)	(-)
Six fish groups	10 vs. 20 cm/s	0.0078	0.0063
	10 vs. 35 cm/s	NS	0.0021
	20 vs. 35 cm/s	NS	0.0049

3.3.2. Space use

Horizontal probability density, P_H, revealed that at the lowest velocity, fish tended to stay close to the upstream honeycomb grid while they relocated towards the downstream end of the flume and closer to the side walls as flow velocity increased (Figure 3.4). For similar velocities, all group sizes tended to occupy similar areas. However, larger groups occupied a broader area compared to smaller groups at all three

flow velocities, and the central region was sparsely occupied, particularly for single fish. Fish occupied the largest area at the medium velocity (20 cm/s) for all group sizes. As far as vertical space use P_V is concerned, fish mostly remained within the first 5 cm from the flume bottom and swam closer to the bottom as velocity increased (Figure 3.5). Notably, the vertical analysis underestimated the actual number of fish at the bottom of the flume (see Chapter 2, Section 0). Velocity had a significant effect (Friedman test, p-value = 0.0231) on the longitudinal and vertical median positions (identified as $\widetilde{P_H}(x)$ and $\widetilde{P_V}(z)$, respectively) of the fish for all group sizes (Table 3.2). In contrast, lateral median position ($\widetilde{P_H}(y)$) was only affected by velocity for six-fish groups. Group size did not significantly affect the median position of fish (Kruskal-Wallis test, p-value > 0.05).

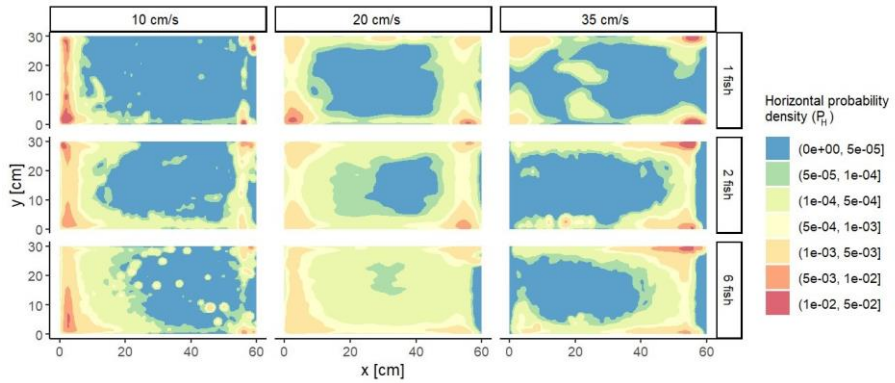

Figure 3.4: Horizontal probability density (P_H) illustrating fish distribution at various flow velocities (10, 20, and 35 cm/s) and group sizes (1, 2, and 6). P_H is obtained through Kernel density estimation, utilising fish positions (x, y) from all experiments conducted under the respective velocity and group size conditions. In each panel, water is flowing from left to right.

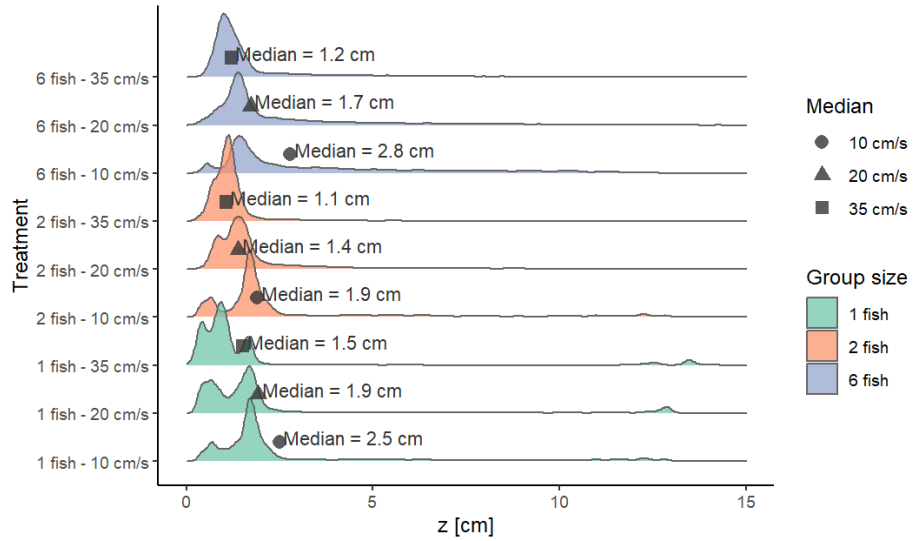

Figure 3.5: Vertical probability density (P_V) for different group sizes and velocities (bottom of the flume: z=0; free water surface: z=15cm). Each P_V is retrieved with Kernel density estimation of all vertical fish positions detected at the respective flow velocity and group size.

Table 3.2: Effect of group size and velocity on medians of probability distributions in the three coordinates (x, y, z) representing longitudinal, lateral and vertical positions, respectively. The effect of group size (between-subject factor) was assessed with Kruskal-Wallis test, while the impact of velocity (within-subject factor) was analysed through Friedman test (df=2 for all tests). NS stands for non-significant (p-value ≥ 0.05).

Factor	*Cluster*	$\widetilde{P_H}(x)$	$\widetilde{P_H}(y)$	$\widetilde{P_V}(z)$
Group size [1]	10 cm/s	NS	NS	NS
	20 cm/s	NS	NS	NS
	35 cm/s	NS	NS	NS
Velocity [2]	Single fish	0.0327	NS	< 0.001
	Two-fish group	0.0052	NS	0.0113
	Six-fish group	< 0.001	0.029	< 0.001

[1]*Between-subject factor: Kruskal-Wallis test,* [2]*Within-subject factor: Friedman test*

3.3.3. Bottom-filtered analysis

Upon visual comparison of bottom-filtered EA* and TL* (Figure 3.6) with the unfiltered data (Figure 3.3), it can be seen that the main trend remained consistent. Similar to the unfiltered analysis of the entire water volume, group size did not impact EA* and TL* at the lowest velocity (10 cm/s). However, EA* increased with larger group sizes at both medium and high velocities (20 and 35 cm/s, respectively), replicating the findings observed in the unfiltered analysis on EA. Similarly, the trend of TL* with group size was similar to that observed in the entire volume analysis (TL). Specifically, six-fish and two-fish groups exhibited comparable TL* values at the medium velocity, significantly higher than those of single fish. Conversely, at the highest velocity, an increase in TL* was observed with larger group sizes, perfectly mirroring the results from the unfiltered

analysis. By comparing differences in means, it can be seen that the variation in EA* stays below 7.5% (|EA-EA*|,

Table 3.3), while in TL is below 1.5 cm (|TL-TL*|,

Table 3.3). Also, regarding horizontal space use ($P_H^*(x,y)$, Figure 3.7), the distributions are highly similar to what observed in the entire volume ($P_H(x,y)$, Figure 3.4). Variations in medians positions are less than 2 cm for the longitudinal position ($|P_H^*(x) - P_H(x)|$,

Table 3.3) 1.1 cm for the lateral one ($|P_H^*(y) - P_H(y)|$,

Table 3.3).

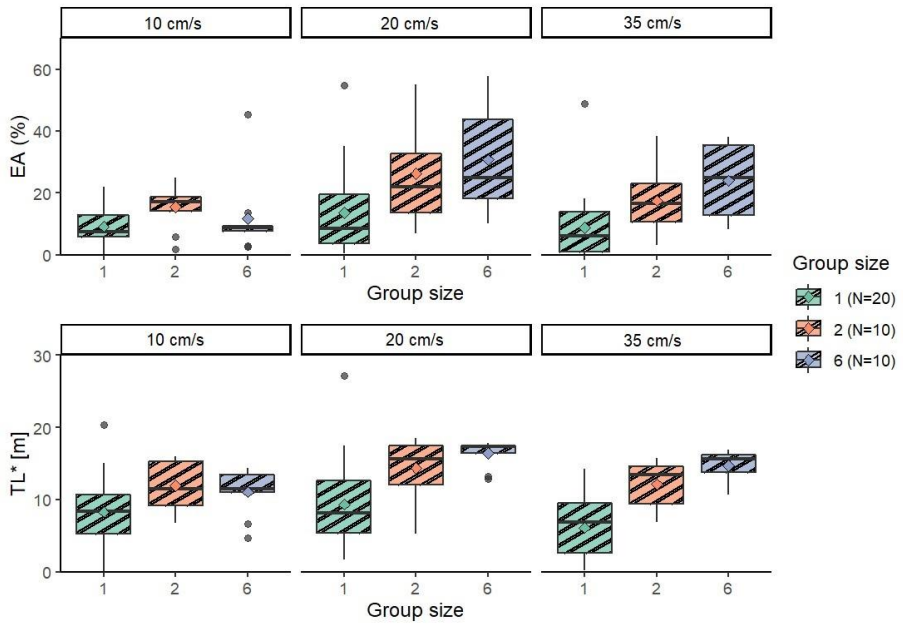

Figure 3.6: Bottom-filtered explored area (EA*, top) and trajectory length (TL*, bottom) for each fish in the horizontal plane. Filtering was implemented by considering only frames in which no fish was detected above 5cm from the flume bottom. Values are averaged for the group, and the rhombi represent the means. TL* is adjusted for the number of non-filtered frames ($TL^* = TL \cdot F_{tot} / F_{bottom}$, with F_{tot} being the total number of frames in each trial and F_{bottom} the number of frames in which no fish was detected above 5cm in the vertical coordinate).

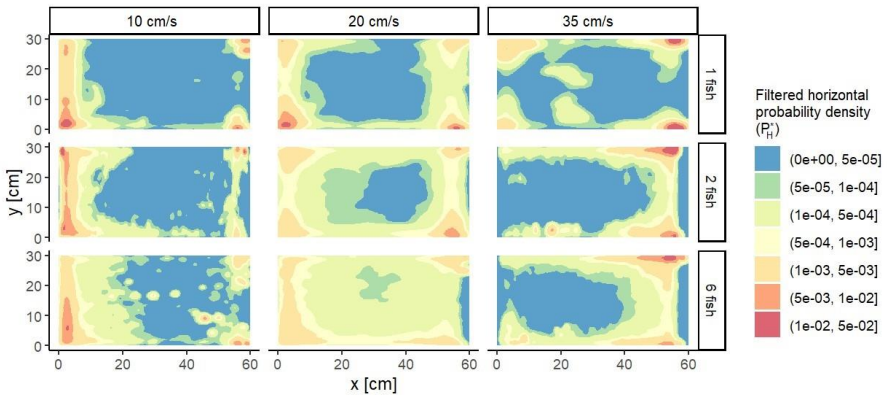

Figure 3.7: Bottom-filtered horizontal probability density (P_H*) illustrating fish distribution at various flow velocities (10, 20, and 35 cm/s) and group sizes (1, 2, and 6). Filtering was performed by disregarding frames in which at least one fish was detected above 5cm from the flume bottom. Water is flowing from left to right.

Table 3.3: Mean explored area and trajectory length for unfiltered analysis (EA, and TL) and for bottom-filtered analysis (EA*, and TL*), and median values of longitudinal and lateral position ($\tilde{P}_H(x)$ and $\tilde{P}_H(y)$ for unfiltered data, and $\tilde{P}_H^*(x)$and $\tilde{P}_H^*(y)$ from bottom filtered analysis). Filtering was performed by removing frames in which at least one fish was detected above 5 cm in the water column.

	Mean explored area (%)		Mean trajectory length [m]		Median longitudinal position [cm]		Median lateral position [cm]	
	$\overline{EA}$	$\overline{EA^*}$	$\overline{TL}$	$\overline{TL^*}$	$\tilde{P}_H(x)$	$\tilde{P}_H^*(x)$	$\tilde{P}_H(y)$	$\tilde{P}_H^*(y)$
1 fish, 10 cm/s	9.9	9.2	8.3	9.8	2.57	2.67	9.75	9.19
1 fish, 20 cm/s	14.0	13.6	9.4	10.2	4.55	3.87	5.47	4.83
1 fish, 35 cm/s	8.8	8.8	6.1	6.5	55.3	55.3	23.7	23.7
2 fish, 10 cm/s	20.1	15.6	12.0	12.4	3.72	3.54	15.6	14.5
2 fish, 20 cm/s	27.1	26.4	14.4	14.5	35.0	33.6	8.88	8.19
2 fish, 35 cm/s	17.6	17.5	11.2	12.2	52.3	52.4	28.0	27.8
6 fish, 10 cm/s	19.1	11.7	11.1	10.4	4.09	3.82	9.08	9.08
6 fish, 20 cm/s	38.3	30.8	16.5	16.3	16.9	18.9	11.4	10.8
6 fish, 35 cm/s	26.1	24.0	14.7	14.6	49.6	49.8	26.7	26.7

3.4 Discussion

Overall, the results demonstrate that individuals within groups exhibited higher exploratory behaviour (i.e. higher EA) and swimming activity (i.e. higher TL) compared to fish swimming alone and that these increments were found to be affected by flow velocity. Preferred swimming areas - denoted by horizontal and vertical space use (P_H and P_V, respectively) - were mainly governed by flow velocity and not that much by group size. We observed that fish spontaneously swam in the near-bottom region, and the horizontal bottom-filtered analysis revealed no discernible differences in the primary trends of EA, TL, and P_H when compared to the unfiltered data (Figure 3.4 and Figure 3.7). Thus, we could reasonably assume that the estimation of EA, TL and P_H through a

two-dimensional reconstruction of trajectories was highly representative of fish behaviour within the entire volume.

In order to make them feasible, treatments were conducted consecutively with an increase in flow velocity within a trial. This carries a potential time dependency, where higher velocities always occur after a longer time in the flume. Even though it cannot be definitely established, there are good reasons to believe that the observed effect of velocity is real and not a hidden effect of time in the flume. Previous studies on guppies (*Poecilia reticulata*) and mosquitofish (*Gambusia holbrooki*), revealed that the influence of time on exploratory behaviour becomes significant after three [176] and four hours [100], respectively. Fish are also unlikely to fatigue under the time and velocities of the experiment. Most of similar-sized *Telestes muticellus*, were able to swim for longer than 30 min at 35m/s in a dedicated (single fish) fatigue test [177]. Given this and the relatively short duration of our experimental trial, it is reasonable to infer that flow velocity exerted a more pronounced influence within the studied timeframe compared to the role of time itself.

Horizontal density distribution showed that locations of high fish density areas were unaffected by group size but were significantly influenced by velocity (Table 3.2). As flow velocity increased, fish moved towards the downstream grid and near-wall regions (Figure 3.4). On the vertical plane, they positioned themselves closer to the flume bottom as velocity increased, regardless of group size (Figure 3.5). It is interesting to note that flow velocity played a role in its horizontal and vertical positioning (P_H and P_V, respectively). These space use patterns are likely driven by the need to seek less challenging flow conditions, as local velocities near the walls or bottom boundary layer are typically lower than in central regions or the intermediate layers of the water column [178,179].

The increase of exploration and swimming activity with increasing group size is a form of social facilitation, intended as the promotion and enhancement of exploration and swimming activity in the presence of conspecifics [100,101,157,158]. This type of social facilitation is reported in the literature for other species in standing water, including minnows (*Phoxinus phoxinus*) [157], goldfish (*Carassius auratus*) [157], perch (*Perca fluviatilis*) [101], three-spine sticklebacks (*Gasterosteus aculeatus*) [158], and juvenile mosquitofish (*Gambusia holbrooki*) [100]. In the present study, we sought to expand upon the current understanding by investigating social facilitation in running waters. Interestingly, we found that the effect of group size on EA and TL seemed to vary with flow velocity, with minimal discernibility at the lowest velocity and becoming more prominent at higher velocities (Figure 3.3 and Table 3.1). But why should flow velocity affect social facilitation? We propose three hypotheses explaining the observed behaviour: competition, stress reduction, and energy saving from coordinated swimming.

Competitive displacement may emerge among individuals seeking energetically favourable areas, typically characterised by lower flow velocities (e.g., near the flume walls [179]). Fish swimming alone tended to linger in these areas, resulting in reduced variability in their positional patterns compared to when swimming in groups (Figure 3.4). Swimming with conspecifics introduces the potential for competition over these desirable locations, which may lead to greater spatial variability and increased exploration by each individual as they interchangingly displace each other while seeking the best spots. Flow velocity could mediate the intensity of such competition by exacerbating the necessity for seeking refuge in these areas at higher velocities. However, if competition were the sole factor at play, the impact on EA and TL would be less prominent. When examining medium and high velocities, we observe a substantial increase in average EA and TL, nearly doubling in two-fish groups compared to single-fish groups (Figure 3.3). Given the near symmetry of the flume and flow-field conditions, even if both fish were to closely compete for a single favourable area and displace each other, it is unlikely that the overall covered area and trajectory would double as there would be plenty of other very favourable spots available nearby. Hence, it is improbable that competition alone accounted for the observed increase in EA and TL with group size at medium and high velocities.

The second hypothesis relates to the calming effect of social facilitation, which reduces stress levels in fish when in the presence of conspecifics, as reported in the literature [77]. Results from physiological analysis indicated that fish swimming in group exhibited lower cortisol and oxidative stress levels in muscle tissue compared to solitary counterparts [164]. It plausible that under low flow conditions, fish may experience reduced stress, thereby diminishing the significance of the presence of others (i.e. group effects such as social facilitation). Conversely, higher velocities might induce greater stress in fish, amplifying the impact of conspecific presence. Studies on rainbow trout (*Oncorhynchus mykiss*) and turbot (*Scophthalmus maximus*) have shown a relationship between stress and flow velocity, with increasing stress responses and oxidative damage observed for increasing flow velocities [180,181].

The third explanation centres around the energetic benefits of swimming in a group [182]. Qualitative observations indicated a common tendency for fish to swim and school together, with some exceptions like solitary individuals near the downstream grid. Coordinated swimming (schooling) allows fish to take advantage of hydrodynamic features, resulting in energy conservation strategies [41–44,183,184]. As fish save energy from locomotion, they may allocate these reserves towards exploring their environment and exhibiting increased movement within the available arena when in groups, especially under challenging hydrodynamic conditions, namely at 20 and 35 cm/s.

To support this hypothesis, we assess the level of hydrodynamic challenge faced by the fish in our experiments by comparing the employed flow velocities with the so-called *Maximum Sustained Velocity* (U_{ms}) as estimated using the formula suggested by Videler [185]. U_{ms} identifies the crossover between the *sustained* and *prolonged swimming range*. The former indicates a range of velocities (lower than U_{ms}) that fish can maintain for more than 200 mins, using only aerobic energy reserves [186]. The latter instead identifies a range of velocities (higher than U_{ms}) that are much more energetically demanding as they involve the use of anaerobic energy reserves, which eventually run out, hence leading fish to fatigue [186]. Note that the reliability of Videler's formula to estimate U_{ms} is generally low as it does not account for effects of important factors such as e.g. temperature and species [187]. However, it can be used to provide some reference values that are valid for discussion purposes. In our case (i.e. fish with a body length of 5 cm), the Videler's equation estimates $U_{ms} = 27$ cm/s. This implies that the lowest velocity utilised in our experiments (10 cm/s) was probably well within the sustained range and hence not energetically challenging for the fish. In contrast, the medium (20 cm/s) and highest investigated velocities (35 cm/s) are either close to or greater than the estimated U_{ms} meaning they could belong to the more challenging prolonged range Considering that the fish swam at the highest velocity for 10 minutes, it is reasonable to infer that they were far from fatigue, given that U_{ms} provides an order of magnitude for fish swimming continuously for 200 minutes (note that boundaries between different ranges are far from being sharp).

These considerations support the hypothesis that energetics might have played a role in modulating the effect of flow velocity on social facilitation observed in our study. Nevertheless, it remains challenging to discern whether the effects of flow velocity on social facilitation are attributable to a single mechanism or a combination of the three herein identified. Furthermore, certain individuals and groups may be primarily influenced by one mechanism, while others may experience a greater impact from another due to the substantial intraspecific variability observed in wild animals [188]. Nonetheless, these findings fulfil a dual purpose by highlighting the critical role of flow velocity in behavioural research on rheophilic species [189–192] and by emphasising the importance of integrating group behaviour in the field of ecohydraulics.

Overall, these findings have practical applications for conservation management purposes. Despite the extensive construction of fish passages to mitigate freshwater migratory fish decline, numerous structures suffer from low passage efficiency [5,6]. An effective exploratory behaviour is crucial for fish to search and locate entrances of fishways at river barriers [5,6,30], and while extensive research has been done to assess and enhance the attractiveness of such entrances for individual fish [5,162,193], very little was known about the case of shoaling fish [67,108]. Furthermore, gaining insights into the spatial utilisation of shoals, both horizontally and vertically, in confined environments can

provide relevant evidence for the efficient design of fish passages, resulting in optimised configurations that align with fish spatial preferences, particularly under challenging flow conditions [194].

