## Swansea University E-Theses

## Collective Behaviour: From Cells to Humans

Duteil, Mathieu

How to cite:
Duteil, Mathieu (2018) Collective Behaviour: From Cells to Humans. Doctoral thesis, Swansea University. http://cronfa.swan.ac.uk/Record/cronfa50750

Use policy:
This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence: copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder. Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

Please link to the metadata record in the Swansea University repository, Cronfa (link given in the citation reference above.)
http://www.swansea.ac.uk/library/researchsupport/ris-support/

# Prifysgol Abertawe Swansea University 

# Collective Behaviour: From Cells to Humans 

Mathieu Duteil

Submitted to Swansea University in fulfilment of the requirements for the degree of Doctor of Philosophy

To Jacques Duteil

## SUMMARY

Living in organised groups is a strategy that can be observed in a multitude of diverse species. Among such species, the behaviour of an individual on their own is not the same as within a group: the environment is modified by the presence of more subjects, individuals interact with each other, and from those interactions complex patterns of behaviour can emerge. Some species of animals almost exclusively exist as groups, and as a consequence, studying them in a social context is the only way to understand their behaviour in nature.

This is the idea that drives all the research presented in this thesis: the particular behaviour exhibited by the group is so robust that it will emerge even in a very simplified environment. By observing the individual and the group in those simplified experimental conditions, it is possible to deduce rules that might govern the interaction. The importance of interactions in the group's behaviour can then be demonstrated by implementing a computer model of agents following those rules and comparing it with natural and experimental behaviour.
This thesis presents different examples of such analyses, and gives illustrations of the range of questions that can be answered through this method. Groups of stem cells, juvenile sea bass and human beings were successively observed and tracked in suitable environments, with or without perturbation. The data extracted from those experiments were then processed so as to correct recording errors, and individual and collective behaviours were derived from those data, returning new insights on the nature of the interaction at the individual level, their consequences at the global level, as well as the effects of the interaction on both. Finally, I present the computer models derived from those analyses.
Many systems in nature share this property of global behaviours emerging from deterministic local interaction, and as a consequence studies of this kind could shed light on important questions, of which cancer treatment, ocean acidification and human organisations are but a few examples.

## DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.


## STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s). Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.


## STATEMENT 2

I hereby give consent for my thesis, if accepted, to be made available online in the University's Open Access Repository and for inter-library loan. I give consent for the title and summary to be made available to outside organisations.


Date $\qquad$ .23/02/19

## TABLE OF CONTENTS

SUMMARY ..... iii
DECLARATION ..... iv
ACKNOWLEDGEMENTS ..... vii
DISCLAIMER ..... ix
LIST OF FIGURES AND TABLES ..... xi
ABBREVIATIONS ..... xiv
CHAPTER 1 - GENERAL INTRODUCTION .....  1
1.1. Systems .....  1
1.2. Systems in biology: the group ..... 6
1.3. Data collection methods ..... 8
1.4. Tracking ..... 9
1.5. Importance of the environment ..... 14
1.6. Computer models ..... 18
1.7. Chapters organisation and objectives. ..... 22
CHAPTER 2 - DOUBLE PLATFORM ANALYSIS OF MESENCHYMAL STEM CELLS GROWTH ..... 25
2.1. Introduction ..... 25
2.1.1. Stem cells ..... 25
2.1.2. Methods of observation. ..... 27
2.1.3. Research objectives. ..... 50
2.2. Methods ..... 50
2.2.1. Cell harvesting and culture ..... 50
2.2.2. Data acquisition ..... 51
2.2.3. Data processing ..... 54
2.3. Results ..... 66
2.3.1. Results of the tracking ..... 66
2.3.2. Results on cell motility ..... 70
2.4. Models ..... 71
2.4.1. Microscopy data-driven model ..... 71
2.4.2. Flow cytometry data-driven model ..... 88
2.4.3. Third model ..... 95
2.5. Discussion ..... 98
2.6. Conclusion ..... 101
CHAPTER 3 - IMPACT OF OCEAN ACIDIFICATION ON THE SHOALING BEHAVIOUR OF EUROPEAN SEA BASS ..... 104
3.1. Introduction ..... 104
3.1.1. A changing ocean ..... 104
3.1.2. Sea bass ..... 114
3.1.3. Research question ..... 116
3.2. Material and methods. ..... 117
3.2.1. Animals ..... 117
3.2.2. Observations ..... 118
3.2.3. Tracking ..... 120
3.2.4. Postprocessing ..... 122
3.2.5. Data extraction ..... 126
3.3. Data analysis ..... 130
3.4. Results ..... 131
3.5. Computer model ..... 134
3.6. Discussion ..... 141
3.7. Conclusion ..... 143
CHAPTER 4 - NO I IN TEAM? IMPACT OF INDIVIDUAL CHARACTERISTICS ON HUMAN GROUP PERFORMANCE IN A FORAGING TASK ..... 146
4.1. Introduction ..... 146
4.2. Methods ..... 149
4.2.1. Study subjects and foraging task ..... 149
4.2.2. Subject movement data ..... 152
4.2.3. Data analysis ..... 158
4.3. Results ..... 159
4.4. Discussion ..... 163
4.5. A human Multi-Armed Bandit problem ..... 166
4.6. Conclusion ..... 171
CHAPTER 5 - SYNOPSIS; THE FUTURE ..... 173
APPENDIX ..... 179
REFERENCES ..... 187