This Chapter offers insights into the interplay between flow velocity and social facilitation on fish exploration and swimming activity. We found that social facilitation occurs even with as few as two fish, but its effects vary depending on flow velocity. This means that caution should be exerted when interpreting findings from experiments conducted in aquariums and tanks with stagnant waters, as they may not entirely capture the effects of social facilitation on fish behaviour, particularly for rheophilic species. Since river habitats often exhibit diverse flow conditions, neglecting hydrodynamics could lead to unrepresentative or biased outcomes. Conversely, a comprehensive understanding of collective behaviour becomes imperative when investigating the behaviour of gregarious species inhabiting flowing waters. In essence, this study underscores the necessity of accounting for the intricate relationship between hydrodynamics and social interactions to gain a more holistic perspective on fish behaviour.

Chapter 4. Flow velocity and boundary effects on fish interactions

Abstract

The origin of coordinated collective motion in fish has been a longstanding inquiry among researchers. In the realm of freshwater migratory fish, understanding the intricacies of conspecific interactions holds significant implications for conservation efforts, especially as populations face steep declines due to river fragmentation. To address this, a thorough comprehension of environmental factors is essential. Previous studies examining the influence of flow velocity on shoaling tendencies have produced conflicting results, and there is a noticeable data scarcity regarding behaviours at relatively high velocities (i.e. investigated velocities were of the order of about four body lengths per second at most), which correspond to conditions that may be encountered during migration through fish passages. The impact of boundaries, such as riverbanks and concrete walls in fish passages, on fish interaction in flowing waters remains largely unexplored despite the recognition that certain species migrate along riverbanks to conserve energy. In this Chapter, we aim to address gaps in the current literature by scrutinizing the interaction dynamics of pairs of *Telestes muticellus* across three distinct flow velocities and in different regions of a confined open channel flume. Specifically, we investigated the influence of velocity and boundaries on pair interactions, examining both longitudinal and lateral motions across different spatial patterns. Furthermore, we quantified the response time for the longitudinal and lateral components at the three different flow velocities.

4.1 Introduction

Coordinated collective motion is a fascinating phenomenon observed ubiquitously in the animal kingdom, from formations of flocking birds to the intricate patterns of swarming ants and schooling fish [8]. These coordinated patterns arise from individual social interactions, forming the basis for these intriguing behavioural phenomena. In the case of migratory freshwater fishes, the investigation of these interactions finds important applications. Since the 1970s, these species have witnessed a global population decline of 76%, with river fragmentation identified as the primary cause [14]. As previously explained in Chapter 1, many fish species migrate in coordinated groups [65], and understanding the social dynamics driving their movements can contribute to improving the planning of conservation strategies and fish passage design [195].

Throughout the years, multiple approaches have attempted to elucidate the principles governing school formation and coordination, including hydrodynamic [45,50], particle physics [196], and data-driven approaches [36,197], each offering unique insights into the mechanisms driving coordinated swimming behaviour. However, most of these studies were carried out considering standing water conditions. This represents a strong limitation because, while migrating, fish are most often exposed to complex hydrodynamic features and the current understanding of how individuals process and combine social and hydrodynamic information is relatively poor [198].

Prior studies have yielded conflicting results regarding the effect of flow velocity on shoaling tendencies among various species. Guppies (*Poecilia reticulata*) displayed reduced shoaling time with increasing flow velocities [199], whereas wild zebrafish (*Danio rerio*), chubs (*Leuciscus cephalus*), and minnows (*Phoxinus phoxinus*) showed increased shoaling tendency with increasing flow velocities [51,200,201]. Notably, these studies have consistently defined shoaling as fish being less than four Body Lengths (BL) apart, except Shelton et al. (2020) [200], who employed five BL, and Allouche & Gaudin (2001) [201], where the metric used to define shoaling was not reported. Although a threshold of four BL is widely accepted in the literature to characterise fish shoaling in standing waters [33,202,203], its *a priori* applicability to running waters is questionable. In general, the effect of flow velocity on the so-called decorrelation distance between two fish (i.e. the distance beyond which the two fish stop interacting and hence the distance that can be taken as a threshold to define whether two fish are shoaling or not) has never been investigated.

The impact of flow velocity on swimming formation has also been explored, specifically for red nose tetra fish (*Hemigrammus bleheri*), zebrafish (*Danio rerio*), and minnows (*Phoxinus phoxinus*) [49,51,52,204]. Across these studies, a preference for a side-by-side configuration granting energetic benefits over high flows was consistently observed. Nevertheless, it is worth noting that the flow conditions examined in these studies ranged

between 0 and 2BL, with an exception reaching 4.3 BL/s [49,52]. These intervals are somewhat limited, and there is little knowledge about behaviours at higher velocities. This is particularly relevant for investigating scenarios in which fish encounter hydrodynamically challenging environments, such as fish passages [205,206].

Experimental observations on the effect of flow velocity on fish interaction in shoals are also quite limited. Chicoli et al. (2014) found that mutual responses among individuals of giant danio (*Devario aequipinnatus*) were more pronounced in still water compared to flowing conditions [39]. They suggested that this might be attributed to a weaker signal sensed by the lateral line as masked by ambient flow, making the transmission of hydrodynamic information from neighbouring fish more challenging. De Bie et al. (2020) calculated information transfer among pairs of minnows (*Phoxinus phoxinus*) in three different conditions – one with no flow, one with low flow (5.5 cm/s, 1 BL/s), and one with high flow (11 cm/s, 2 BL/s) – but no evident differences between treatments were discussed for this metric [51]. To measure information transfer, they computed cross-correlation functions of fish ground velocity, discovering that the rear fish was much more likely to react to the front one than the other way around, which suggests information transfer flowed predominantly from front to back. Additionally, they identified peaks in velocity cross-correlation functions to determine the timescale over which fish responded to each other, finding it to range between 0.3 and 0.5 seconds with no reported differences between treatments.

Ultimately, nearly all previous works either simulated or observed shoaling behaviour in open waters, neglecting the potential impact of boundaries such as side walls. Several species are known to migrate along the river banks, including salmon (*Oncorhynchus nerka*), and barbel (*Luciobarbus comizo, L. microcephalus*) [207–209]. This behaviour is primarily induced by the tendency of fish to avoid excessive velocities and upwellings in the main channel to save energy for migration [210]. Collective behaviour is often modelled as a multi-body dynamical problem whereby attractive and repulsive forces between agents dictate the social dynamics of the group. The effects of boundaries on group behaviour are often modelled as attraction forces. For instance, in a study performed in still water tanks, Herbert-Read et al. (2011) measured the attraction force exerted by lateral walls, finding it to be additive to the one exerted by group mates [211]. This phenomenon is termed thigmotaxis (or "wall-hugging") and can be attributed to excessive stress [212]. In the case of running waters, the attraction force of lateral walls could also represent an energy-saving strategy, as near-wall regions are characterised by reduced flow velocities where fish spend less energy in locomotion, a concept often referred to as wall holding [179,211]. To our knowledge, no study has investigated the effect of boundaries on fish interaction in running waters.

The aim of this work is to investigate interactions of fish exposed to flow velocities ranging from 2 to 7 BL/s throughout a confined swimming arena and hence accounting for boundary effects. Adhering to the reductionist approach [33,49,51], we analyse the trajectories of fish pairs within three flow conditions in four distinct regions of the flume. This deliberate choice aligns with the requirement to measure fundamental interaction metrics, focusing on the simplest subsystem of a shoal. Our objective is to characterise interactions in terms of (i) shoaling time, (ii) mutual relative positioning, and (iii) cross-correlation functions. With the latter, we aim to quantify interactions across varying relative distances and orientations between fish, considering different boundary regions and estimating response times. These metrics offer insights into the spatial scale at which interactions are significant and the speed and efficiency of information transfer within a shoal, thereby impacting their decision-making processes [213].

4.2 Materials and methods

4.2.1 Animals, experimental protocol, trajectory data

Italian riffle dace (*Telestes muticellus*) is a social fish species native to the streams and rivers of the Italian peninsula, which, despite its abundance, has been relatively understudied [162]. This study involved a cohort of 20 wild juvenile Italian riffle dace, with an average length of 5.14 cm (SD ± 0.34 cm). Experiments were conducted in May 2021 within an open-channel flume featuring transparent walls. The flume was equipped with a combined pump-inverter system, enabling precise flow rate control and a downstream weir of adjustable height, allowing for control of water depth. The experiment protocol entailed introducing fish pairs into the flume with a bulk flow velocity (U_B estimated as flow rate per unit wet area) of 10 cm/s (U_{B10}). This flow velocity was kept constant for 15 minutes, with the duration divided into an initial 5-minute habituation period followed by a 10-minute trial period. Subsequently, the flow velocity was incrementally raised to 20 cm/s (U_{B20}) and maintained at this level for an additional 10 minutes. Finally, pairs were exposed to the highest flow velocity of 35 cm/s (U_{B35}) for 10 minutes. Water depth was constantly maintained at 15 cm throughout the entire experiment. Comprehensive details regarding the animals and the experimental setup can be found in Chapter 3. The analysis of vertical positioning, as presented in Chapter 3, indicated that fish predominantly remained within the first 5 cm from the flume bottom. This behaviour allowed us to approximate their trajectory as two-dimensional within the horizontal plane.

4.2.2 Overview of data analysis

Video recordings were analysed only in the last 7 minutes of each trial velocity. This was done to avoid potential disturbances in fish behaviour due to the velocity transient. The fish's positional data (expressed in cm, with a precision of ± 3mm and a sampling rate of 50 frames per second) were acquired using an AI-based approach that was employed to analyse video recordings captured from the camera positioned below the channel, as detailed in Chapter 2.

The reference system was defined with its origin on the head of the Leading fish (L). The x-axis was aligned with the flow direction, while the y-axis was oriented transversely to the water flow within the horizontal plane (Figure 4.2). Data were analysed to compute shoaling times, relative positioning of fish and cross-correlation functions as described in Sections 4.2.3, 4.2.4, and 4.2.5. Cross-correlation functions were utilised to quantify fish interactions and estimate decorrelation distances (D_d) and response times (t_r), (i.e. the time after which a fish, on average, responds to the movement of the other fish). The swimming arena was divided into four swimming regions (Back, Front, Centre and Side,

Figure 4.1) in order to assess the effect of boundaries on the metrics listed above (definition of the swimming region is detailed in Section 4.2.4).

Data analysis was carried out using R version 4.0.5 [166]. Plots were created using the *ggplot2* package (version 3.4.0, [168]), and statistical analysis was performed using the *rstatix* package (version 0.7.0, [169]) in R, with a significance threshold set at $p < 0.05$.

4.2.3 Shoaling time and relative position

The shoaling time was computed as the fraction of time for which fish stayed within a threshold distance that, in order to ensure the robustness of the results, was varied as 2, 4, 6, and 8 BL. Shoaling time data exhibited did not fulfil the assumptions for parametric tests (Shapiro-Wilk test with p-value < 0.05). Hence, the Friedman test was employed to explore the influence of flow velocity (U_B) on shoaling times, considering U_B as a within-subject factor to account for repeated measures across different flow velocities [173]. Plots of the relative position in the swimming arena were generated using the *geom_density_2d* command in the *ggplot2* v.3.4.0 35 [168] package in R.

4.2.4 Swimming regions

The arena was partitioned into four distinct swimming regions (SR): Back, Front, Centre, and Side (Figure 4.1). The Front and Back regions were delineated as areas within a 5 cm (equivalent to 1 BL) distance from the upstream and downstream grids, respectively. The Side region was defined as the lateral area located within 2.5 cm (equivalent to 0.5 BL) from the side walls but remained at least 5 cm away from the upstream or downstream grids (already designated as part of the Front and Back regions). The Centre region was subsequently identified as the central area not included in the other defined regions. These values were selected based on observations from the fish position distribution analysis previously described in Chapter 3. More specifically, the median distance from lateral walls at U_{B35} was found to be below 0.5 BL, while the median distance from the upstream or downstream at U_{B10} was below 1 BL.

After defining the swimming regions (SR), we calculated the time fish spent in each region for each experiment (t_{SR}, expressed in s). To obtain a representation devoid of bias arising from the varied area extensions of SR, we also computed the fraction of time fish spent in each SR, normalised for the area extension of each SR (t^*_{SR}). We employed the Friedman test to investigate the impact of SR and U_B on both t_{SR} and t^*_{SR} [173]. The Friedman test was also utilised to assess the effect of SR on correlation ($c_{u,n}(SR)$ and $c_{v,n}(SR)$, refer to Section 4.2.5 for details).

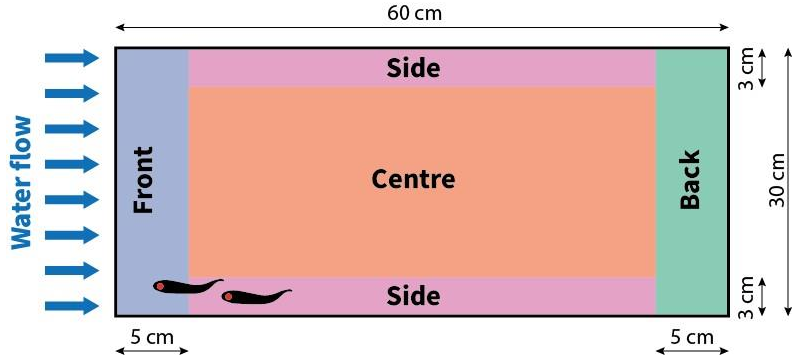

Figure 4.1: Schematic representation of the bottom-view of the test section , along with the delimitation of Swimming Regions (SR: Front, Centre, Side and Back). Water flows from left to right.

4.2.5 Cross-correlation functions

To quantify the interaction between fish pairs, we employed cross-correlation functions of fish velocity components along the longitudinal (u) and lateral coordinate (v) [33,214]. The fish were identified as the Lead (L) and the Rear (R), depending on their longitudinal coordinate, and the mean correlation was estimated from the ensemble average of the N experiments:

$$C_u(\vec{r},\tau) \overset{\text{def}}{=} \frac{1}{N}\sum_{n=1}^{N} c_{u,n}(\vec{r},\tau) \quad (N = 10)$$

Where τ is a time lag, $\vec{r} = \begin{pmatrix} d \\ \theta \end{pmatrix}$ represents the vector distance between the two fish, whereby d is the absolute distance between the fish and θ is the angle between them, ranging from 0 to 90°, defined as 0° for fish staying side-by-side and 90° when in an in-line configuration (Figure 4.2).

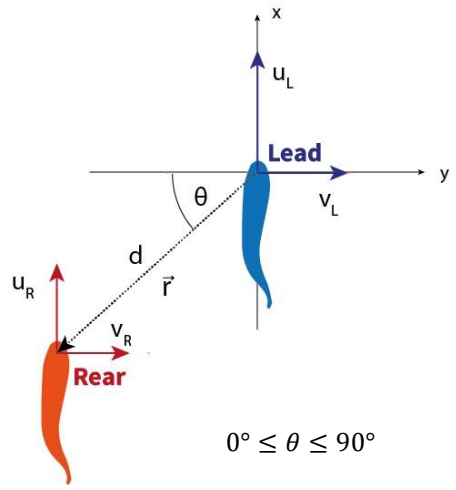

Figure 4.2: Coordinate reference system

For each time lag τ and each n^{th} experiment, $c_{u,n}(\vec{r},\tau)$ was estimated as a Pearson correlation coefficient, a widely used statistic used to quantify the linear relationship or degree of association between two variables [215]. The cross-correlation function between the longitudinal velocity of the Lead fish ($u_{L,n}$) and the Rear fish ($u_{R,n}$) is therefore computed as:

$$c_{u,n}(\vec{r},\tau) = \frac{\sum_t \left(u_{L,n}(t) - \overline{u_{L,n}}\right)\left(u_{R,n}(\vec{r},t+\tau) - \overline{u_{R,n}}(\vec{r})\right)}{\sqrt{\sum_t \left(u_{L,n}(t) - \overline{u_{L,n}}\right)^2}\sqrt{\sum_t \left(u_{R,n}(\vec{r},t+\tau) - \overline{u_{R,n}}(\vec{r})\right)^2}}$$

Where t is time and the overbar symbol refers to time averaging.

To determine the spatial scale of fish interactions and quantify the decorrelation distance D_d, cross-correlation functions $c_{u,n}(\vec{r},\tau)$ with zero-time lag (τ =0) were calculated at chosen relative distances between fish (i.e. at chosen distances d and regardless of θ). Specifically, data were discretised into two body length (BL) increments across the range of 0 to 14 BL, and the mean correlation $C_u(d)$ was computed:

$$C_u(d) = \frac{1}{N}\sum_{n=1}^{N} c_{u,n}(2(i-1) < d < 2i, \tau = 0) \qquad i = [1 \dots 7]$$

A correlation threshold was set as 0.05, designating any correlation values below this threshold as statistically insignificant. This threshold was selected based on a correlation analysis between the velocity time series of two unrelated fish associated with different randomly selected experiments, revealing that 95% of random correlations fell

below 0.05. Employing a conservative approach, the decorrelation distance D_d was defined as the distance beyond which the mean correlation coefficient remained consistently below 0.05 for each velocity component and U_B (i.e., $C_u(d > D_d) < 0.05$ and $C_v(d > D_d) < 0.05$). For all analyses further described, this decorrelation distance was employed as a threshold to filter out the time intervals in which fish were too far apart to interact.

The effect of relative orientation on fish interactions was assessed by computing zero-lag correlation functions $c_{u,n}(\vec{r}, \tau = 0)$ at time intervals when the fish were positioned at specific intervals of θ, and a relative distance below the decorrelation distance D_d:

$$C_u(\theta) = \frac{1}{N}\sum_{n=1}^{N} c_{u,n}(\pi/12\,(i-1) < \theta < \pi/12\,i, d < D_d, \tau = 0) \qquad i = [1 \dots 6]$$

In order to examine the effect of the swimming regions (SR), N cross-correlation functions were computed for each SR_i, at a zero-time lag and a relative distance below the decorrelation distance D_d:

$$c_{u,n}(SR) = c_{u,n}(SR_i, d < D_d, \tau = 0) \qquad SR_i \in [\text{Back}, \text{Centre}, \text{Front}, \text{Side}]$$

Cross-correlation functions of fish velocities were computed across different time lags ($\tau = [-5 \dots 5]$ s) for any $\vec{r}$, where the relative distance (d) was less than the decorrelation distance D_d. This analysis aimed to quantify response times, defined as the τ values corresponding to peaks in $c_{u,n}(\tau, d < D_d)$ [33]. The average response time (t_r) was then calculated as:

$$t_{r,u} = \frac{1}{N}\sum_{n=1}^{N} \tau_{MAX,u,n}$$

$$\tau_{MAX,u,n} : \; c_{u,n}(\tau_{MAX}) = \max_{\tau}(c_{u,n}(\tau))$$

Data pertaining to correlation functions and response times did not fulfil the assumptions for patametric tests (Shapiro-Wilk test with p-value < 0.05). Consequently, the Friedman test was employed to assess the impact of flow velocity as a within-subject factor, accounting for repeated measures across varying U_B [173].

4.3 Results

4.3.1 Position and shoaling time

Shoaling time is found to be statistically unaffected by U_B, irrespective of the threshold distance applied to consider fish shoaling (Friedman test, p-value > 0.05 for 2, 4, 6, and 8 BL). Specifically, for a threshold distance of 4 BL, average shoaling times (mean ± standard deviation) constitute 79% ± 35%, 85% ± 31%, and 83% ± 30% of the total time in the flume for U_{B10}, U_{B20}, and U_{B35} respectively.

Analysis of the relative position within the entire flume region reveals distinct patterns: at U_{B10} and U_{B20}, fish tended to adopt a side-by-side configuration, while at U_{B35}, an in-line configuration was prevalent (Figure 4.3). However, examination of relative positions within different swimming regions (Figure 4.4) indicates that the relative orientation is dictated by a combined effect of the specific region where fish were swimming and velocity. In particular, the relative positioning of fish swimming in the Front and Centre region is evidently affected by U_B: at U_{B10} and U_{B20}, the predominant configuration was side-by-side, while at the highest flow velocity, pairs also started to adopt an in-line configuration. Relative positioning in the Back region is also influenced by U_B, albeit to a lesser extent.

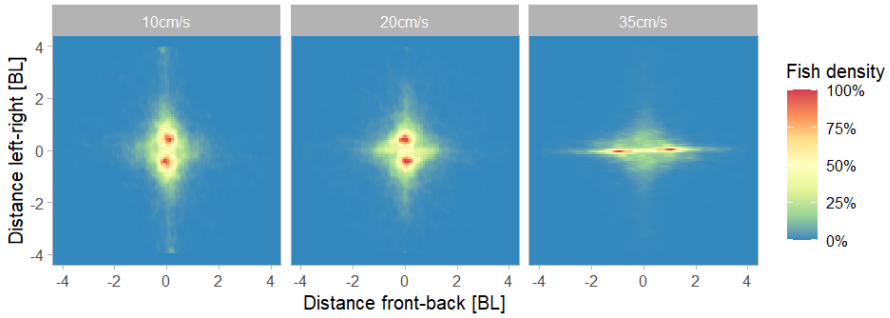

Figure 4.3: Relative position of fish pairs for different flow velocities U_{B10}, U_{B20}, and U_{B35} (10, 20 and 35 cm/s of bulk velocity, respectively). Water is flowing from left to right.

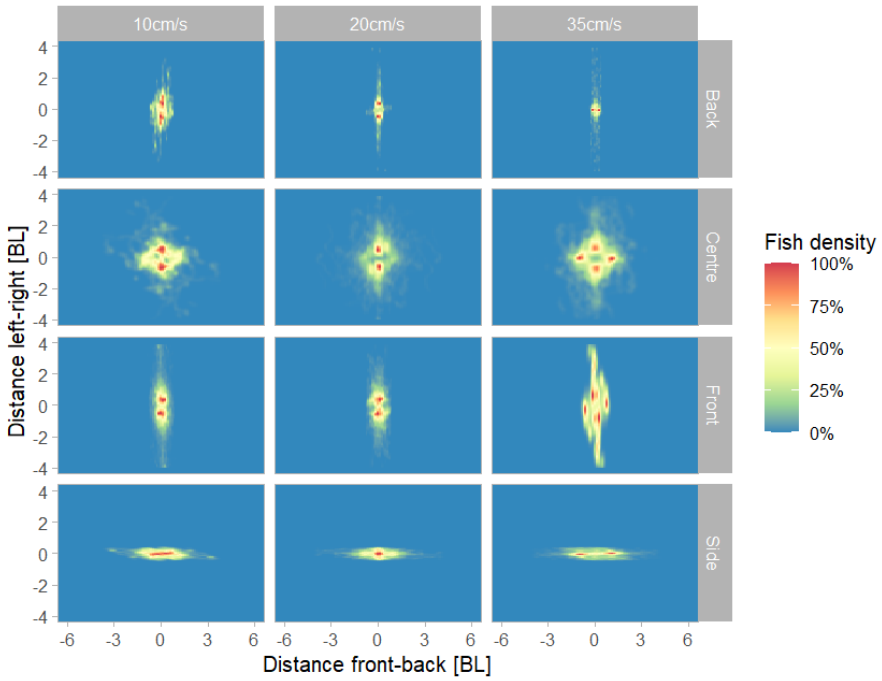

Figure 4.4: Relative positions of fish pairs for different swimming regions and different flow velocities velocities U_{B10}, U_{B20}, and U_{B35} (10, 20 and 35 cm/s of bulk velocity, respectively). Front and Back regions are defined within 5 cm (1 BL) of upstream and downstream grids, respectively. Areas within 2.5 cm (0.5 BL) from lateral walls are defined as Side (remaining at least 5 cm away from grids), and the Centre region is the central area not covered by other defined regions. Water is flowing from left to right

The analysis of the time spent by fish in each region, denoted as t_{SR} (Figure 4.5, left), reveals that at U_{B10}, fish predominantly occupied the Front region. At U_{B20}, the distribution was more even, while at U_{B35}, fish primarily swam next to the lateral walls, specifically in the Side region. These tendencies are confirmed by the normalised fraction of time spent in each region, denoted as t^*_{SR} (Figure 4.5, right), which accounts for the extension of each SR. The only notable difference between the figures is the fraction of time spent in the Centre region, as it has the most extension, indicating that fish tended to swim near boundaries.

Statistical analyses of both t_{SR} and t^*_{SR} confirm that at U_{B10} and U_{B35}, fish tended to swim in specific regions (Table 4.1), specifically in the Front and Side regions (Figure 4.5). Conversely, the effect was not statistically significant at U_{B20} (Table 4.1). The analyses also indicate that the U_B significantly influenced the utilisation of the Front, and Side regions by fish.

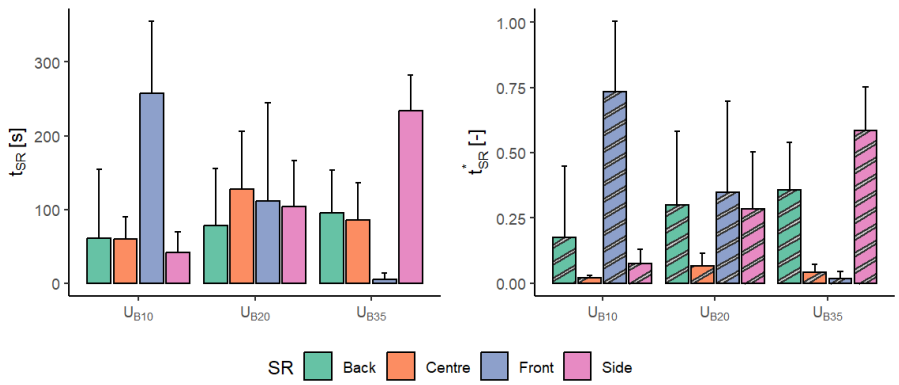

Figure 4.5: Swimming region usage as time spent by fish swimming in each region t_{SR} (a) and fraction of time normalised for the extension of the SR areas t^*_{SR} (b) for different flow velocities (U_{B10} = 10 cm/s, U_{B20}= 20cm/s, and U_{B35}= 35cm/s). Front and Back regions are defined within 5 cm (1 BL) of upstream and downstream grids, respectively. Areas within 2.5 cm (0.5 BL) from lateral walls are defined as Side (remaining at least 5 cm away from grids), and the Centre region is the central area not covered by other defined regions. N experiments = 10.

Table 4.1: Effect of swimming region (SR) and flow velocity (U_B) on fish swimming region usage. Namely, time spent by fish swimming in each region (t_{SR}) and the fraction of time normalised for the extension of the SR areas (t^*_{SR}). The analysis is conducted with the Friedman test (df=2 for all tests). NS indicates non-significant (p-value ≥ 0.05).

Factor	*Cluster*	t_{SR} *(p-values)*	t^*_{SR} *(p-values)*
SR	U_{B10} (10 cm/s)	0.003	< 0.001
	U_{B20} (20 cm/s)	NS	NS
	U_{B35} (35 cm/s)	< 0.001	< 0.001
U_B	Back	NS	NS
	Centre	0.045	NS
	Front	0.003	0.002
	Side	< 0.001	< 0.001

4.3.2 Relative distance and orientation effects on fish interaction

The mean longitudinal cross-correlation function computed over different relative distances, $C_u(d)$ displays positive values, albeit lower than 0.2 (Figure 4.6, top left). Such values drop below 0.05 when the distance between fish exceeds 5 BL for U_{B10} and U_{B35}, while for U_{B20}, they drop beneath this value for d> 6 BL. For the lateral velocity component, $C_v(d)$ drops below the threshold value at 3 BL for U_{B35}, 4 BL for U_{B10}, and 6 BL for U_{B20} (Figure 4.6, top right). For the following analysis, we conservatively used 6 BL as the distance of decorrelation D_d.

The effects of relative orientation on fish interaction are captured by $C_u(\vartheta)$ and $C_v(\vartheta)$ (Figure 4.6, bottom panels). $C_u(\vartheta)$ exhibits a trend at U_{B10}, displaying an increasing correlation with increasing ϑ, reaching above 0.25 when fish displayed an in-line configuration. A similar but less pronounced behaviour is observed at 20 cm/s. However, the correlation exhibits no discernible trend at U_{B35}, remaining consistently below 0.15.

Conversely, $C_v(\vartheta)$ displays opposite trends at U_{B10} and U_{B20}, reaching the largest values at θ=0° (side-by-side configuration) and registering correlations of 0.3 and 0.25 for U_{B20} and U_{B10}, respectively. These correlations gradually decrease, falling below 0.15 at higher angles (i.e. approaching an in-line configuration). In the case of U_{B35}, a similar trend is observed, albeit generally lower, with the maximum correlation staying below 0.2 for θ =0°.

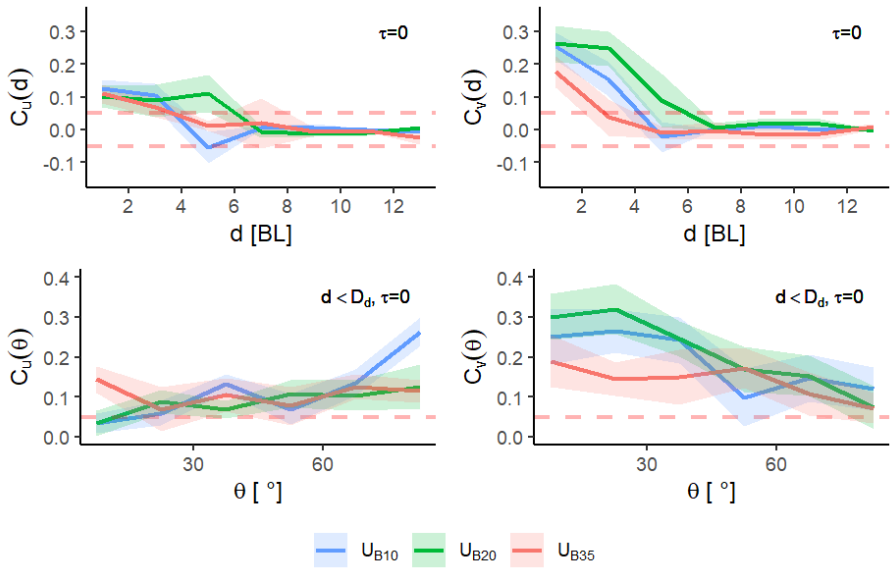

Figure 4.6: Mean zero-time lag correlation between fish velocities for the three flow velocities U_B (U_{B10} = 10 cm/s in blue, U_{B20}= 20cm/s in green and U_{B35}= 35cm/s in red). Top panels are for correlation for different distances between fish on the longitudinal ($C_u(d)$, panel a), and lateral ($C_v(d)$, panel b). Bottom panels represent correlation for different orientations ($C_u(\vartheta)$ and $C_v(\vartheta)$, panels c and d, respectively). θ=0° corresponds to a side-by-side configuration, while θ=90° stands for an in-line arrangement. The coloured bands represent the standard deviation (N=10 experiments).

4.3.3 Boundary effects on interactions

To assess whether boundary effects influenced interactions between pairs, we calculated N zero-time lag correlation functions on the longitudinal ($c_u(SR)$) and lateral ($c_v(SR)$) velocity components for each swimming region (SR, Figure 6). Longitudinal motion exhibits significant correlation (i.e., with a median higher than 0.05) solely in the Centre and Side regions (Figure 4.7a), although at U_{B20}, a significant correlation is observed only in the Side regions. Conversely, for lateral motion (Figure 4.7b), a significant correlation is found in the Centre and Front regions (except for tU_{B35}, at which correlation is significant only in the Centre). Statistical analysis reveals that the effect of the swimming region is significant across all U_B for both longitudinal and lateral motion (Friedman test, p-value < 0.05).

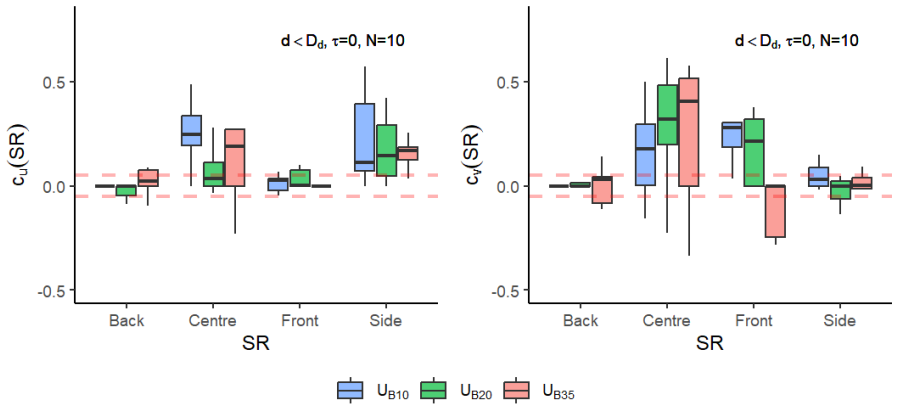

Figure 4.7: Zero-time lag correlation functions in different swimming regions (SR) on the longitudinal (a) and lateral (b) velocities for the three flow velocities U_B (U_{B10} = 10 cm/s in blue, U_{B20}= 20cm/s in green and U_{B35}= 35cm/s in red). Areas within 2.5 cm (0.5 BL) from lateral walls are defined as Side (remaining at least 5 cm away from grids), and the Centre region is the central area not covered by other defined regions. Number of experiments N=10.

4.3.4 Response time

Cross-correlation functions help to depict the response time of interaction by showing peaks in correlation at certain time lags (τ). Negative lags depict a correlation comparing the velocities of the Lead fish at the current time with the delayed velocity of the Rear fish (L responds to R). On the other hand, positive lags correspond to a delay in the Lead fish (R responds to L). $C_u(\tau)$ shows two prominent peaks, one for the positive τ and one for the negative (Figure 4.8a). The peaks occur at lags of magnitude 0.47 ± 0.31, 0.47 ± 0.44, and 0.54 ± 0.5 sec for flow velocities of U_{B10}, U_{B20}, and U_{B35}, respectively. Statistical analysis reveals that U_B did not affect such response times (Friedman test, p-value >0.05). By contrast, $C_v(\tau)$ displays only one major peak on the positive lags for all flow velocities (Figure 4.8b). In this case, the peaks are located at 0.73 ± 0.49 s lags for U_{B10}, 0.54 ± 0.29 s for U_{B20}, and 0.28 ± 0.12 s for U_{B35}. Statistical analysis shows that U_B significantly affects the response time, as estimated from C_v (Friedman χ^2= 12.8, df = 2, p-value <0.05).

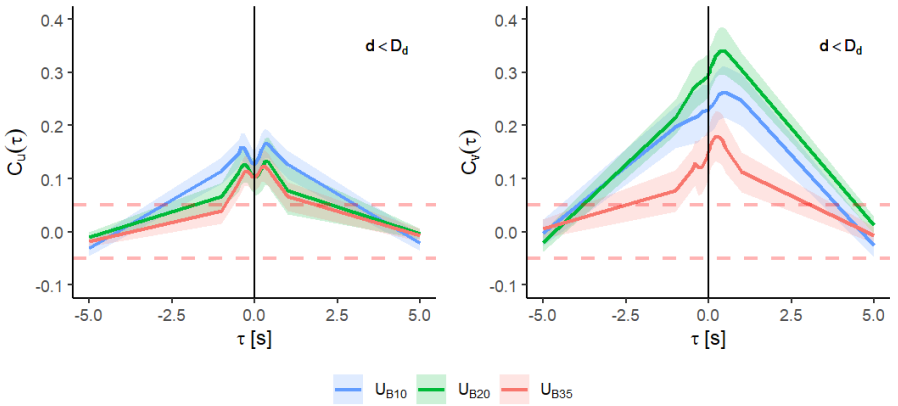

Figure 4.8: Mean cross-correlation functions of longitudinal (a) and lateral (b) velocity time series between the Leading fish and the Rear fish for the three flow velocities. Positive lags (τ) means that the Lead is delayed (Rear responds to Lead), and negative lags represent the Lead fish responding to the Rear. The coloured bands represent the standard deviation (N=10 experiments).

4.4 Discussion

The results presented in Section 4.3.1 indicate that flow velocity (U_B) has no appreciable effect on shoaling time. The available literature on this matter presents inconsistent findings, with some studies indicating either an increase [51,200,201] or a reduction in shoaling time with increasing flow velocities [199,216]. Notably, the work by De Bie et al. (2020) [51] stands out as the sole research work that specifically examined and isolated the impact of flow velocity in a controlled flume experiment, unlike the other studies that included additional variables such as habitat [200,201], predation [201] and parasitism [199]. De Bie et al. (2020) reported that pairs of minnows (*Phoxinus phoxinus*) displayed an increase in shoaling time with increasing flow velocity. However, it is crucial to note that their findings were specific to time windows associated with fish swimming in the central area of the chosen arena, meaning that data pertaining to other regions of the flume were filtered out of the analysis. This represents a significant difference that makes the work by De Bie et al. (2020) and ours (where the entire swimming arena was considered in the computation of shoaling time) hardly comparable.

It is also plausible that, in the experiments herein discussed, the lack of observable alterations in shoaling time under varying flow conditions may be attributed to shoaling being significantly induced by a need to reduce stress, which, contrary to, e.g. energy saving, might be not necessarily dependent of the hydrodynamic challenge fish are exposed to. This hypothesis is supported, to some extent, by the findings of Schumann et al. (2023), who conducted physiological analyses of the fish investigated in our experimental campaign [164]. Their results revealed a reduction in cortisol and oxidative stress levels as group size increased, indicating a calming effect on the fish induced by the presence of conspecifics [77–80].

For the highest investigated flow velocity, fish clearly tended to swim in-line in the Side region (specifically 56% of the time, Figure 4.5a), suggesting that, in such hydrodynamically challenging conditions, space use was dictated by energy-saving strategies. Near the walls, the mean velocity decreases, leading to lower drag experienced by fish and consequently higher attractiveness for this region - a phenomenon observed in studies involving single fish as well [179,184,217].

At the lowest flow velocity instead, fish preferred side-by-side positioning in the Front region, accounting for 61% of the total time (Figure 4.5a). This observation might be linked to hydrodynamic conditions created by the honeycomb grid (unexplored in this study), potentially attracting fish for energy-saving reasons. Alternatively, the tendency of pairs to position themselves immediately downstream of the honeycomb grid could be driven by behavioural factors. It is plausible that the low water flow we imposed might

resemble natural conditions that induce fish to adopt drift-foraging behaviours. Drift foraging involves fish lingering in low-velocity areas alongside faster currents, where abundant drifting invertebrates are transported [218,219]. While scientific literature on the foraging habits of *Telestes muticellus* is limited, observations of drift-foraging behaviour in other benthic cyprinids, such as *Phoxinus phoxinus* [220], lend support to this hypothesis.

At the intermediate velocity, no statistically significant differences were found in the usage of different swimming regions (Table 4.1). For these hydrodynamic conditions, fish swam uniformly around the swimming arena, and the attractive forces exerted by lateral walls or the upstream grid were insignificant. This observation aligns with previous analyses presented in Chapter 3, where peak exploration was reported for 20 cm/s. A possible explanation for this behavior is that the medium velocity represents an optimal speed for horizontal movement [221]. At the lowest velocity, fish tended to utilize the water column more, whereas at the highest velocity, they sought refuge near the lateral walls [179](see further discussion below).

Within the central region - where the influence of lateral walls and grids is negligible - a transition from predominantly side-by-side to a mixed in-line and side-by-side configuration was observed for increasing flow velocities. These results are in contrast with those of Ashraf et al. (2016) [49] who, in an experimental protocol similar to ours, report that (i) fish spontaneously tended to occupy the central region of the channel at all flow conditions, always avoiding the lateral walls and; (ii) fish progressively adopted a side-by-side configuration (referred to as the "phalanx" configuration by the authors) for increasing flow velocities. The authors hypothesised that such a side-by-side configuration minimised energy expenditure via the synchronisation of tail-beat frequency.

Contrary to point (i), our results indicate that as flow velocity increases, fish tend to avoid the central part of the channel, showing a clear tendency to swim right next to the lateral walls (Figure 4.5). Moreover, our results contradict point (ii) because, for increasing flow velocities, fish tend to deviate from a side-by-side configuration in any swimming region (Figure 4.4). It should be noted that the experiments presented by Ashraf et al. (2016) were performed under significantly different conditions, which perhaps explains such different results. Firstly, they employed a very shallow channel (i.e. the ratio between channel height and fish body length was about 0.6, whereas, in the present experiments, it was about 3). In the shallow conditions employed by Ashraf et al. (2016), the region next to the lateral walls might have provided a limited hydrodynamic advantage to fish, as its body create a significant blockage. Consequently, a non-negligible flow diversion/acceleration is generated around the fish's body, leading to increased skin friction, generation of lift forces, and ultimately, body destabilisation (similar flow-induced forces are discussed in Przybilla et al. (2010) [222]). Secondly, the

swimming arena employed by Ashraf et al. (2016) was very short (about 20 cm, i.e. 5 BL long, whereas our channel was 60 cm, i.e. 12 BL long). This limited length resulted in the development of a very thin boundary layer in the lateral region (as well as the top and bottom of the channel), making it less viable for hydrodynamic benefits.