## ACKNOWLEDGEMENTS

This thesis is not only the result of four years of work, but also the culmination of my life in academia so far. 10 years ago, I decided that scientific research would be the best career for me from a self-actualisation perspective, but the path has sometimes proven difficult. This is why I would like to present my most heartfelt thanks to my parents, who have been there to support me every step of the way, encouraging me to persevere when previous endeavours fell through. I have experienced too many ways in which a PhD can end in failure, and with that in mind, I want to say that Andrew King and Rowan Brown are the best advisers I have ever had. Not only did they trust me with this long term project, but they also guided me throughout, giving me important help and feedback on every aspect of my work. I have at the very least considered every advice they have given me. There is no doubt in my mind that this thesis would not exist had they not been such great mentors.

Throughout my studies and my career, I have always challenged myself to work on multidisciplinary projects always grounded in new fields of knowledge. In that regard, this project certainly did not disappoint. Four years ago, I only knew the most basic facts concerning the subjects presented here, and I had seriously underestimated the amount I would have to learn here. Fortunately, I was surrounded with fascinating people who were always eager to share their knowledge. In particular, I would like to thank Ines Fürtbauer, who has greatly improved my skills in statistics. She was helped in that by the groundwork covered by the very clear course taught by James Bull and Luca Borger. I would also like to thank Luca for all the events he organised to prepare us students to various aspects of academic work. Hannah Williams and Rory Wilson helped me a lot when exploring the world of animal tagging, and Daniel Strömbom has taught me everything I wanted to know about group simulations. I also want to thank Sahar Seyed Salamat for explaining ocean acidification to me. Naomi Everett's help was invaluable in navigating the trickiest rules of the English language. And I cannot overestimate the help of Alexandra Elbakyan for finding the papers I needed.

The work on cell biology was certainly the part that was most alien to me at the beginning, and what I was dreading the most. I am frankly amazed by how much Rachel Errington, Marie Wiltshire and Rachel Howard-Jones have managed to teach me on that
subject: I can even believe that I understand flow cytometry now! I would like to thank Dimitra Georgopoulou for sharing my confusion every time, until we successfully identified what we did not understand. I wish her the best in her work, and I look forward to read what will come from it!

I will keep fond memories of my time in Swansea, like the various anecdotes from fieldwork that made me very happy that I rarely had to leave my desk (honestly, I am a bit envious of the fascinating work done by the biologists here), the meetings at the Pub On The Pond, the games nights with Jacques-Olivier Laloë and Gaëlle Fehlmann, the fun art project in London with Heather Barnett, the wonderful landscapes of Wales, the conservation work done with Ben Sampson, the food packages sent by my sister Ségolène, the visits from my parents to help me move (I swear that there has often been a nice weather here, you just always happened to choose the worst day to come!). Wherever I go next, I can only hope to form memories as good as those I have from Swansea!