Moreover, Ashraf et al. (2016) employed flow velocities ranging from 0.77 to 4 BL/s, whereas the range investigated herein is between 2 and 7 BL/s. Notably, at flow velocities of 10 and 20 cm/s (equivalent to 2 and 4 BL/s), we also report a tendency of fish to swim side-by-side, except for the Side region (see Figure 4.3 and Figure 4.4). However, such a tendency disappears at 35 cm/s (i.e. 7 BL/s). It is plausible that the side-by-side configuration ceases to be optimal at such fast flows, prompting fish to find more hydrodynamic benefits by swimming in proximity to the lateral walls. This holds particularly true in deep flows, as water depth can prevent excessive local blockage and allow for the development of a thicker boundary layer. Such a hypothesis would reconcile our results with those of De Bie et al. (2020), who also observed the "phalanx" configuration at their highest flow but utilised a velocity range limited to 1 and 2 BL/s.

In order to estimate the degree of hydrodynamic challenge experienced by the fish, the so-called *Maximum Sustained Velocity* (U_{ms}) can be instrumental. U_{ms} marks the transition between sustained (i.e. the regime that can be sustained indefinitely without inducing fatigue, relying solely on aerobic energy reserves) and prolonged (i.e. a regime in which the fish utilises the anaerobic energy reserves, which are finite and eventually leads to fatigue) swimming ranges. As previously detailed in Chapter 3, U_{ms} can be estimated by the Viedeler equation [185], and in our case, is U_{ms}= 27 cm/s (5.4 BL/s), which surpasses the velocities employed in previous studies. This suggests that the broader velocity range employed in our study is more likely to encompass various regimes, each presenting different challenges to the energetics of the fish. Specifically, the lowest velocity in our experiments (2 BL/s) likely fell well within the sustained range, posing no significant energetic challenge for the fish. In contrast, the medium (4 BL/s) and highest investigated velocities (7 BL/s) are either close to or exceed the estimated U_{ms}, suggesting that they may belong to the significantly-more energetically-demanding prolonged range.

In the analysis of fish interactions, we found that, up to a relative distance of 6 BL, the correlation between fish velocities remained significant (Figure 4.6 a,b), though this result is contingent upon the threshold we defined for negligible correlation (0.05). Interestingly, the maximum correlation occurs at the medium velocity (20 cm/s), coinciding with the velocity at which fish exhibited the highest exploration rate (refer to Chapter 3). Moreover, results reveal a higher correlation between fish in lateral than longitudinal motions. In lateral motions, there is a more rapid decrease in correlation at the highest flow intensity (35 cm/s, Figure 4.6b). This observation is likely attributed to

fish spatial preferences rather than hydrodynamic factors. As fish tended to stay close to the lateral walls at high flow velocities, their lateral motion was constricted, resulting in a lower correlation.

Similar conclusions can be drawn by looking at the dependence of correlation on the relative orientation (Figure 4.6 c,d). In fact, results confirm higher correlations of the longitudinal component when fish adopt an in-line configuration, typically observed in the Side region. Conversely, longitudinal correlations are lower for a side-by-side configuration, where fish tend to better coordinate through their lateral movements. The lateral coordination is presumably facilitated by communication supported by visual cues, and the side-by-side positioning ensures that both fish are in each other's field of view. In the in-line configuration, the leading fish cannot exploit such kind of communication.

In general, as fish tend to swim in specific areas based on flow conditions (Figure 4.5, Table 4.1), their spatial choices and arrangements affect the degree of interaction. At lower velocities, fish predominantly spend time side-by-side in the Front region (Figure 4.4 and Figure 4.5) and consequently interact more over lateral motion (Figure 4.6d and Figure 4.7). In contrast, in faster flows, they tend to favour an in-line formation next to the lateral walls (Figure 4.4 and Figure 4.5), therefore promoting coordination through their longitudinal motion (Figure 4.7). This phenomenon can be characterised as a "Positioning bias", suggesting that the spatial distribution of fish within the environment significantly shapes their interactions. Interestingly, at the intermediate velocity - where fish distributed themselves most evenly across the arena - they displayed the highest correlation for both velocity components (Figure 4.6 a,b), representing an optimum not only for exploration but also in terms of interaction.

The cross-correlation functions reported in Figure 4.8 provide hints about how communication between fish occurs. Our results reveal a bidirectional interaction in longitudinal motion, as depicted by the nearly symmetrical double peaks (Figure 4.8a). Unexpectedly, this bidirectional interaction persists even at the highest flow velocity (U_{B35}), where fish predominantly adopt an in-line configuration (see Figure 4.3). When fish swim in an in-line arrangement, the Rear fish is likely to respond to the Lead fish's longitudinal motion (i.e. peak in $C_u(\tau)$ for positive τ) as a result of a visual cue. However, interpreting the reverse (i.e. the peak of $C_u(\tau)$ for negative τ) is challenging, as it is unclear how the Lead fish can respond to the motion of the Rear fish, especially in faster flows when the Rear fish is mostly directly behind the other, potentially outside its field of view. The utilisation of the lateral line could be an explanation, given that cyprinids are reported to possess numerous superficial neuromasts on the tail fin [223]. These neuromasts, constituting the lateral line, serve as specialised sensory organs that enable fish to detect changes in water movement and pressure [224]. However, it is possible that at high flow velocities, the hydrodynamic footprint of the Rear fish becomes masked by

mean flow and turbulence, as suggested in previous studies [39,51]. Interestingly, video recordings revealed that in the Side region, the Rear fish occasionally collided with the Lead fish, prompting a swift longitudinal acceleration response. It is plausible that this collision-induced acceleration serves as the predominant factor behind the observed bidirectional interaction.

Conversely, lateral motion demonstrated that interaction followed a dominant direction whereby the Rear fish responded predominantly to the Lead fish ($C_v(\tau)$ in Figure 4.8b), with the response time influenced by flow velocity. This suggests heightened alertness in fish exposed to faster flows. Higher stress responses and oxidative damage at elevated flow velocities were observed for rainbow trout (*Oncorhynchus mykiss*) and turbot (*Scophthalmus maximus*) [180,181], potentially leading to heightened alertness [225]. It would be interesting to investigate further the impact of flow velocity on response time in other riverine species using dedicated experimental protocols.

All the findings presented in this Chapter complement studies that exclusively analysed fish interactions in open fields, away from walls. Studies conducted exclusively in such settings run the risk of potential misinterpretation, failing to accurately represent the natural behaviour of fish when swimming in proximity to walls as normally encountered in the form of river banks or concrete structures in fish passages. Additionally, our results provide valuable insights for Agent-based Models (ABM) developers seeking to incorporate upstream migration of fish groups [53,54]. We successfully quantified key parameters, including threshold distances for interactions (Figure 4.6), trends with relative positioning (Figure 4.3, Figure 4.4), response times (Figure 4.8), and the influence of boundaries on interactions (Figure 4.7). As ABMs gain prominence in conservation management, integrating evidence-based interaction rules - such as the effect of flow velocity and boundaries on fish collective behaviour - can significantly enhance their simulation applicability [226]. This is particularly pertinent in engineered sections like fish passages, where knowledge of fish behavioural patterns is essential to ensure high passage efficiency [5].

As a final remark, it is important to recognise the inherent limitations of correlation analysis. This method focuses solely on linear relationships, yet animal interactions can frequently be non-linear [196]. Moreover, correlation coefficients are limited in their ability to establish direct cause-and-effect relationships or provide precise quantification of information transfer magnitude and direction. Looking ahead, an intriguing path for future research could involve delving into these interactions using different approaches rooted in information theory. Such investigations have the potential to provide a comprehensive understanding of information transfer and leadership dynamics among fish as a function of different environmental factors.

Chapter 5. Measuring direct information transfer in animals

Abstract

This Chapter introduces a novel approach, termed "state-based transfer entropy", to investigate information transfer between animal pairs from their trajectory data. The research focuses on overcoming limitations associated with traditional linear cross-correlation analysis, such as linearity and the inability to capture genuine information transfer. The study adopts information theory, specifically transfer entropy, as a powerful tool to analyse the causal relationships governing fish interactions. Classical computations of transfer entropy from continuous variables face challenges in estimating their probability density distribution. Our approach overcomes this issue by classifying states based on a physical representation of the underlying system dynamics. In the validation process, the results obtained through the state-based transfer entropy approach are compared against findings from traditional cross-correlation analysis, presented in the previous Chapter. The aim is to demonstrate the efficacy and reliability of the state-based transfer entropy approach in uncovering meaningful insights into the causal dynamics within fish pairs.

5.1 Introduction

Information transfer, in the context of animal behaviour, refers to the exchange of meaningful signals or cues between individuals, influencing their behaviour and decision-making processes [227]. Understanding the mechanisms and patterns of information transfer among fish can offer insights into the formation of coordinated shoals, including navigation during collective migration [65,213]. In Chapter 4, linear cross-correlation analysis was utilised to assess fish interaction across three flow velocities in terms of distance of decorrelation, boundary effects, and response time. Linear cross-correlation analysis is a simple and widely used method but faces inherent limitations, primarily its linearity and the inability to capture genuine information transfer. Considering the prevalent nonlinear dynamics in animal behaviour [228], these limitations become substantial, necessitating alternative analytical tools to capture the intricacies inherent in animal interactions.

In this context, the application of information theory has emerged as a powerful tool, providing a data-driven framework to explore the causal relationships that govern the behaviours of living organisms [229]. Information theory is a branch of applied mathematics and electrical engineering developed by Shannon (1948) that involves quantifying and analysing the amount of information contained in signals and data [230]. It deals with concepts such as entropy and mutual information and offers a mathematical foundation for understanding information and its transmission. Mutual information quantifies the shared information between two variables and has been used in some studies of animal behaviour [231,232]. However, its inherent symmetry poses challenges in discerning the directionality of information flow – an essential aspect of understanding causal relationships.

Transfer Entropy (TE), first introduced by Schreiber (2000), shares some of the properties of mutual information but takes the dynamics and direction of information into account, assessing the flow of information from one variable to another over time [233]. TE considers the temporal sequence of events, providing an understanding of how one variable causally influences the future states of another. TE has emerged as a relevant metric for quantifying information transfer in fish, finding application in various studies examining collective information flow within shoals [234–236], leadership dynamics [237], fish-robot interactions [229,238], and prey-predator dynamics [239].

One of the significant drawbacks in utilising TE for analysing information transfer from experimental results lies in the challenge of estimating the probability density distribution (PDF) of continuous variables. Commonly employed estimators, such as the Gaussian-distribution model, Kernel estimation, and the Kraskov-Stogbauer-Grassberger

(KSG) technique, come with their own set of limitations [240]. The Gaussian distribution model, while parameter-free, assumes linearity – a critical limitation in the context of animal behaviour [241]. Kernel-estimators aim to replicate the PDF of a variable, providing model-free flexibility for measuring nonlinear relationships [233,242]. However, they are highly susceptible to the parameter choice for binning frequency. The Kraskov-Stogbauer-Grassberger (KSG) technique represents an advancement in kernel estimation, incorporating dynamic binning frequency adjustments based on the density of the samples, thereby mitigating errors in PDF estimation [243]. Despite their effectiveness in computing TE on continuous data, these estimators are associated with high computational costs and implement complex algorithms, often opaque to the user, potentially leading to misinterpretations of TE results.

In response to these challenges, Staniek and Lehnertz (2008) theorised an extension of TE employing a symbolic approach to discretise continuous variables (i.e., symbolic transfer entropy) [244]. This approach involves transforming continuous data into symbolic sequences by categorising the data into discrete symbols or states based on specific criteria, such as ranking the continuous variables as higher or lower. Symbolic TE provides a robust and computationally feasible alternative for capturing information transfer in complex systems, such as those found in biological or behavioural studies. Porfiri et al. (2017) applied this method to study fish information transfer [245]. Their approach, however, holds some limitations, as it is highly dependent on noise, with the risk of misinterpretation or loss of information. For instance, if a variable is nearly constant but noisy, the symbolic representation will still rank higher or lower states, potentially introducing misleading information into the analysis.

This Chapter presents a novel TE-based approach termed “state-based transfer entropy” to investigate information transfer between fish pairs. This method offers a simplified and robust solution to the challenges posed by continuous variables. It builds upon the symbolic TE introduced by Staniek and Lehnertz (2008) [244], but instead of ranking the variables as higher or lower, state-based TE classifies states that embody a physical representation of underlying system dynamics. To explore its potential and identify the limitations imposed by classical linear correlation analyses, the results obtained through this new approach are compared against findings presented in Chapter 4.

5.2 Methods

5.2.1 Entropy and mutual information

The concept of *entropy* originates from thermodynamics but has been applied in various fields, including information theory. In the context of information theory, entropy is a measure of the uncertainty or randomness associated with a random variable: the greater the uncertainty or the randomness, the higher the entropy.

For continuous random variables, calculating entropy entails integrating their probability density function (PDF). In frequent scenarios where the exact PDF of a continuous variable is unknown, such as in experimental research on natural phenomena, estimators can be applied to discretise the variable and approximate PDF [241–243]. This approach transforms the problem of entropy calculation into a sum of probabilities for an approximated discrete variable. For a discrete random variable X with probability distribution $P(x)$, the *entropy* $H(X)$ is calculated as:

$$H(X) = -\sum_{x \in X} P(x)\log\left[P(x)\right]$$

The choice of the logarithm base determines the unit of measurement: based 2 yields "bits", while a natural base e yields "nats".

In 1948, Shannon introduced the concepts of *joint entropy*, *conditional entropy* and *mutual information*, which serve as the pillars of information theory (Figure 5.1) [230].

For two discrete variables X and Y with a joint probability distribution of $P(x,y)$ (i.e. the probability of both occurring simultaneously), the *joint entropy* $H(X,Y)$ is given by:

$$H(X,Y) = -\sum_{\substack{x \in X \\ y \in Y}} P(x,y)\log\left[P(x,y)\right]$$

Conditional entropy $H(X|Y)$, instead, measures the information needed to describe the outcome of X when the value of another random variable Y is known. It can be computed as:

$$H(X|Y) = -\sum_{\substack{x \in X \\ y \in Y}} P(x,y)\log\left[\frac{P(x,y)}{P(y)}\right] = H(X,Y) - H(Y)$$

Mutual information $I(X;Y)$ between two discrete variables X and Y quantifies their mutual dependence and is given by:

$$I(X;Y) = -\sum_{x,y} P(x,y) \log\left[\frac{P(x,y)}{P(x)P(y)}\right]$$

It directly measures the amount of information, either in "nats" or "bits", gained about one variable by observing the other. An essential property of mutual information is its symmetry, meaning that $I(X,Y) = I(Y,X)$.

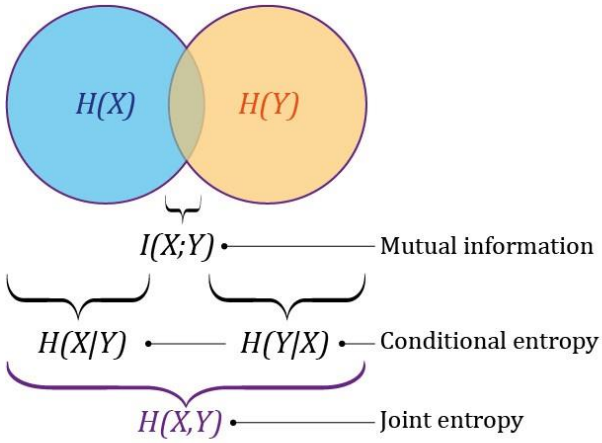

Figure 5.1: Information diagram for entropies in Information Theory

5.2.2 Transfer entropy

Transfer entropy (TE) is grounded in the concept of *conditional mutual information* from Y to X given Z ($I(X;Y|Z)$). This quantifies how much information the *source* variable Y provides about the *destination* variable X, given the known value of a third variable Z (*condition*). When we consider two discrete variables assuming specific values over time ($X(t), Y(t)$), this metric can quantify the impact of variable Y in predicting the state of variable X (see representation in Figure 5.2) [233]. In this case, the *source* is variable Y at the time step t (Y_t), the *destination* is variable X at the future time step (X_{t+1}), and the *condition* is the current state of destination variable X (X_t). This formulation leads to the concept of TE from the variable Y to the variable X denoted as:

$$TE_{Y\to X} = I(X_{t+1}; Y_t | X_t) = H(X_{t+1}|X_t) - H(X_{t+1}|X_t, Y_t)$$

This metric captures the directed influence or predictability between variables and is asymmetrical ($TE_{Y\to X} \neq TE_{X\to Y}$). Mathematically, TE is computed as:

$$TE_{Y\to X} = \sum_{\substack{x_{t+1}\in X \\ x_t\in X \\ y_t\in Y}} P(x_{t+1}, x_t, y_t) \log\left[\frac{P(x_{t+1}|x_t, y_t)}{P(x_{t+1}|x_t)}\right]$$

By rearranging terms using the relation $P(A|B)\cdot P(B) = P(A,B)$, the explicit formula for TE becomes:

$$TE_{Y\to X} = \sum_{x,y} P(x_{t+1}, x_t, y_t) \log\left[\frac{P(x_{t+1}, x_t, y_t)P(x_t)}{P(x_{t+1}, x_t)P(x_t, y_t)}\right]$$

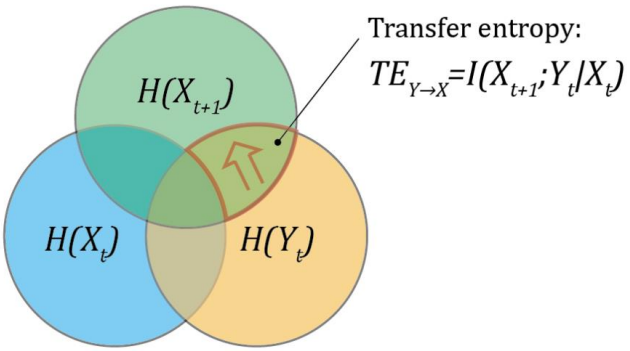

Figure 5.2: Information diagram for transfer entropy. The orange circle denotes the entropy associated with the source variable (Y), the blue circle represents the entropy linked to the current state of the destination variable (X), and the green circle represents the entropy associated with its future state. Transfer entropy from the source variable Y to the destination variable X is depicted by the red contoured shape.

5.2.3 Effective, net and total transfer entropies

Depending on the nature of the discrete variables, which may not be purely random, calculating TE could yield misleading results. In order to mitigate this issue, it is advisable to compute *Effective Transfer Entropy* (ETE), which removes potential biases from the quantification of information transfer [246]. It is advisable to run multiple synthetic simulations to ensure that ETE is representative of the noise in the destination variable. ETE is calculated by subtracting from TE an average TE from "synthetic" sources ($Y_{sinth,i}$) entirely uncorrelated with the destination variable. Being N_s the number of synthetic simulations, ETE is calculated as follows:

$$ETE_{Y\to X} = TE_{Y\to X} - \frac{1}{N_s}\sum_{i}^{N_s} TE_{Ysynth,i\to X}$$

An example of a synthetic source involves utilising the trajectory of a fish from a distinct experiment. It is noteworthy that ETE is invariably a non-negative value ($ETE \geq 0$). Thus, in cases where the mean synthetic TE surpasses the actual TE, the ETE will be zero.

In situations where measuring the directionality of information transfer is of interest, such as identifying follower/leader dynamics, the *net Transfer Entropy* ($netTE$) can be computed:

$$netTE_{Y\to X} = TE_{Y\to X} - TE_{X\to Y}$$

Similarly, the *total Transfer Entropy* ($totETE$) represents the total amount of information transmitted within the system:

$$totTE_{Y\leftrightarrow X} = TE_{Y\to X} + TE_{X\to Y}$$

Additionally, *net* and *total Effective Transfer Entropies* ($netETE$, $totETE$) can be computed by analysing the difference and the sum of ETE from Y to X and from X to Y, respectively.

5.2.4 State-based approach

To address the challenges of computing TE on continuous variables, we propose a novel approach: the state-based transfer entropy. This approach is inspired by the symbolic TE introduced by Staniek and Lehnertz (2008) [244], which approximates the system's behaviour as a sequence of discrete symbols based on a given criteria. In contrast with criteria that categorise variables as higher or lower than their previous value, such as the one Porfiri et al. (2017) applied to fish trajectories [245], we suggest defining states as a physical representation of fish motion.

In its simplest form, state-based TE involves a three-state system, where fish motion is categorised as forward, backwards, or stationary (states 1, -1, and 0 in Figure 5.3a, respectively):

$$\begin{cases} if\ u < -u_{th} \ \to state\ "-1" \\ if\ u > u_{th} \ \to state\ "1" \\ if -u_{th} \leq u \leq u_{th} \ \to state\ "0" \end{cases}$$

In this approach, discrete variables are represented as vectors with discrete elements $[X_t, Y_t] \in [-1,0,1]$. The intermediate "stationary" state, defined by the threshold velocity u_{th}, categorises fish movement. Longitudinal velocities below this threshold are considered insignificant, allowing the capture of meaningful changes in motion and responses between fish. This choice not only reduces system instability to noise - especially crucial when the movement along the flow's longitudinal axis (x-axis) is nearly zero - but also reflects actual swimming behaviour. Within a stationary reference system, in fact, these states can be associated with bursting, drifting, and station holding, respectively.

This discretised framework proves particularly useful when analysing information transfer from animal trajectories. It directs focus toward changes in motion in response to a group mate rather than emphasizing velocity magnitude, which may not be particularly significant in this context. Importantly, this approach eliminates challenges associated with binning and significantly reduces computation costs.

Likewise, a three-state system can be defined for the direction transversal to the flow, with the definition of a threshold velocity on the lateral component (v_{th}, see Figure 5.3b):

$$\begin{cases} if\ v < -v_{th} \rightarrow state\ "-1" \\ if\ v > v_{th} \rightarrow state\ "1" \\ if -v_{th} \leq v \leq v_{th} \rightarrow state\ "0" \end{cases}$$

This configuration allows the system to capture a significant lateral motion of the fish, distinguishing whether it swims towards the right side of the flume, the left side, or remains neutral.

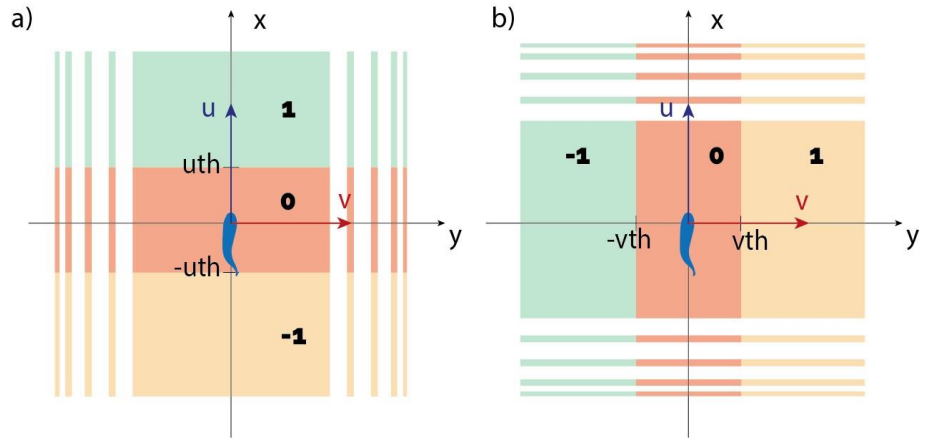

Figure 5.3: Definition of three-state system on the longitudinal direction (a), and lateral direction (b)

5.2.5 Data analysis

Data analysis was carried out using R version 4.0.5 [166]. A detailed description of the experimental protocol and trajectory data can be found in Chapter 3 and Chapter 4. Similar to the linear cross-correlation analysis discussed in Chapter 4, the reference system was defined with its origin on the head of the Leading fish (L). The x-axis is aligned with the flow direction, and the y-axis is oriented transversely to the water flow within the horizontal plane (see Figure 4.2).

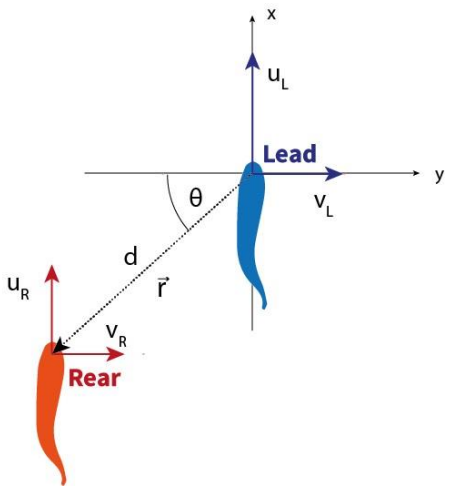

Figure 5.4: Coordinate reference system

As in the analysis presented in Chapter 4, we explored information transfer among fish pairs across various spatial configurations and response times for the three flow velocities employed (bulk velocity, U_B). Specifically, we investigated total and net effective transfer entropies for different relative distances (d), orientations (θ) and response times (τ): $totETE_u(d,\theta,\tau)$, $totETE_v(d,\theta,\tau)$, $netETE_u(d,\theta,\tau)$, $netETE_v(d,\theta,\tau)$. The effects of d and θ were investigated as of Chapter 4, with ETE being computed for distance intervals $2(i-1) < d < 2i$ BL with $i \in [1 \dots 7]$ and orientation intervals $\pi/12\,(i-1) < \theta < \pi/12\,i$ with $i \in [1 \dots 6]$.

Response time was investigated by introducing a time lag (τ) into the source variable:

$$TE_{Y,\tau\to X} = I(X_{t+1}; Y_{t-\tau}|X_t)$$

Effective transfer entropies with varying time lags (τ) were computed to quantify response times, spanning between 0 to 5 seconds. Similarly to the approach taken in Chapter 4 with cross-correlation analysis, the peaks in $totETE$ were used as estimators

to depict the mean response time of fish interacting with one another [33]. The response time t_r was calculated as the mean lags at which the maximum $totETE$ occurred for each experiment (n):

$$t_{r,u} = \frac{1}{N}\sum_{n=1}^{N} \tau_{MAX,u,n}$$

$$\tau_{MAX,u,n} : TE_{Y,u,n,\tau_{MAX}\to X,u,n} = \max_{\tau}(TE_{Y,u,n,\tau\to X,u,n})$$

5.2.6 Parameter selection

In order to compare the results obtained from TE and linear cross-correlation analysis, we measured $totETE$ and $netETE$ for two distinct three-state systems, one for the longitudinal component (u, Figure 5.3a), and one for the lateral velocity component (v, Figure 5.3b). Velocities of the Lead and Rear fish represented source and destination variables accordingly. For each of the three flow velocities (U_{B10}, U_{B20}, and U_{B35} representing 10, 20 and 35 cm/s respectively), calculations of $totETE$ and $netETE$ involved averaging the results across the ten experiments ($N = 10$):

$$\begin{cases} totETE_u = \frac{1}{N}\sum_{n}^{N}(TE_{u_{L,n}\to u_{R,n}} + TE_{u_{R,n}\to u_{L,n}}) \\ netETE_u = \frac{1}{N}\sum_{n}^{N}(TE_{u_{L,n}\to u_{R,n}} - TE_{u_{R,n}\to u_{L,n}}) \\ totETE_v = \frac{1}{N}\sum_{n}^{N}(TE_{v_{L,n}\to v_{R,n}} + TE_{v_{R,n}\to v_{L,n}}) \\ netETE_v = \frac{1}{N}\sum_{n}^{N}(TE_{v_{L,n}\to v_{R,n}} - TE_{v_{R,n}\to v_{L,n}}) \end{cases}$$

To determine threshold velocities for defining states (u_{th}, v_{th}), a sensitivity analysis was conducted on $totETE$. Initial values of synthetic simulations (N_s) was set to 30; details on the retrieval of synthetic simulations are provided later in this Section. For every U_B, $totETE_u$ and $totETE_v$ were computed with u_{th} and v_{th} ranging from 0 to 10 cm/s ($u_{th}, v_{th} \in$ [0, 0.1, 0.5, 1, 1.5, 2, 5, 10] cm/s). Time frames in which fish were more distant than the decorrelation distance (D_d = 7 Body Lengths, BL), retrieved in Chapter 4, were filtered out for sensitivity analyses. Threshold velocities u_{th} and v_{th} were identified as the velocities corresponding to peaks in $totETE_u$ and $totETE_v$, respectively.

Once that u_{th} and v_{th} were determined, a second sensitivity analysis was conducted to identify the optimal number of synthetic simulations (N_s) to compute ETE. Values of $totETE$ were calculated for N_s ranging from 0 to 100 ($N_s \in$ [0, 5, 10, 15, 20, 30, 50,

100]). Synthetic simulations quantify the amount of noise in the destination variable by calculating information transfer between this variable and another completely unrelated one. For each of the 30 experiments (i.e., ten experiments at U_{B10}, U_{B20}, and at U_{B35}), synthetic simulations were generated by randomly sampling fish trajectories from the remaining 29. As the synthetic source is unrelated to the destination, the Lead fish or the Rear trajectories were sampled regardless. For the n^{th} experiment, synthetic TE was calculated using, as the source variable, the state time series of a fish from another experiment (m with $m \neq n$) randomly sampled. For instance, synthetic TE for the longitudinal component of the Rear fish in the n^{th} experiment is computed as:

$$TE_{synth \to uR,n} = TE_{u,m \to uR,n} \quad ; \; m \neq n$$

Where u, m represents the longitudinal state vector of either the Lead or Rear fish from the m^{th} experiment. This definition provides a pool of 58 total trajectories to randomly sample from. When N_s exceeded this value, an option to reverse trajectories in time was added, effectively doubling this number. The ETE of the Lead fish to the Rear fish in the n^{th} experiment is computed by substituting TE with the mean synthetic TE calculated for a number of simulations N_s:

$$ETE_{uL,n \to uR,n} = TE_{uL,n \to uR,n} - \frac{1}{N_s} \sum_{i}^{N_s} TE_{synth,i \to uR,n}$$

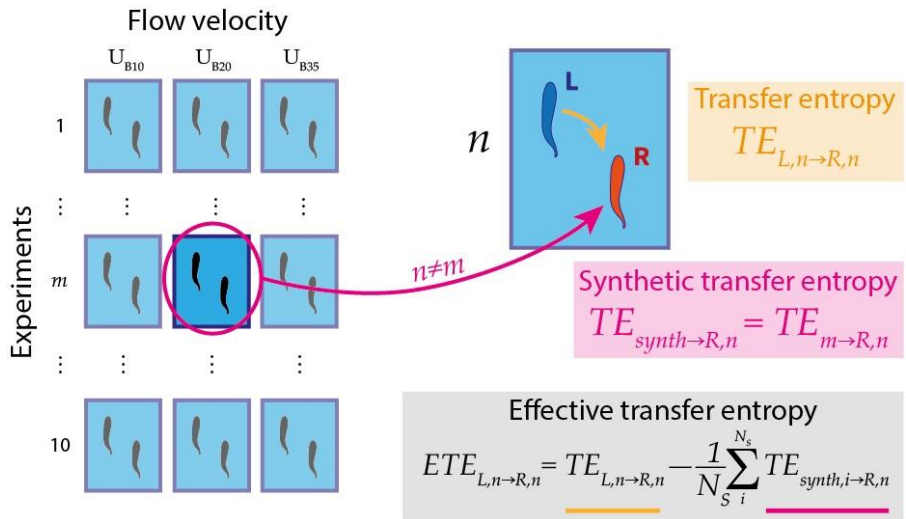

Figure 5.5: Synthetic simulations for calculation of effective transfer entropy (ETE) from the Lead fish to the Rear one in the n^{th} experiment. The ETE (grey box) is calculated by subtracting a synthetic transfer entropy (TE_{synth}, pink box) from the transfer entropy (TE, yellow box). The synthetic transfer entropy is calculated using trajectory data from another experiment ($m \neq n$) as source variable.

5.3 Results

5.3.1 Parameter selection

The sensitivity analysis on threshold velocities (u_{th}, v_{th} Figure 5.6) revealed a discernible trend in $totETE$ for both velocity components. $totETE$ exhibited an initial increase until values near one cm/s, followed by a subsequent decrease. In both longitudinal and lateral components, the apex of $totETE$ was observed at U_{B20}, succeeded by U_{B10}, while the highest velocity (U_{B35}) exhibited the lowest $totETE$. For subsequent analyses, the values of u_{th} and v_{th} were set to one cm/s for all flow velocities, as it approximately corresponds to the condition at which $totETE$ reached a maximum for all U_B and both velocity components. This choice resulted in state 0, representing "stationary" motion, being the most frequently observed state in both components (Figure 5.7).

The results from the second sensitivity analysis on the number of synthetic simulations (Figure 5.8) demonstrated that $totETE$ decreased with increasing N_s until reaching a plateau at N_s = 20. This consistent trend was observed across all U_B in both the longitudinal (Figure 5.8a) and lateral components (Figure 5.8b). In order to enhance computational efficiency, the number of synthetic simulations for ETE computation was optimised and set at N_s = 20. In alignment with the sensitivity analysis on threshold velocities, $totETE$ exhibited its highest values at U_{B20}, succeeded by U_{B10} and U_{B35}, respectively.

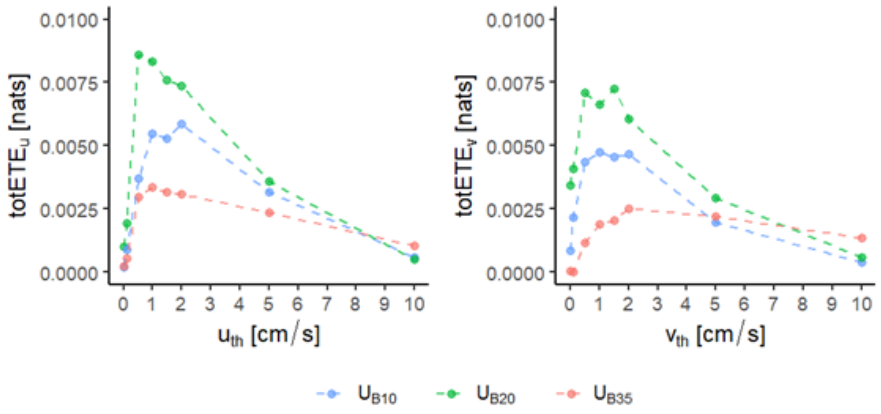

Figure 5.6: Sensitivity analysis on $totETE$ for threshold velocities on the longitudinal (a) and lateral (b) components

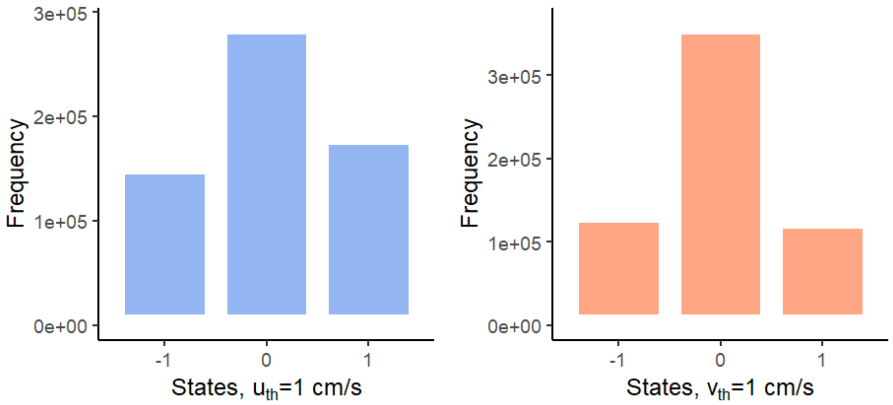

Figure 5.7: State frequency for threshold velocities of 1 cm/s for longitudinal (a) and lateral (b) components

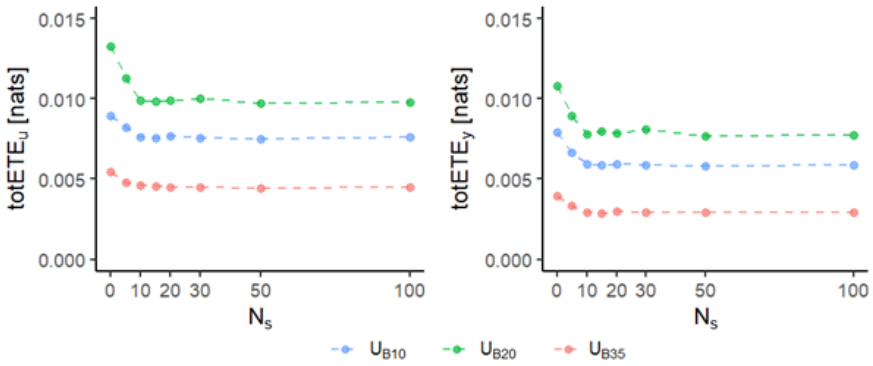

Figure 5.8: Sensitivity analysis for the number of synthetic simulations (N_s) for the computation of total effective transfer entropy ($totETE$) on the longitudinal (a) and lateral (b) components

5.3.2 Relative distance and orientation

Analysing the total effective transfer entropy between fish pairs (Figure 5.9) provides significant insights. Regarding the effect of relative distance (d), the longitudinal velocity component exhibits an expanded range of positive $totETE_u$ values with increasing U_B (Figure 5.9a). At the U_{B10}, $totETE_u$ is high when fish are less than 2 BL apart, diminishing rapidly to zero at 6 BL. This trend persists U_{B20}, extending information transfer until a distance of 8 BL. For U_{B35}, though $totETE_u$ is generally lower, some transfer persists up to when fish are 10 BL apart. A similar pattern is observed for the lateral component (Figure 5.9b), although short-range $totETE_v$ appears slightly lower than for the longitudinal component.

Regarding relative orientation (Figure 5.9 c,d), the effect is less pronounced than relative distance (Figure 5.9 a,b). At U_{B10}, the maximum value of $totETE_u$ occurs at θ =

90° (Figure 5.9c), indicating fish aligned in an in-line pattern. At both U_{B10} and U_{B20}, $totETE_u$ shows oscillations without a discernible trend. Conversely, at U_{B35}, $totETE_u$ peaks around 40°. For the lateral velocity component, this peak is slightly extended at around 40-50° (Figure 5.9d). At U_{B10} and U_{B20}, $totETE_v$ is higher for low angles, corresponding to fish in a side-by-side configuration.

Examining the net effective transfer entropy offers insights into the directionality of information flow (Figure 5.10). A positive $netETE$ indicates predominantly Front-to-Rear information flow, while a negative value of $netETE$ suggests a stronger Rear-to-Front flow. For the longitudinal velocity component, $netETE_u$ is close to zero for short-range distances (i.e., less than 2 BL), remaining overall positive for distances between 2 and 6 BL. Subsequently, it returns to zero, coinciding with totETE being zero as well. A similar trend is observed for the lateral component, except for U_{B35}, where $netETE_v$ is slightly below zero for relative distances between 2 and 7 BL.

With regards to the effect of relative orientation (θ) on the direction of information flow, it can be seen that in the longitudinal component, $netETE_u$ remains approximately zero for U_{B10} and U_{B20} (Figure 5.10c). However, at U_{B35}, a dominant information flow emerges from the Rear fish to the Lead one, particularly evident at an angle θ around 40°. In contrast, for the lateral component, U_{B10} exhibits oscillations of $netETE_v$ around zero, indicating no prevalent information flow (Figure 5.10d). At U_{B20}, $netETE_v$ turns negative for angles between 20 and 60°. For U_{B35}, the information flow is directed from Rear to Lead for angles below 20°, while for higher angles, the Rear fish appears to respond to the Lead fish.

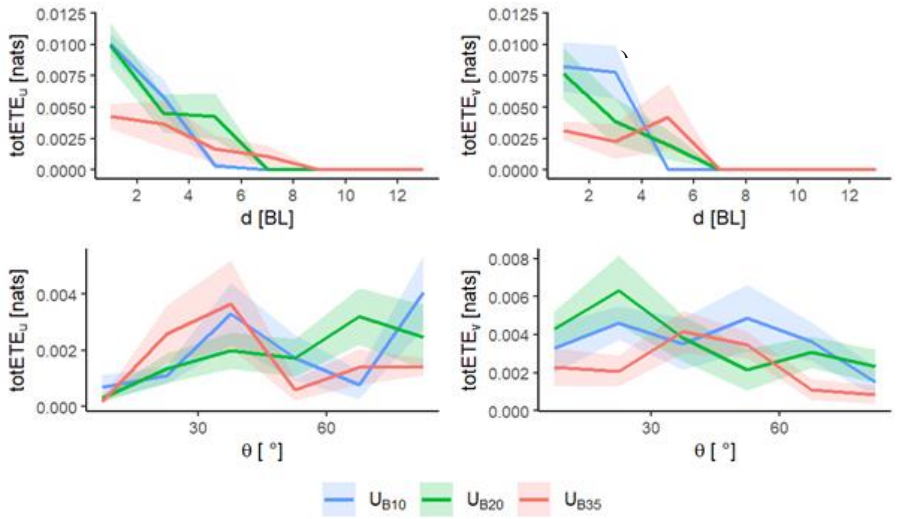

Figure 5.9: Total effective transfer entropy on longitudinal (a,c) and lateral (b,d) components for different relative distances (a,b) and orientations (c,d). θ=0° corresponds to a side-by-side configuration, while θ=90° stands for an in-line arrangement. The coloured bands represent the standard deviation (N=10 experiments).