## DISCLAIMER

The following people and institutions contributed to the publication of work undertaken as part of this thesis:

| M.D. | Mathieu Duteil | College of Engineering <br> Department of Biosciences, College of Science, Swansea University |
| :---: | :---: | :---: |
| A.P.-E. | Alfonso Pérez | Instituto Cajal, Consejo Superior de Investigaciones Científicas, Madrid, Spain <br> Department of Physics, Physics of Living Systems Group, Massachusetts Institute of Technology, Cambridge, MA, USA |
| A.J.K. | Andrew King | Department of Biosciences, College of Science, Swansea University |
| D.G. | Dimitra Georgopoulou | College of Engineering <br> Department of Biosciences, College of Science, Swansea University |
| E.C.P. | Ed Pope | Department of Biosciences, College of Science, Swansea University |
| G.G.d.P. | Gonzalo de Polavieja | Champalimaud Research, Champalimaud Center for the Unknown, Lisbon, Portugal |
| I.F. | Ines Fürtbauer | Department of Biosciences, College of Science, Swansea University |
| J.R.U. | James R. <br> Usherwood | Structure \& Motion Laboratory, Royal Veterinary College, Hatfield |
| J.P.M. | Julia P. Myatt | Biosciences, University of Birmingham |
| M.W. | Marie Wiltshire | Department of Pathology, University of Wales College of Medicine, Cardiff |
| M.R.B. | Rowan Brown | College of Engineering, Swansea University |
| N.O. | Nathan Oesch | SENRG, Department of Experimental Psychology, Oxford |
| R.J.E. | Rachel <br> Errington | Department of Medical Biochemistry, University of Wales College of Medicine, Cardiff |


| R.A.H.-J. | Rachel <br> Howard-Jones | Tissue Engineering and Reparative Dentistry, Cardiff <br> Institute of Tissue Engineering and Repair, School of <br> Dentistry, Cardiff University |
| :--- | :--- | :--- |
| R.I.M.D. | Robin Dunbar | SENRG, Department of Experimental Psychology, <br> Oxford |
| S.S. | Seirian Sumner | School of Biological Sciences, University of Bristol |
| S.H. | Stephen Hailes | Department of Computer Science, University College <br> London |

## Chapter 2

The cell cultures were prepared by M.W. and R.A.H.-J., with D.G. and M.D. The observation protocols were set up by R.J.E., with D.G. and M.D. Data were extracted by M.R.B. and M.D. The different growth models are the work of M.D. M.D. wrote the manuscript with input from M.R.B.

## Chapter 3

The work presented in this chapter has been published in a peer reviewed journal ([1], see appendix). A.J.K. and E.C.P. conceived and performed the experiments and collected the video data. A.P.-E., G.G.d.P. and M.D. tracked fish trajectories from video. M.D. processed the data and extracted parameters for analyses, and constructed the simulation model with input from M.R.B., I.F., A.J.K. M.D. analysed the data. M.D. and A.J.K. led writing of the paper with input from all authors, who read and approved the final manuscript. M.D. wrote the chapter, with input from A.J.K., I.F. and M.R.B.

## Chapter 4

A.J.K., S.S., N.O. and R.I.M.D. designed the experiment. A.J.K., J.P.M., S.H. and J.R.U. performed the experiments and collected the data, which were processed by M.R.B. and M.D. The statistical analysis was conducted by I.F. and M.D. M.D. wrote the manuscript with input from A.J.K., I.F. and M.R.B.