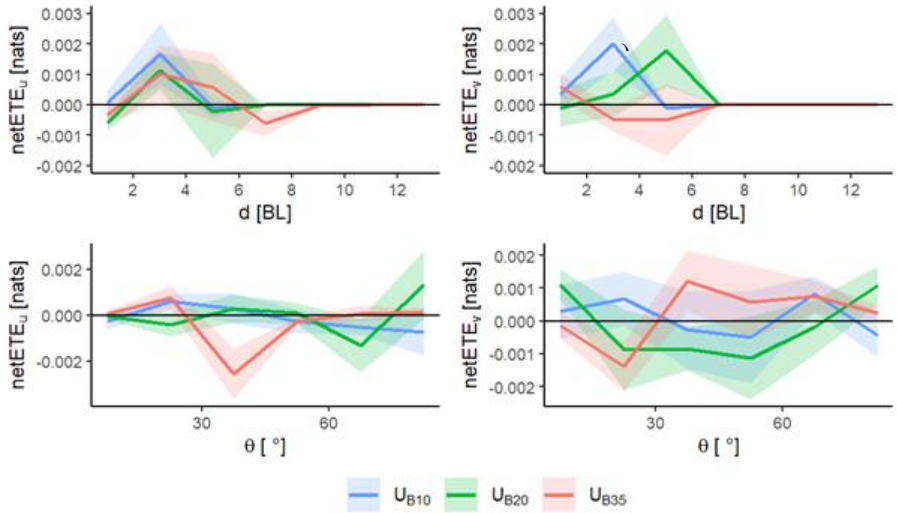

Figure 5.10: Net effective transfer entropy for the three flow velocities U_B (U_{B10} = 10 cm/s in blue, U_{B20}= 20cm/s in green and U_{B35}= 35cm/s in red). Left panels (a,c) represent the longitudinal component, and right (b,d) represent the lateral one. Top panel different relative distances (a,b) and orientations (c,d). θ=0° corresponds to a side-by-side configuration, while θ=90° stands for an in-line arrangement. The coloured bands represent the standard deviation (N=10 experiments).

5.3.3 Response time

The analysis of response times reveals distinct patterns for the longitudinal and lateral components (Figure 5.11 a,b, respectively). For the longitudinal component, the

highest ETE_u is observed at U_{B20}, while at U_{B10} and U_{B35}, ETE_u shows comparable values. Overall, $ETE_{u_L \to u_R}$ from the Lead fish to the Rear one (dashed lines, Figure 5.11a) is comparable with $ETE_{u_R \to u_L}$ in the opposite direction (solid lines, Figure 5.11a). The response times (t_r) for longitudinal motion are found to be 0.29 ± 0.32 s, 0.23 ± 0.35 s, and 0.13 ± 0.09 s for U_{B10}, U_{B20}, and U_{B35} respectively.

On the lateral component, information transfer is generally higher, with the highest ETE_v observed at U_{B20}, followed by U_{B10} and U_{B35}, respectively. Information flows predominantly from the Lead to the Rear ($ETE_{v_L \to v_R} > ETE_{v_R \to v_L}$ dashed lines and solid lines, respectively, Figure 5.11b) for all U_B. The response times for the lateral component are 0.22 ± 0.19 s for U_{B10}, 0.31 ± 0.30 for U_{B20}, and 0.17 ± 0.15 for U_{B35}. Statistical analysis indicates no significant effect of U_B on t_r for either the longitudinal or lateral components (Friedman test, p-value > 0.05).

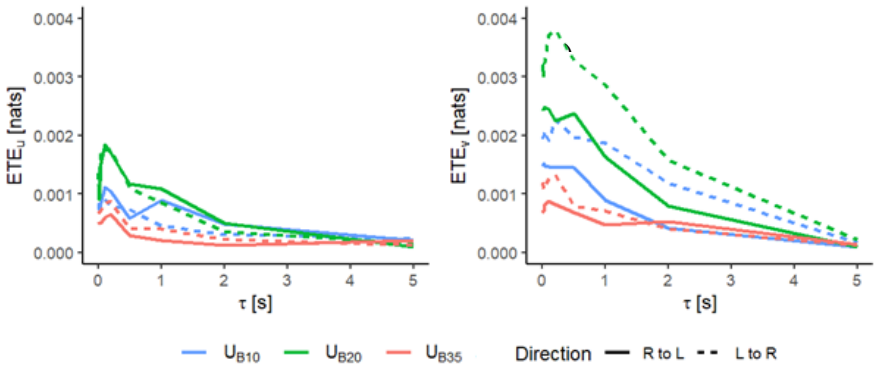

Figure 5.11: Effective transfer entropies at various time lags (τ) for the three flow velocities U_B (U_{B10} = 10 cm/s in blue, U_{B20}= 20cm/s in green and U_{B35}= 35cm/s in red). The left (a) panel represent the longitudinal component, and the right (b) panel represent the lateral component. Solid lines denote effective transfer entropy computed with the Rear fish's state as the source variable and the Lead fish's state as the destination variable. Dashed lines indicate ETE with the information flow reversed from the Lead fish towards the Rear one.

5.4 Discussion

In this study, we introduced a state-based Transfer Entropy (TE) approach to quantify information transfer, providing an alternative to linear cross-correlation analysis. Unlike linear cross-correlation, our approach captures non-linear patterns and allows for the description of directional flow. The state-based TE, while requiring the fine-tuning of parameters such as velocity thresholds (u_{th}, v_{th}, Figure 5.6) and the number of synthetic simulations (N_s, Figure 5.8), provides a straightforward method for investigating information transfer dynamics. Note that the choice of threshold velocities carries a meaningful interpretation, as the "stationary" state emerged as the most frequent state in both components (see Figure 5.7). This observation aligns with expectations, considering that the interacting fish likely perceive states involving motion as significant events.

Comparing state-based TE results with those obtained from linear cross-correlation analysis (Chapter 4) reveals both similarities and differences. Results from $totETE$ on relative distance and orientation (Figure 5.9) show similar major trends to mean linear correlation. Despite these similarities, some differences can be observed. Firstly, the distance at which fish pairs ceased to interact appeared to be more extended when measured through state-based TE than cross-correlation analysis. This disparity could stem from nonlinear interactions influencing the observed patterns [228], or it might be generated from the lack of a threshold for $totETE$ to ascertain the significance of information transfer—a strategy we previously employed in linear cross-correlation analysis (utilizing a threshold of 0.05). Secondly, results from the response time exhibited no statistical influence of flow velocity, and the peaks in response times appeared to be less clear (i.e. more diffused) compared to those observed in the cross-correlation analysis.

The most notable contribution of the state-based TE approach lies in its ability to distinguish the direction of information flow - from the Lead fish to the Rear fish and vice versa. Values of $netETE$ not only describe the directionality of information flow but also quantify its magnitude. For short-range distances (i.e., less than 2BL), information flow is bidirectional, as indicated by nearly zero $netETE_u$ and $netETE_v$ (Figure 5.10 a and b, respectively) while $totETE_u$ and $totETE_v$ display peaks (Figure 5.9 a,b, respectively). This implies that when fish are in close proximity, the Lead fish respond to the Rear as much as the Rear responds to the Lead. This might seem counterintuitive, as the Rear fish is often outside the Lead fish's field of view. However, at such short distances, the Lead fish can still sense the Rear one even when it is not visible. This phenomenon is enabled by the lateral line, which is composed of neuromasts that extend to the fish's tail [223], allowing the Lead to detect of the trailing neighbour through the pressure field generated by the flow diversion around its body.

As the distance between fish increases, the predominant direction of information flow shifts from front to back. This directionality is particularly evident in both the longitudinal and lateral components, with the exception of the lateral component at the highest flow velocity (Figure 5.10b). In this case, $netETE_v$ approaches zero, albeit slightly negative. Importantly, the associated $totETE_v$ is found to be relatively limited (Figure 5.9b), implying that information flow in the lateral component at high velocities is diminished, regardless of its direction. These findings provide insights into the nonlinear and bidirectional nature of information transfer among fish pairs, reinforcing and extending our earlier discussions in Chapter 4.

Regarding response time, the findings appear less conclusive compared to those of the linear cross-correlation analysis, showing considerable individual variation. Unlike the observations in Chapter 4, no discernible effect of U_B on response time was observed. This discrepancy might arise from nonlinear dynamics in fish interactions, which could not be captured by linear cross-correlation analysis. Despite the observed variability, the peaks on the longitudinal component of ETE_u were found to be indistinguishable between the front-to-back and back-to-front directions. This observation confirms the bi-directionality of information flow along the longitudinal component, as hypothesised in Chapter 4. In contrast, the lateral component exhibited distinct peaks only in the front-to-back direction of ETE_v, whereas the reverse direction displayed a more ambiguous response. The overall response times, as depicted by the state-based TE, ranged from 0.13 to 0.31 s.

Looking ahead, it would be valuable to fully explore other features of the state-based TE approach, with particular emphasis on its inherent flexibility in extending the number of states as necessary, thus enabling more intricate categorisations. Conventional methods of computing TE quantify information transfer between only two variables, often limited to velocity or acceleration on one component. In contrast, the state-based TE categorises fish motion into a finite number of states, which allows the consideration of more components. For instance, future studies could explore the dynamics of information transfer within a system defined by both longitudinal and lateral components simultaneously, such as a five-state system (Figure 5.12 as an example).

Such system could provide a more comprehensive yet discretised physical representation of fish motion in the horizontal plane. This proves especially valuable when studying fish in an open channel flow, where the longitudinal and lateral components carry distinct meanings that can evolve with changes in velocity. Thus, our approach provides a practical method to extend TE analysis to continuous data, offering comprehensive insights into the information flow between variables. It is important to acknowledge that discretising the system into finite transition states may introduce some loss of detail. However, the state-based TE approach enables a comprehensive analysis

by encompassing states that reflect the fish's physical swimming behaviours, such as bursting, drifting, and station holding for the longitudinal component.

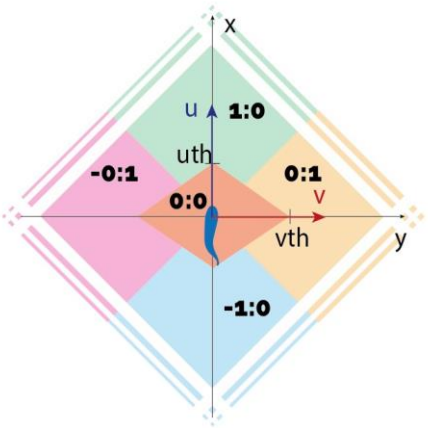

Figure 5.12: Definition of five-state system for the representation of fish motion within the horizontal plane

In conclusion, the choice of an approach for studying animal interactions should be guided by the specific goals of the study. If the primary aim is to quantify interaction, classical linear cross-correlation analysis, as demonstrated in Chapter 4, serves this purpose effectively. However, TE-based approaches are more appropriate, albeit more complex, when the goal is to evaluate actual information transfer and determine its directionality, particularly in situations such as studying leader personality traits or predator-prey interactions. The method proposed in this study seamlessly integrates the robustness of TE in computing information transfer among variables with the simplicity of a symbolic approach. This combination offers a valuable tool for researchers seeking a comprehensive understanding of information flow in complex behavioural dynamics.

Conclusions

In this study, we delved into the intricate relationship between flow velocity and collective behaviour. A comprehensive review of existing literature was conducted, exploring various facets where this interplay holds practical significance, including energy expenditure, navigation, stress, exploration and predation. We then presented a specialised methodology grounded on deep-learning that successfully allowed the tracking fish groups in videos with an unstable background. This method was then applied to extract data from an experimental campaign involving groups of *Telestes muticellus* under three different flow velocities. The results revealed a combined effect of flow velocity and social facilitation (objective (i)). Increased group size enhanced exploration and swimming activity among fish, but this effect was found to be modulated by flow velocity. Importantly, social facilitation did not play a significant role under low flow conditions, emphasizing the importance of incorporating hydrodynamics into the analysis of riverine fish behaviour. Linear cross-correlation analysis was then utilised to assess interaction between fish pairs (objective (ii)). The analysis of linear interaction between fish velocities indicated that interactions predominantly occur within a relative distance of 6BL, with lateral motion exhibiting more dominance than longitudinal motion, especially in side-by-side configurations. The innovative state-based transfer entropy approach further validates these findings, providing additional insights into information flow dynamics (objective (iii)). Interestingly, the results indicate that, at medium velocities, fish display the highest levels of exploration, swimming activity, and information transfer.

Moving forward, there are several avenues for future research that can enhance our understanding of the intricate dynamics between flow conditions and collective behaviour in riverine fish. Firstly, exploring additional hydrodynamic parameters beyond flow velocity, particularly turbulence, could offer a more nuanced perspective on how environmental factors shape fish behaviour. Secondly, extending our interaction analysis to encompass shoals with greater numbers than fish pairs will provide a more comprehensive view of group dynamics.

Furthermore, translating our experimental findings into practical applications, such as developing design guidelines for fish passage or integrating them into existing models simulating fish migration, holds immense potential for promoting the conservation of migratory species. To achieve this, further steps are needed to convert and apply fundamental research into practical guidelines. For example, experimental studies investigating group dynamics at simulated fish passages could be the first step in understanding how the findings presented in this work apply to more complex environments. Addressing these aspects in future studies will can contribute to the development of more effective strategies for fish passage design and the overall preservation of freshwater fish populations.

References

1. Deinet, S. *et al.* The Living Planet Index for Migratory Freshwater Fishes: 2024 update. (2024).

2. Larinier, M. Dams and fish migration. *Dams, Ecosyst. Funct. Environ. Restor.* 30 (2001).

3. Katopodis, C., Kells, J. A. & Acharya, M. Nature-like and conventional fishways: Alternative concepts? *Can. Water Resour. J.* **26**, 211–232 (2001).

4. Larinier, M. Pool Fishways, Pre-Barrages and Natural Bypass Channels. *Bull. Français la Pêche la Piscic.* **364**, 54–82 (2002).

5. Noonan, M. J., Grant, J. W. A. & Jackson, C. D. A quantitative assessment of fish passage efficiency. *Fish Fish.* **13**, 450–464 (2012).

6. Bunt, C. M., Castro-Santos, T. & Haro, A. Performance of fish passage structures at upstream barriers to migration. *River Res. Appl.* **28**, 457–478 (2012).

7. Silva, A. T. *et al.* The future of fish passage science, engineering, and practice. *Fish Fish.* **19**, 340–362 (2018).

8. Krause, J. & Ruxton, G. D. *Living in Groups*. *Transactional Analysis Journal* vol. 43 (Oxford University Press, 2013).

9. Mozzi, G., Manes, C., Nyqvist, D., Domenici, P. & Comoglio, C. Aggregation in Riverine Fish: A Review from a Fish Passage Perspective. *Int. Sch. Hydraul.* 265–280 (2024) doi:10.1007/978-3-031-56093-4_21.

10. Mozzi, G. *et al.* The interplay of group size and flow velocity modulates fish exploratory behaviour. *Sci. Rep.* **14**, 13186 (2024).

11. Ceballos, G. *et al.* Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Sci. Adv.* **1**, 9–13 (2015).

12. Strayer, D. L. & Dudgeon, D. Freshwater biodiversity conservation: Recent progress and future challenges. *J. North Am. Benthol. Soc.* **29**, 344–358 (2010).

13. Northcote, T. G. Migratory strategies and production in freshwater fishes. *Ecol. Freshw. fish Prod.* 326–359 (1978).

14. Deinet, S. *et al.* The Living Planet Index (LPI) for migratory freshwater fish. 55 pp (2020).

15. Zarfl, C., Lumsdon, A. E., Berlekamp, J., Tydecks, L. & Tockner, K. A global boom in hydropower dam construction. *Aquat. Sci.* **77**, 161–170 (2015).

16. Grill, G. *et al.* Mapping the world's free-flowing rivers. *Nature* **569**, 215–221 (2019).

17. Lehner, B. *et al.* High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Front. Ecol. Environ.* **9**, 494–502 (2011).

18. Birnie-Gauvin, K. *et al.* River connectivity reestablished: Effects and implications of six weir removals on brown trout smolt migration. *River Res. Appl.* **34**, 548–554 (2018).

19. Watson, J. M., Coghlan, S. M., Zydlewski, J., Hayes, D. B. & Kiraly, I. A. Dam Removal and Fish Passage Improvement Influence Fish Assemblages in the Penobscot River, Maine. *Trans. Am. Fish. Soc.* **147**, 525–540 (2018).

20. Gough, P., Fernández Garrido, P. & Van Herk, J. Dam Removal: A viable solution for the future of our European rivers. *Dam Remov. Eur* (2018).

21. Berga, L. The Role of Hydropower in Climate Change Mitigation and Adaptation: A Review. *Engineering* **2**, 313–318 (2016).

22. López-Moreno, J. I. *et al.* Dam effects on droughts magnitude and duration in a transboundary basin: The Lower River Tagus, Spain and Portugal. *Water Resour. Res.* **45**, (2009).

23. Dixit, A. Floods and vulnerability: Need to rethink flood management. *Nat. Hazards* **28**, 155–179 (2003).

24. Edenhofer, O. *et al. Renewable Energy Sources and Climate Change Mitigation: Special Report of the Intergovernmental Panel on Climate Change*. (Cambridge University Press, 2011).

25. Roscoe, D. W. & Hinch, S. G. Effectiveness monitoring of fish passage facilities: Historical trends, geographic patterns and future directions. *Fish Fish.* **11**, 12–33 (2010).

26. Quaranta, E., Katopodis, C. & Comoglio, C. Effects of bed slope on the flow field of vertical slot fishways. *River Res. Appl.* **35**, 656–668 (2019).

27. Castro-Santos, T. Optimal swim speeds for traversing velocity barriers: An analysis of volitional high-speed swimming behavior of migratory fishes. *J. Exp. Biol.* **208**, 421–432 (2005).

28. Katopodis, C. & Williams, J. G. The development of fish passage research in a historical context. *Ecol. Eng.* **48**, 8–18 (2012).

29. Ead, S. A., Katopodis, C., Sikora, G. J. & Rajaratnam, N. Flow regimes and structure in pool and weir fishways. *J. Environ. Eng. Sci.* **3**, 379–390 (2004).

30. Clay, C. H. *Design of Fishways and Other Fish Facilities*. *Design of Fishways and Other Fish Facilities* (CRC Press, 1995). doi:10.1201/9781315141046.

31. Odeh, M. Fish passage innovation for ecosystem and fishery restoration. *Am. Fish. Soc.* 24 (1999).

32. Castro-Santos, T., Cotel, A. & Webb, P. W. Fishway evaluations for better bioengineering: an integrative approach. in *Challenges for diadromous fishes in a dynamic global environment. American Fisheries Society, Symposium* vol. 69 557–575 (2009).

33. Katz, Y., Tunstrøm, K., Ioannou, C. C., Huepe, C. & Couzin, I. D. Inferring the structure and dynamics of interactions in schooling fish. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 18720–18725 (2011).

34. Hemelrijk, C. K. & Hildenbrandt, H. Schools of fish and flocks of birds: Their shape and internal structure by self-organization. *Interface Focus* **2**, 726–737 (2012).

35. Kasumyan, A. O. & Pavlov, D. S. Evolution of schooling behavior in fish. *J. Ichthyol.* **58**, 670–678 (2018).

36. Tunstrøm, K. *et al.* Collective States, Multistability and Transitional Behavior in Schooling Fish. *PLoS Comput. Biol.* **9**, (2013).

37. Pitcher, T. J. Heuristic definitions of fish shoaling behaviour. *Anim. Behav.* **31**, 611–613 (1983).

38. Strandburg-Peshkin, A. *et al.* Visual sensory networks and effective information

transfer in animal groups. *Curr. Biol.* **23**, R709–R711 (2013).

39. Chicoli, A. *et al.* The effects of flow on schooling Devario aequipinnatus: School structure, startle response and information transmission. *J. Fish Biol.* **84**, 1401–1421 (2014).

40. Pavlov, D. S. & Kasumyan, A. O. Patterns and mechanisms of schooling behaviour in fish: A review. *J. Ichthyol.* **40**, 163–231 (2000).

41. Fish, F. E., Fegely, J. F. & Xanthopoulos, C. J. Burst-and-coast swimming in schooling fish (Notemigonus crysoleucas) with implications for energy economy. *Comp. Biochem. Physiol. -- Part A Physiol.* **100**, 633–637 (1991).

42. Herskin, J. & Steffensen, J. F. Energy savings in sea bass swimming in a school: measurements of tail beat frequency and oxygen consumption at different swimming speeds. *J. Fish Biol.* **53**, 366–376 (1998).

43. Svendsen, J. C., Skov, J., Bildsoe, M. & Steffensen, J. F. Intra-school positional preference and reduced tail beat frequency in trailing positions in schooling roach under experimental conditions. *J. Fish Biol.* **62**, 834–846 (2003).