## LIST OF FIGURES AND TABLES

## Figures

Figure 1: Examples of emergent systems ..... 5
Figure 2: Stokes’ shift ..... 28
Figure 3: Different fluorescent proteins and nanoparticles in a petri dish ..... 30
Figure 4: An application of multi-colour fluorescence ..... 31
Figure 5: Structure of a quantum dot nanoparticle ..... 33
Figure 6: Principle of a bright field, phase-contrast, and fluorescence microscopy ..... 39
Figure 7: Three delivery methods for flow cytometry ..... 40
Figure 8: Hydrodynamic focusing ..... 41
Figure 9: Collection optics for forward scatter signal ..... 43
Figure 10: Simplified model of a flow cytometer ..... 44
Figure 11: Order of microscopy observation through a well-plate ..... 53
Figure 12: Stitching of bright field and fluorescence microscopy images. ..... 55
Figure 13: Microscopy image, before and after processing ..... 57
Figure 14: Average fluorescence of all the images collected. ..... 57
Figure 15: Distribution of fluorescence across images ..... 58
Figure 16: Comparison of bright field and fluorescence images ..... 59
Figure 17: Different error types that can be obtained from automated detection. ..... 61
Figure 18: An attempt at cell detection by Cell-IQ analyser ..... 62
Figure 19: Examples of composite FOV ..... 63
Figure 20: A cell and its daughters before, during and after mitosis ..... 64
Figure 21: Example of FOV being tracked ..... 65
Figure 22: Distribution of inter-mitotic times ..... 67
Figure 23: Evolution of total fluorescence over time ..... 68
Figure 24: Fluorescence of the parent and daughter cells. ..... 69
Figure 25: Distribution of diffusion coefficients of the DPP ..... 71
Figure 26: Probability mass function of the inter-mitotic times ..... 73
Figure 27: Distribution of possible repartitions of QD upon mitosis ..... 74
Figure 28: Evolution of the cell coverage of 9 FOV throughout the experiment ..... 75
Figure 29: Growth rate of the cells observed through microscopy ..... 76
Figure 30: Fraction of the cells that do not go through mitosis. ..... 78
Figure 31: Flowchart describing the cell proliferation model ..... 79
Figure 32: Fluorescence in the flow cytometry experiment and in the MDDM ..... 81
Figure 33: Final distributions of each of the four parameters of the MDDM ..... 84
Figure 34: NDF functions that best match the flow cytometry results ..... 85
Figure 35: Fluorescence fom flow cytometry, and according to the MDDM ..... 87
Figure 36: Growth rate of the cells in the flask ..... 89
Figure 37: Model of the fraction of cells that do not grow through mitosis ..... 90
Figure 38: Final distributions of each of the three parameters of the FCDDM ..... 92
Figure 39: Three examples of NDF models obtained from the FCDDM, and the model directly obtained from the flow cytometry data ..... 93
Figure 40: Evolution of the distribution of fluorescence, as observed experimentally through flow cytometry, and according to the FCDDM ..... 94
Figure 41: Cost function obtained when the NDF is defined as a step function ..... 96
Figure 42: Evolution of the distribution of fluorescence, as observed experimentally through flow cytometry, and according to the step model of the NDF ..... 97
Figure 43: Effect of pH on the concentration of carbonates ..... 107
Figure 44: Radiations transmitted by the atmosphere. ..... 110
Figure 45: European sea bass and its habitat ..... 114
Figure 46: System used to maintain pCO 2 and temperature throughout the growth of the fish. ..... 117
Figure 47: Observation setup for the sea bass experiment. ..... 119
Figure 48: The interface of idTracker ..... 121
Figure 49: Sea bass tracking ..... 122
Figure 50: Example of erroneous data to be removed from the tracking ..... 124
Figure 51: Trajectory from Figure 50, after the data have been processed. ..... 125
Figure 52: Distribution of speeds for all conditions ..... 126
Figure 53: Most visited parts of the arena in all conditions ..... 127
Figure 54: Exploration pattern. ..... 128
Figure 55: Distribution of inter-fish distances for each condition ..... 128
Figure 56: Neighbourhood area of a fish and the fish connected to it. ..... 129
Figure 57: Graphical procedures for model diagnostic. ..... 131
Figure 58: Comparison across the three PC of the three conditions. ..... 133
Figure 59: Parameters that describe the state of a fish in model 1 ..... 135
Figure 60: Distribution of the model 1 parameters in the experimental data. ..... 136
Figure 61: How model 2 handles impossible trajectories ..... 138
Figure 62: Comparison of the trajectories of real fish, model 1 and model 2 ..... 138
Figure 63: Differences in behaviour between real and simulated fish ..... 141
Figure 64: The foraging arena ..... 151
Figure 65: Distribution of speeds in the corrected data set ..... 153
Figure 66: Two examples of trajectory from the foraging experiment ..... 154
Figure 67: Usable and accurate data ..... 156
Figure 68: Position of all subjects at one instant ..... 156
Figure 69: Distribution of the variables that describe subjects' behaviour ..... 157
Figure 70: Graphical procedures for model diagnostic ..... 159
Figure 71: Two of the "landscapes" employed by Reverdy et al. ..... 170