44. Marras, S. *et al.* Fish swimming in schools save energy regardless of their spatial position. *Behav. Ecol. Sociobiol.* **69**, 19–226 (2015).

45. Weihs, D. Hydromechanics of fish schooling. *Nature* **241**, 290–291 (1973).

46. Gazzola, M., Tchieu, A. A., Alexeev, D., De Brauer, A. & Koumoutsakos, P. Learning to school in the presence of hydrodynamic interactions. *J. Fluid Mech.* **789**, 726–749 (2016).

47. Verma, S., Novati, G. & Koumoutsakos, P. Efficient collective swimming by harnessing vortices through deep reinforcement learning. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 5849–5854 (2018).

48. Filella, A., Nadal, F., Sire, C., Kanso, E. & Eloy, C. Model of Collective Fish Behavior with Hydrodynamic Interactions. *Phys. Rev. Lett.* **120**, 35–37 (2018).

49. Ashraf, I., Godoy-Diana, R., Halloy, J., Collignon, B. & Thiria, B. Synchronization and collective swimming patterns in fish (Hemigrammus bleheri). *J. R. Soc. Interface* **13**, (2016).

50. Li, L. *et al.* Vortex phase matching as a strategy for schooling in robots and in fish. *Nat. Commun.* **11**, 1–9 (2020).

51. De Bie, J., Manes, C. & Kemp, P. S. Collective behaviour of fish in the presence and absence of flow. *Anim. Behav.* **167**, 151–159 (2020).

52. Ashraf, I. *et al.* Simple phalanx pattern leads to energy saving in cohesive fish schooling. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 9599–9604 (2017).

53. Milner-Gulland, E. J., Fryxell, J. M. & Sinclair, A. R. E. *Animal Migration: A Synthesis*. (OUP Oxford, 2011).

54. Couzin, I. D. Collective animal migration. *Curr. Biol.* **28**, R976–R980 (2018).

55. Berdahl, A., Torney, C. J., Ioannou, C. C., Faria, J. J. & Couzin, I. D. Emergent sensing of complex environments by mobile animal groups. *Science (80-.).* **339**, 574–576 (2013).

56. Berdahl, A., Westley, P. A. H. & Quinn, T. P. Social interactions shape the timing of spawning migrations in an anadromous fish. *Anim. Behav.* **126**, 221–229 (2017).

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

58. Simons, A. M. Many wrongs: The advantage of group navigation. *Trends Ecol. Evol.* **19**, 453–455 (2004).

59. Codling, E. A., Pitchford, J. W. & Simpson, S. D. Group navigation and the 'many-wrongs principle' in models of animal movement. *Ecology* **88**, 1864–1870 (2007).

60. Bode, N. W. F. *et al.* Distinguishing social from nonsocial navigation in moving animal groups. *Am. Nat.* **179**, 621–632 (2012).

61. Huse, G., Railsback, S. & Feronö, A. Modelling changes in migration pattern of herring: collective behaviour and numerical domination. *J. Fish Biol.* **60**, 571–582 (2002).

62. Fagan, W. F., Cantrell, R. S., Cosner, C., Mueller, T. & Noble, A. E. Leadership, social learning, and the maintenance (or collapse) of migratory populations. *Theor. Ecol.* **5**, 253–264 (2012).

63. De Luca, G., Mariani, P., MacKenzie, B. R. & Marsili, M. Fishing out collective memory of migratory schools. *J. R. Soc. Interface* **11**, 20140043 (2014).

64. Kao, A. B., Miller, N., Torney, C., Hartnett, A. & Couzin, I. D. Collective learning and optimal consensus decisions in social animal groups. *PLoS Comput. Biol.* **10**, e1003762 (2014).

65. Berdahl, A. *et al.* Collective animal navigation and migratory culture: From theoretical models to empirical evidence. *Philos. Trans. R. Soc. B Biol. Sci.* **373**,

(2018).

66. Goodwin, R. A. *et al.* Fish navigation of large dams emerges from their modulation of flow field experience. *Proc. Natl. Acad. Sci.* **111**, 5277–5282 (2014).

67. Okasaki, C., Keefer, M. L., Westley, P. A. H. & Berdahl, A. Collective navigation can facilitate passage through human-made barriers by homeward migrating Pacific salmon: Collective navigation improves passage. *Proc. R. Soc. B Biol. Sci.* **287**, (2020).

68. Berdahl, A., Westley, P. A. H., Levin, S. A., Couzin, I. D. & Quinn, T. P. A collective navigation hypothesis for homeward migration in anadromous salmonids. *Fish Fish.* **17**, 525–542 (2016).

69. Lehtonen, J. & Jaatinen, K. Safety in numbers: the dilution effect and other drivers of group life in the face of danger. *Behav. Ecol. Sociobiol.* **70**, 449–458 (2016).

70. Turner, G. F. & Pitcher, T. J. Attack abatement: a model for group protection by combined avoidance and dilution. *Am. Nat.* **128**, 228–240 (1986).

71. Morgan, M. J. & Godin, J. J. Antipredator benefits of schooling behaviour in a cyprinodontid fish, the banded killifish (Fundulus diaphanus). *Z. Tierpsychol.* **70**, 236–246 (1985).

72. Wrona, F. J. & Dixon, R. W. J. Group size and predation risk: a field analysis of encounter and dilution effects. *Am. Nat.* **137**, 186–201 (1991).

73. Fernandes, M. de O. & Volpato, G. L. Heterogeneous growth in the Nile tilapia: Social stress and carbohydrate metabolism. *Physiol. Behav.* **54**, 319–323 (1993).

74. Sloman, K. A., Motherwell, G., O'connor, K. I. & Taylor, A. C. The effect of social stress on the Standard Metabolic Rate (SMR) of brown trout, Salmo trutta. *Fish Physiol. Biochem.* **23**, 49–53 (2000).

75. Sloman, K. A., Metcalfe, N. B., Taylor, A. C. & Gilmour, K. M. Plasma cortisol concentrations before and after social stress in rainbow trout and brown trout. *Physiol. Biochem. Zool.* **74**, 383–389 (2001).

76. Ejike, C. & Schreck, C. B. Stress and Social Hierarchy Rank in Coho Salmon. *Trans. Am. Fish. Soc.* **109**, 423–426 (1980).

77. Nadler, L. E., Killen, S. S., McClure, E. C., Munday, P. L. & McCormick, M. I. Shoaling reduces metabolic rate in a gregarious coral reef fish species. *J. Exp. Biol.* **219**, 2802–2805 (2016).

78. Lefrançois, C., Ferrari, R. S., Moreira Da Silva, J. & Domenici, P. The effect of progressive hypoxia on spontaneous activity in single and shoaling golden grey mullet Liza aurata. *J. Fish Biol.* **75**, 1615–1625 (2009).

79. Queiroz, H. & Magurran, A. E. Safety in numbers? Shoaling behaviour of the Amazonian red-bellied piranha. *Biol. Lett.* **1**, 155–157 (2005).

80. Schleuter, D., Haertel-Borer, S., Fischer, P. & Eckmann, R. Respiration rates of Eurasian perch Perca fluviatilis and ruffe: lower energy costs in groups. *Trans. Am. Fish. Soc.* **136**, 43–55 (2007).

81. Parker, F. R. Reduced Metabolic Rates in Fishes as a Result of Induced Schooling. *Trans. Am. Fish. Soc.* **102**, 125–131 (1973).

82. Culbert, B. M., Gilmour, K. M. & Balshine, S. Social buffering of stress in a group-living fish. *Proc. R. Soc. B Biol. Sci.* **286**, 20191626 (2019).

83. Allen, P. J., Barth, C. C., Peake, S. J., Abrahams, M. V. & Anderson, W. G. Cohesive social behaviour shortens the stress response: The effects of conspecifics on the stress response in lake sturgeon Acipenser fulvescens. *J. Fish Biol.* **74**, 90–104 (2009).

84. Mushtaq, M. Y. *et al.* Effect of acute stresses on zebra fish (Danio rerio) metabolome measured by NMR-based metabolomics. *Planta Med.* **80**, 1227–1233 (2014).

85. Gaikwad, S. *et al.* Acute stress disrupts performance of zebrafish in the cued and spatial memory tests: The utility of fish models to study stress–memory interplay. *Behav. Processes* **87**, 224–230 (2011).

86. Otsuka, A., Shimomura, Y., Sakikubo, H., Miura, K. & Kagawa, N. Effects of single and repeated heat stress on anxiety-like behavior and locomotor activity in medaka fish. *Fish. Sci.* 1–10 (2022).

87. Gregory, T. R. & Wood, C. M. The effects of chronic plasma cortisol elevation on the feeding behaviour, growth, competitive ability, and swimming performance of juvenile rainbow trout. *Physiol. Biochem. Zool.* **72**, 286–295 (1999).

88. Wilson, D. S., Coleman, K., Clark, A. B. & Biederman, L. Shy-bold continuum in pumpkinseed sunfish (Lepomis gibbosus): an ecological study of a psychological trait. *J. Comp. Psychol.* **107**, 250–260 (1993).

89. Wolf, M. & Weissing, F. J. Animal personalities: consequences for ecology and evolution. *Trends Ecol. Evol.* **27**, 452–461 (2012).

90. Toms, C. N., Echevarria, D. J. & Jouandot, D. J. A Methodological Review of Personality-Related Studies in Fish: Focus on the Shy-Bold Axis of Behavior. *Int. J. Comp. Psychol.* **23**, (2010).

91. Hasenjager, M. J., Hoppitt, W. & Dugatkin, L. A. Personality composition determines social learning pathways within shoaling fish: Group personality and social learning. *Proc. R. Soc. B Biol. Sci.* **287**, (2020).

92. Jolles, J. W., Boogert, N. J., Sridhar, V. H., Couzin, I. D. & Manica, A. Consistent Individual Differences Drive Collective Behavior and Group Functioning of Schooling Fish. *Curr. Biol.* **27**, 2862-2868.e7 (2017).

93. Brown, G. E., Bongiorno, T., Dicapua, D. M., Ivan, L. I. & Roh, E. Effects of group size on the threat-sensitive response to varying concentrations of chemical alarm cues by juvenile convict cichlids 1. (2005) doi:10.1139/Z05-166.

94. Chapman, B. B., Morrell, L. J., Krause, J., Biology, C. & Fisheries, I. Unpredictability in food supply during early life influences boldness in fish. (2010) doi:10.1093/beheco/arq003.

95. Kareklas, K., Arnott, G., Elwood, R. W. & Holland, R. A. Plasticity varies with boldness in a weakly- electric fish. *Front. Zool.* 1–7 (2016) doi:10.1186/s12983-016-0154-0.

96. Klefoth, T., Skov, C., Krause, J. & Arlinghaus, R. The role of ecological context and predation risk-stimuli in revealing the true picture about the genetic basis of boldness evolution in fish. 547–559 (2012) doi:10.1007/s00265-011-1303-2.

97. Roy, T. & Bhat, A. Repeatability in boldness and aggression among wild zebrafish (Danio rerio) from two differing predation and flow regimes. *J. Comp. Psychol.* **132**, 349 (2018).

98. Tang, Z. & Fu, S. J. Effects of habitat conditions on the boldness and sociability of wild-caught fish (Zacco platypus) along a river. *J. Ethol.* **39**, 379–391 (2021).

99. Galhardo, L., Vitorino, A. & Oliveira, R. F. Social familiarity modulates personality trait in a cichlid fish. *Biol. Lett.* **8**, 936–938 (2012).

100. Ward, A. J. W. Social facilitation of exploration in mosquitofish (Gambusia holbrooki). *Behav. Ecol. Sociobiol.* **66**, 223–230 (2012).

101. Magnhagen, C. & Bunnefeld, N. Express your personality or go along with the group: What determines the behaviour of shoaling perch? *Proc. R. Soc. B Biol. Sci.* **276**, 3369–3375 (2009).

102. Cote, J., Fogarty, S., Brodin, T., Weinersmith, K. & Sih, A. Personality-

dependent dispersal in the invasive mosquitofish: Group composition matters. *Proc. R. Soc. B Biol. Sci.* **278**, 1670–1678 (2011).

103. Albayrak, I., Boes, R. M., Kriewitz-Byun, C. R., Peter, A. & Tullis, B. P. Fish guidance structures: Hydraulic performance and fish guidance efficiencies. *J. Ecohydraulics* **5**, 113–131 (2020).

104. Ruggerone, G. T. Consumption of migrating juvenile salmonids by gulls foraging below a Columbia River dam. *Trans. Am. Fish. Soc.* **115**, 736–742 (1986).

105. Schilt, C. R. Developing fish passage and protection at hydropower dams. *Appl. Anim. Behav. Sci.* **104**, 295–325 (2007).

106. Gowans, A. R. D., Armstrong, J. D., Priede, I. G. & Mckelvey, S. Movements of Atlantic salmon migrating upstream through a fish-pass complex in Scotland. *Ecol. Freshw. Fish* **12**, 177–189 (2003).

107. Petersen, J. H., Gadomski, D. M. & Poe, T. P. Differential predation by northern squawfish (Ptychocheilus oregonensis) on live and dead juvenile salmonids in the Bonneville Dam Tailrace (Columbia River). *Can. J. Fish. Aquat. Sci.* **51**, 1197–1204 (1994).

108. Lemasson, B. H., Haefner, J. W. & Bowen, M. D. Schooling increases risk exposure for fish navigating past artificial barriers. *PLoS One* **9**, (2014).

109. Landeau, L. & Terborgh, J. Oddity and the 'confusion effect'in predation. *Anim. Behav.* **34**, 1372–1380 (1986).

110. Pitcher, T. J., Green, D. A. & Magurran, A. E. Dicing with death: predator inspection behaviour in minnow shoals. *J. Fish Biol.* **28**, 439–448 (1986).

111. Kelley, J. L. & Magurran, A. E. Learned predator recognition and antipredator responses in fishes. *Fish Fish.* **4**, 216–226 (2003).

112. Mccormick, M. I. & Manassa, R. Predation risk assessment by olfactory and visual cues in a coral reef fish. *Coral Reefs* **27**, 105–113 (2008).

113. Smith, R. J. F. Alarm signals in fishes. *Rev. Fish Biol. Fish.* **2**, 33–63 (1992).

114. Agostinho, A. A., Agostinho, C. S., Pelicice, F. M. & Marques, E. E. Fish ladders: Safe fish passage or hotspot for predation? *Neotrop. Ichthyol.* **10**, 687–696 (2012).

115. Brown, C. & Laland, K. N. Social learning in fishes : a review. (2003).

116. Lamba, A., Cassey, P., Segaran, R. R. & Koh, L. P. Deep learning for

environmental conservation. *Curr. Biol.* **29**, R977–R982 (2019).

117. Christin, S., Hervet, É. & Lecomte, N. Applications for deep learning in ecology. *Methods Ecol. Evol.* **10**, 1632–1644 (2019).

118. Barreiros, M. de O., Dantas, D. de O., Silva, L. C. de O., Ribeiro, S. & Barros, A. K. Zebrafish tracking using YOLOv2 and Kalman filter. *Sci. Rep.* **11**, 1–14 (2021).

119. Lopez-Marcano, S. *et al.* Automatic detection of fish and tracking of movement for ecology. *Ecol. Evol.* **11**, 8254–8263 (2021).

120. Huang, Z. J., He, X. X., Wang, F. J. & Shen, Q. A Real-Time Multi-Stage Architecture for Pose Estimation of Zebrafish Head with Convolutional Neural Networks. *J. Comput. Sci. Technol.* **36**, 434–444 (2021).

121. Pereira, E. *et al.* Temporal patterns of the catadromous thinlip grey mullet migration in freshwater. *Ecohydrology* **14**, e2345 (2021).

122. Goodwin, R. A., Nestler, J. M., Anderson, J. J., Weber, L. J. & Loucks, D. P. Forecasting 3-D fish movement behavior using a Eulerian–Lagrangian–agent method (ELAM). *Ecol. Modell.* **192**, 197–223 (2006).

123. Caravaggi, A. *et al.* A review of camera trapping for conservation behaviour research. *Remote Sens. Ecol. Conserv.* **3**, 109–122 (2017).

124. Nelson, X. J. & Fijn, N. The use of visual media as a tool for investigating animal behaviour. *Anim. Behav.* **85**, 525–536 (2013).

125. Wratten, S. *Video Techniques in Animal Ecology and Behaviour*. (Springer Science & Business Media, 1993).

126. Leibe, B. & Schiele, B. Analyzing appearance and contour based methods for object categorization. in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.* vol. 2 II–409 (IEEE, 2003).

127. Kalinke, T., Tzomakas, C. & von Seelen, W. A texture-based object detection and an adaptive model-based classification. in *Procs. IEEE Intelligent Vehicles Symposium* vol. 98 341–346 (Citeseer, 1998).

128. Li, Y., Wang, S., Tian, Q. & Ding, X. Feature representation for statistical-learning-based object detection: A review. *Pattern Recognit.* **48**, 3542–3559 (2015).

129. Lau, K.-K. *et al.* An edge-detection approach to investigating pigeon navigation. *J. Theor. Biol.* **239**, 71–78 (2006).

130. Malik, J., Belongie, S., Leung, T. & Shi, J. Contour and texture analysis for image segmentation. *Int. J. Comput. Vis.* **43**, 7–27 (2001).

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

132. Couzin, I. D. & Heins, C. Emerging technologies for behavioral research in changing environments. *Trends Ecol. Evol.* **38**, 346–354 (2023).

133. Mahapatra, D. Analyzing training information from random forests for improved image segmentation. *IEEE Trans. Image Process.* **23**, 1504–1512 (2014).

134. Tzotsos, A. & Argialas, D. Support vector machine classification for object-based image analysis. *Object-based image Anal. Spat. concepts knowledge-driven Remote Sens. Appl.* 663–677 (2008).

135. Xiao, Y. *et al.* A review of object detection based on deep learning. *Multimed. Tools Appl.* **79**, 23729–23791 (2020).

136. Lauer, J. *et al.* Multi-animal pose estimation, identification and tracking with DeepLabCut. *Nat. Methods* **19**, 496–504 (2022).

137. Poudel, R. P. K., Liwicki, S. & Cipolla, R. Fast-scnn: Fast semantic segmentation network. *arXiv Prepr. arXiv1902.04502* (2019).

138. Liu, T. & Stathaki, T. Faster R-CNN for robust pedestrian detection using semantic segmentation network. *Front. Neurorobot.* **12**, 64 (2018).

139. Anantharaman, R., Velazquez, M. & Lee, Y. Utilizing mask R-CNN for detection and segmentation of oral diseases. in *2018 IEEE international conference on bioinformatics and biomedicine (BIBM)* 2197–2204 (IEEE, 2018).

140. Liu, W. *et al.* Ssd: Single shot multibox detector. in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14* 21–37 (Springer, 2016).

141. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. in *Proceedings of the IEEE conference on computer vision and pattern recognition* 779–788 (2016).

142. Wang, C.-Y., Yeh, I.-H. & Liao, H.-Y. M. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *arXiv Prepr. arXiv2402.13616* (2024).

143. Terven, J., Córdova-Esparza, D.-M. & Romero-González, J.-A. A comprehensive

review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Mach. Learn. Knowl. Extr.* **5**, 1680–1716 (2023).

144. Kalman, R. E. A new approach to linear filtering and prediction problems. (1960).

145. Jiang, P., Ergu, D., Liu, F., Cai, Y. & Ma, B. A Review of Yolo algorithm developments. *Procedia Comput. Sci.* **199**, 1066–1073 (2022).

146. Chicco, D. & Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* **21**, 1–13 (2020).

147. Hong, S.-J., Han, Y., Kim, S.-Y., Lee, A.-Y. & Kim, G. Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery. *Sensors* **19**, 1651 (2019).

148. The MathWorks Inc. MATLAB version: 9.13.0 (R2022b). at https://www.mathworks.com (2022).

149. Guerin, B. *Social Facilitation.* (Editions de la Maison des Sciences de l'Homme, 1993).

150. Uzsák, A., Dieffenderfer, J., Bozkurt, A. & Schal, C. Social facilitation of insect reproduction with motor-driven tactile stimuli. *Proc. R. Soc. B Biol. Sci.* **281**, 20140325 (2014).

151. Pays, O. *et al.* The effect of social facilitation on vigilance in the eastern gray kangaroo, Macropus giganteus. *Behav. Ecol.* **20**, 469–477 (2009).

152. Jackson, A. L., Ruxton, G. D. & Houston, D. C. The effect of social facilitation on foraging success in vultures: A modelling study. *Biol. Lett.* **4**, 311–313 (2008).

153. Ryer, C. H. & Olla, B. L. Social mechanisms facilitating exploitation of spatially variable ephemeral food patches in a pelagic marine fish. *Anim. Behav.* **44**, 69–74 (1992).

154. Baird, T. A., Ryer, C. H. & Olla, B. L. Social enhancement of foraging on an ephemeral food source in juvenile walleye pollock, Theragra chalcogramma. *Environ. Biol. Fishes* **31**, 307–311 (1991).

155. Suboski, M. D. *et al.* Alarm reaction in acquisition and social transmission of simulated-predator recognition by zebra danio fish (Brachydanio rerio). *J. Comp. Psychol.* **104**, 101 (1990).

156. Karplus, I., Katzenstein, R. & Goren, M. Predator recognition and social facilitation of predator avoidance in coral reef fish Dascyllus marginatus

juveniles. *Mar. Ecol. Prog. Ser.* **319**, 215–223 (2006).

157. Magurran, A. E. & Pitcher, T. J. Foraging, timidity and shoal size in minnows and goldfish. *Behav. Ecol. Sociobiol.* **12**, 147–152 (1983).

158. Webster, M. M., Ward, A. J. W. & Hart, P. J. B. Boldness is influenced by social context in threespine sticklebacks (Gasterosteus aculeatus). *Behaviour* **144**, 351 (2007).

159. Kent, M. I. A., Lukeman, R., Lizier, J. T. & Ward, A. J. W. Speed-mediated properties of schooling. *R. Soc. Open Sci.* **6**, 181482 (2019).

160. Réale, D., Reader, S. M., Sol, D., McDougall, P. T. & Dingemanse, N. J. Integrating animal temperament within ecology and evolution. *Biol. Rev.* **82**, 291–318 (2007).

161. Palmer, M. & Ruhi, A. Linkages between flow regime, biota, and ecosystem processes: Implications for river restoration. *Science (80-.).* **365**, eaaw2087 (2019).

162. Schiavon, A. *et al.* Survival and swimming performance of a small-sized Cypriniformes (Telestes muticellus) tagged with passive integrated transponders. *J. Limnol.* **82**, (2023).

163. Ashraf, M. U., Nyqvist, D., Comoglio, C. & Manes, C. The effect of in-flume habituation time and fish behaviour on estimated swimming performance. *J. Ecohydraulics* 1–9 (2024).

164. Schumann, S. *et al.* Social buffering of oxidative stress and cortisol in an endemic cyprinid fish. *Sci. Rep.* **13**, 20579 (2023).

165. Nyqvist, D. *et al.* PIT-tagging Italian spined loach (Cobitis bilineata)–methodology, survival, and behavioral effects. *J. Fish Biol.* (2022).

166. R Core Team. R: A Language and Environment for Statistical Computing. at https://www.r-project.org/ (2022).

167. Silverman, B. W. *Density Estimation for Statistics and Data Analysis*. vol. 26 (CRC press, 1986).

168. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

169. Kassambara, A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. at https://rpkgs.datanovia.com/rstatix/ (2023).

170. Terpstra, T. J. The asymptotic normality and consistency of kendall's test against trend, when ties are present in one ranking. *Indag. Math.* **55**, 327–333 (1952).

171. Jonckheere, A. R. A Distribution-Free k-Sample Test Against Ordered Alternatives. *Biometrika* **41**, 133 (1954).

172. Kruskal, W. H. & Wallis, W. A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **47**, 583–621 (1952).

173. Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Am. Stat. Assoc.* **32**, 675–701 (1937).

174. Wilcoxon, F. Individual comparisons by ranking methods. in *Breakthroughs in Statistics: Methodology and Distribution* 196–202 (Springer, 1992).

175. Bonferroni, C. Teoria statistica delle classi e calcolo delle probabilita. *Pubbl. del R Ist. Super. di Sci. Econ. e Commericiali di Firenze* **8**, 3–62 (1936).

176. Mikheev, V. N. & Andreev, O. A. Two-phase exploration of a novel environment in the guppy, Poecilia reticulata. *J. Fish Biol.* **42**, 375–383 (1993).

177. Ashraf, M. U. *et al.* Fish swimming performance: effect of flume length and different fatigue definitions. in *Advances in Hydraulic Research* (GeoPlanet: Earth and Planetary Sciences IN PRESS, 2024). doi:doi.org/10.1007/978-3-031-56093-4_1.

178. Vezza, P. *et al.* Rethinking swimming performance tests for bottom-dwelling fish: the case of European glass eel (Anguilla anguilla). *Sci. Rep.* **10**, 16416 (2020).

179. Kerr, J. R., Manes, C. & Kemp, P. S. Assessing hydrodynamic space use of brown trout, Salmo trutta, in a complex flow environment: a return to first principles. *J. Exp. Biol.* **219**, 3480–3491 (2016).

180. Zelnik, P. R. & Goldspink, G. The effect of exercise on plasma cortisol and blood sugar levels in the rainbow trout, Salmo gairdnerii Richardson. *J. Fish Biol.* **19**, 37–43 (1981).

181. Li, X. *et al.* Effect of flow velocity on the growth, stress and immune responses of turbot (Scophthalmus maximus) in recirculating aquaculture systems. *Fish Shellfish Immunol.* **86**, 1169–1176 (2019).

182. Krause, J. & Ruxton, G. D. *Living in Groups*. (OUP Oxford, 2002).

183. Killen, S. S., Marras, S., Steffensen, J. F. & Mckenzie, D. J. Aerobic capacity influences the spatial position of individuals within fish schools. *Proc. R. Soc. B*

Biol. Sci. **279**, 357–364 (2012).

184. Liao, J. C. A review of fish swimming mechanics and behaviour in altered flows. *Philos. Trans. R. Soc. B Biol. Sci.* **362**, 1973–1993 (2007).

185. Videler, J. J. & Wardle, C. S. Fish swimming stride by stride: speed limits and endurance. *Rev. Fish Biol. Fish.* **1**, 23–40 (1991).

186. Cano-Barbacil, C. *et al.* Key factors explaining critical swimming speed in freshwater fish: a review and statistical analysis for Iberian species. *Sci. Rep.* **10**, 18947 (2020).

187. Hoar, W. S., Randall, D. J. & Donaldson, E. M. *Fish Physiology*. (Academic Press, 1983).

188. Des Roches, S. *et al.* The ecological importance of intraspecific variation. *Nat. Ecol. Evol.* **2**, 57–64 (2018).

189. Freeman, M. C., Bowen, Z. H., Bovee, K. D. & Irwin, E. R. Flow and habitat effects on juvenile fish abundance in natural and altered flow regimes. *Ecol. Appl.* **11**, 179–190 (2001).

190. Kuriqi, A., Pinheiro, A. N., Sordo-Ward, A. & Garrote, L. Influence of hydrologically based environmental flow methods on flow alteration and energy production in a run-of-river hydropower plant. *J. Clean. Prod.* **232**, 1028–1042 (2019).

191. Jowett, I. G., Richardson, J. & Bonnett, M. L. Relationship between flow regime and fish abundances in a gravel-bed river, New Zealand. *J. Fish Biol.* **66**, 1419–1436 (2005).

192. Hockley, F. A., Wilson, C. A. M. E., Brew, A. & Cable, J. Fish responses to flow velocity and turbulence in relation to size, sex and parasite load. *J. R. Soc. Interface* **11**, 20130814 (2014).

193. Silva, A. T., Katopodis, C., Santos, J. M., Ferreira, M. T. & Pinheiro, A. N. Cyprinid swimming behaviour in response to turbulent flow. *Ecol. Eng.* **44**, 314–328 (2012).

194. Quaranta, E., Katopodis, C., Revelli, R. & Comoglio, C. Turbulent flow field comparison and related suitability for fish passage of a standard and a simplified low-gradient vertical slot fishway. *River Res. Appl.* **33**, 1295–1305 (2017).

195. Closs, G. P., Krkosek, M. & Olden, J. D. *Conservation of Freshwater Fishes*. (Cambridge University Press, 2016).

196. Couzin, I. D. & Krause, J. Self-Organization and Collective Behavior in Vertebrates. *Adv. Study Behav.* **32**, 1–75 (2003).

197. Calovi, D. S. *et al.* Disentangling and modeling interactions in fish with burst-and-coast swimming reveal distinct alignment and attraction behaviors. *PLoS Comput. Biol.* **14**, e1005933 (2018).

198. Puckett, J. G., Pokhrel, A. R. & Giannini, J. A. Collective gradient sensing in fish schools. *Sci. Rep.* **8**, 7587 (2018).

199. Hockley, F. A., Wilson, C. A. M. E., Graham, N. & Cable, J. Combined effects of flow condition and parasitism on shoaling behaviour of female guppies Poecilia reticulata. *Behav. Ecol. Sociobiol.* **68**, 1513–1520 (2014).

200. Shelton, D. S. *et al.* Collective behavior in wild zebrafish. *Zebrafish* **17**, 243–252 (2020).

201. Allouche, S. & Gaudin, P. Effects of avian predation threat, water flow and cover on growth and habitat use by chub, Leuciscus cephalus, in an experimental stream. *Oikos* **94**, 481–492 (2001).

202. Hensor, E., Couzin, I. D., James, R. & Krause, J. Modelling density-dependent fish shoal distributions in the laboratory and field. *Oikos* **110**, 344–352 (2005).

203. Magurran, A. E. & Pitcher, T. J. Provenance, shoal size and the sociobiology of predator-evasion behaviour in minnow shoals. *Proc. R. Soc. London. Ser. B. Biol. Sci.* **229**, 439–465 (1987).

204. Lombana, D. A. B. & Porfiri, M. Collective response of fish to combined manipulations of illumination and flow. *Behav. Processes* **203**, 104767 (2022).

205. Duarte, B. A. de F. & Ramos, I. C. R. Reynolds shear-stress and velocity: positive biological response of neotropical fishes to hydraulic parameters in a vertical slot fishway. *Neotrop. Ichthyol.* **10**, 813–819 (2012).

206. Silva, A. T., Santos, J. M., Ferreira, M. T., Pinheiro, A. N. & Katopodis, C. Effects of water velocity and turbulence on the behaviour of Iberian barbel (Luciobarbus bocagei, Steindachner 1864) in an experimental pool-type fishway. *River Res. Appl.* **27**, 360–373 (2011).

207. Morán-López, R. & Tolosa, O. U. Obstacle negotiation attempts by leaping cyprinids indicate bank-side spawning migration routes. *Fish. Res.* **197**, 84–87 (2018).

208. Jellyman, D. J. & Sykes, J. R. E. Diel and seasonal movements of radio-tagged freshwater eels, Anguilla spp., in two New Zealand streams. *Environ. Biol. Fishes*

66, 143–154 (2003).

209. Hughes, N. F. The wave-drag hypothesis: an explanation for size-based lateral segregation during the upstream migration of salmonids. *Can. J. Fish. Aquat. Sci.* **61**, 103–109 (2004).

210. Powers, P. D. & Orsborn, J. F. Analysis of barriers to upstream fish migration. at (1984).

211. Herbert-Read, J. E. *et al.* Inferring the rules of interaction of shoaling fish. *Proc. Natl. Acad. Sci.* **108**, 18726–18731 (2011).

212. Schnörr, S. J., Steenbergen, P. J., Richardson, M. K. & Champagne, D. Measuring thigmotaxis in larval zebrafish. *Behav. Brain Res.* **228**, 367–374 (2012).

213. Ward, A. J. W., Sumpter, D. J. T., Couzin, I. D., Hart, P. J. B. & Krause, J. Quorum decision-making facilitates information transfer in fish shoals. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 6948–6953 (2008).

214. Cavagna, A. *et al.* Scale-free correlations in starling flocks. *Proc. Natl. Acad. Sci.* **107**, 11865–11870 (2010).

215. Pearson, K. VII. Note on regression and inheritance in the case of two parents. *Proc. R. Soc. London* **58**, 240–242 (1895).

216. Garner, P. Effects of variable discharge on the velocity use and shoaling behavior of Phoxinus phoxinus. *J. Fish Biol.* **50**, 1214–1220 (1997).

217. Fish, F. E. Swimming strategies for energy economy. *Fish Locomot. An Eco-ethological Perspect.* 102–134 (2010) doi:10.1201/b10190.

218. Fausch, K. D. Profitable stream positions for salmonids: relating specific growth rate to net energy gain. *Can. J. Zool.* **62**, 441–451 (1984).

219. Hughes, N. F. Selection of positions by drift-feeding salmonids in dominance hierarchies: model and test for Arctic grayling (Thymallus arcticus) in subarctic mountain streams, interior Alaska. *Can. J. Fish. Aquat. Sci.* **49**, 1999–2008 (1992).

220. Grossman, G. D., Rincon, P. A., Farr, M. D. & Ratajczak, R. E. A new optimal foraging model predicts habitat use by drift-feeding stream minnows. *Ecol. Freshw. Fish* **11**, 2–10 (2002).

221. Di Santo, V., Kenaley, C. P. & Lauder, G. V. High postural costs and anaerobic metabolism during swimming support the hypothesis of a U-shaped metabolism–

speed curve in fishes. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 13048–13053 (2017).

222. Przybilla, A., Kunze, S., Rudert, A., Bleckmann, H. & Brücker, C. Entraining in trout: A behavioural and hydrodynamic analysis. *J. Exp. Biol.* **213**, 2976–2986 (2010).

223. Beckmann, M., Erős, T., Schmitz, A. & Bleckmann, H. Number and distribution of superficial neuromasts in twelve common European cypriniform fishes and their relationship to habitat occurrence. *Int. Rev. Hydrobiol.* **95**, 273–284 (2010).

224. Bleckmann, H. & Zelick, R. Lateral line system of fish. *Integr. Zool.* **4**, 13–25 (2009).

225. Pottinger, T. G. The stress response in fish-mechanisms, effects and measurement. *Fish Welf.* 32–48 (2008).

226. Mawer, R. *et al.* Individual based models for the simulation of fish movement near barriers: Current work and future directions. *J. Environ. Manage.* **335**, 117538 (2023).

227. Sumpter, D., Buhl, J., Biro, D. & Couzin, I. Information transfer in moving animal groups. *Theory Biosci.* **127**, 177–186 (2008).

228. Couzin, I. D., James, R., Mawdsley, D., Croft, D. P. & Krause, J. Social Organization and Information Transfer in Schooling Fishes. *Fish Cogn. Behav.* **2**, 166–185 (2007).

229. Porfiri, M. Inferring causal relationships in zebrafish-robot interactions through transfer entropy: a small lure to catch a big fish. *Anim. Behav. Cogn.* **5**, 341–367 (2018).

230. Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).

231. Whetten, A. B. Localized Mutual Information Monitoring of Pairwise Associations in Animal Movement. *arXiv Prepr. arXiv2111.10628* (2021).

232. Owoeye, K., Musolesi, M. & Hailes, S. Characterizing animal movement patterns across different scales and habitats using information theory. *bioRxiv* 311241 (2018).

233. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **85**, 461–464 (2000).

234. Encel, S. A., Schaerf, T. M., Lizier, J. T. & Ward, A. J. W. Locomotion, interactions and information transfer vary according to context in a cryptic fish

species. *Behav. Ecol. Sociobiol.* **75**, 1–12 (2021).

235. Ward, A. J. W. *et al.* Cohesion, order and information flow in the collective motion of mixed-species shoals. *R. Soc. open Sci.* **5**, 181132 (2018).

236. Crosato, E. *et al.* Informative and misinformative interactions in a school of fish. *Swarm Intell.* **12**, 283–305 (2018).

237. Butail, S., Mwaffo, V. & Porfiri, M. Model-free information-theoretic approach to infer leadership in pairs of zebrafish. *Phys. Rev. E* **93**, 42411 (2016).

238. Butail, S., Ladu, F., Spinello, D. & Porfiri, M. Information flow in animal-robot interactions. *Entropy* **16**, 1315–1330 (2014).

239. Hu, F., Nie, L. J. & Fu, S. J. Information dynamics in the interaction between a prey and a predator fish. *Entropy* **17**, 7230–7241 (2015).

240. Lizier, J. T. JIDT: an information-theoretic toolkit for studying the dynamics of complex systems. *Front. Robot. AI* **1**, 11 (2014).

241. Kaiser, A. & Schreiber, T. Information transfer in continuous processes. *Phys. D Nonlinear Phenom.* **166**, 43–62 (2002).

242. Kantz, H. & Schreiber, T. *Nonlinear Time Series Analysis*. vol. 7 (Cambridge university press, 2004).

243. Kraskov, A., Stögbauer, H. & Grassberger, P. Estimating mutual information. *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* **69**, 16 (2004).

244. Staniek, M. & Lehnertz, K. Symbolic transfer entropy. *Phys. Rev. Lett.* **100**, 158101 (2008).

245. Porfiri, M. & Marín, M. R. Symbolic dynamics of animal interaction. *J. Theor. Biol.* **435**, 145–156 (2017).

246. Sensoy, A., Sobaci, C., Sensoy, S. & Alali, F. Effective transfer entropy approach to information flow between exchange rates and stock markets. *Chaos, Solitons and Fractals* **68**, 180–185 (2014).

Appendix

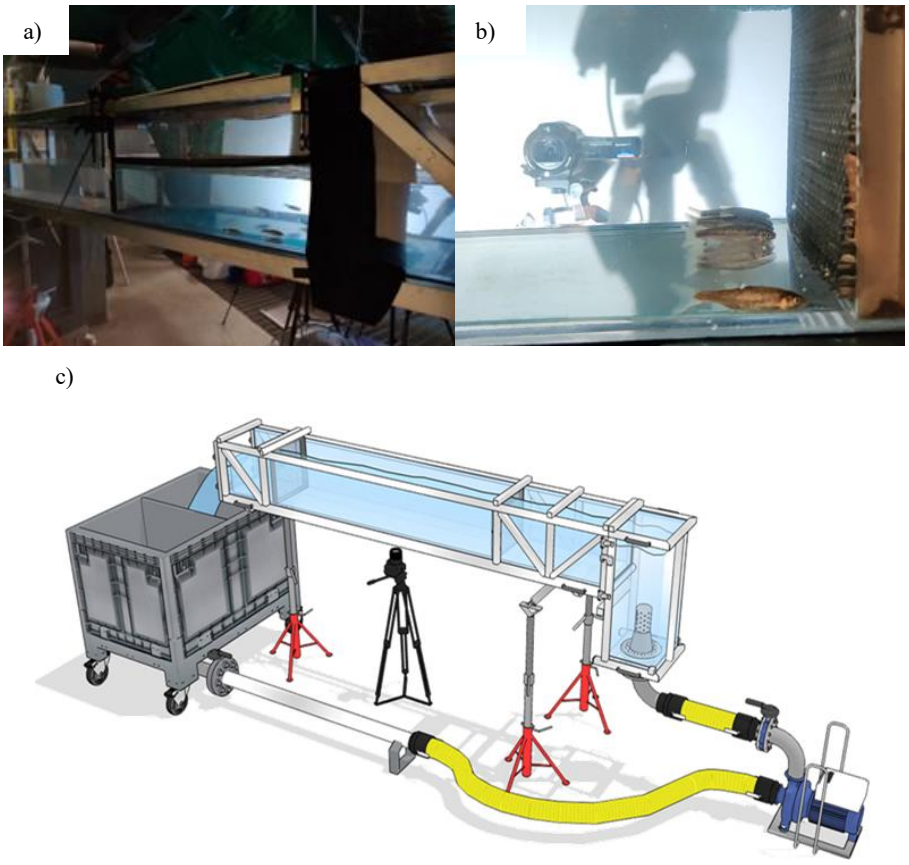

Figure S 1: Experimental apparatus. Picture of the flume in place with swimming fish (a), lateral cameras (b), and technical scheme (c)

Figure S 2: Fish capture. Italian riffle dace (*Telestes muticellus*, a), electrofishing (b), and sorting (c) during the experimental campaign

Figure S 3: Fish storage. Outside artificial pond fed by river water (a), indoor of the hatchery (b), inside storage tanks fed by spring water (c)

Ethical approval

The study was conducted following the Declaration of Helsinki and approved by the Department of Economic Development Protection of Flora and Fauna of the Metropolitan City of Turin, No. 4457 of 29 October 2020, under Italian Decree-Law No. 73.of 19 March 1948, Italian Law No. 56 of 7 April 2014, Italian Law No. 114 of 11 August 2014, Regional Law No. 23 of 29 October 2015, and Italian Legislative Decree No. 26 of 18 August 2000, and by the Director of the Provincial Office for Hunting Fishing Parks and Forests of the Province of Cuneo, No. 3014 of 26 October 2020, under Regional Law No. 37 of 29 December 2006, Provincial Council Decree No. 109 of 13 March 2007. The study was conducted in accordance with relevant guidelines and legislation.

Funding

The research work was carried out with funding received from the European Union Horizon 2020 Research and Innovation Programme under the Marie Sklodowska-Curie Actions, Grant Agreement No. 860800.