## Tables

Table 1: Most common variables used to describe collective motion at the local and global scale, for different types of species. ..... 21
Table 2: Main aspects of the systems studied in the following chapters, and of the methods used to study them. ..... 24
Table 3: Factorial decomposition of the first three principal components ..... 132
Table 4: Results of model testing for differences in the behaviour of fish across experimental conditions ..... 133
Table 5: Results of model testing for differences in the behaviour of real and simulated fish for different behavioural variables ..... 140
Table 6: The BFI-10 questionnaire ..... 150
Table 7: The different analyses planned ..... 159
Table 8: Spearman correlation coefficients between the individual characteristics ..... 160
Table 9: Results of model A ..... 160
Table 10: Results of model B ..... 161
Table 11: Results of model C ..... 163

## ABBREVIATIONS

C: carbon
DEA: differential evolution algorithm
DNA: deoxyribonucleic acid
DNN: distance to the nearest neighbour
DPP/DPPSC: dental pulp progenitor stem cell
EDTA: ethylenediaminetetraacetic acid
FACS: fluorescence-activated cell sorting
FCDDM: flow cytometry data-driven model
FCS: fetal calf serum
FFM: five factor model
FOV: field of view
FRET: fluorescence (or Förster) resonance energy transfer
FSC: forward scatter
GEV: Generalised Extreme Values
GFP: green fluorescence protein
GPS: Global Positioning System
IMT: inter-mitotic time
LMM: Linear Mixed Model
MAB: multi-armed bandit
MAS: multi-agent system
MBTI: Myers-Briggs Type Indicator
MDDM: microscopy data-driven model
Mt C: megatons of carbon
NDF: non-dividing fraction
NEO-PI-R: Revised Neuroticism-Extraversion-Openness Personality Inventory
OA: ocean acidification
OCEAN: Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism
OFT: optimal foraging theory
$\mathrm{pCO}_{2}$ : partial pressure of carbon dioxide
ppm: parts per million

QD: quantum dot
RFID: radio frequency identification
RNA: ribonucleic acid
s.e.: standard error

SSC: side-scatter
std: standard deviation
UCB: upper confidence bound algorithm
UV: ultra-violet
WCST: Wisconsin Card Sorting Test

## CHAPTER 1 - GENERAL INTRODUCTION

### 1.1. Systems

A system can be characterised as an interconnected set of elements that are coherently organised to perform a function [2]. A system can thus be described by its elements, their connections, and its ability to perform a given function. Examples of systems can be found at a variety of scales and in all sorts of contexts: as observed by Boulding [3], every discipline observes some kind of individuals: "electron, atom, molecule, crystal, virus, cell, plant, animal, man, family, tribe, state, church, firm, university, corporation and so on". These examples illustrate an important fact: many entities can be seen both as systems and as elements of a larger system.

Different classifications of systems have been presented, each relying on different aspects. General system theory distinguishes systems based on their structure, and then based on their field of application [3]:

- The first level is that of static systems, such as crystals, in which the interaction between elements is a fixed connection.
- The second level is that of clockwork: a simple dynamic system is programmed to perform an action, but there is no feedback from the environment to verify that the system has reached its desired state: its motions are predetermined. Examples include levers and pulleys.
- The third level is the thermostat: the system tries to maintain a certain state by minimising the difference between its current and desired state. They can thus often be modelled by systems of differential equations.
- The fourth level includes open systems: here the system maintains its existence by information exchange with its environment. Examples include all forms of life.
- The next levels are of a different nature, as Boulding's classification [3] separates plants from animals, from humans, from human organisations.

The different nature of these latter levels implies that those different categories cannot be described by the same tools. Other authors disagree with this assumption. Coupled sets of differential equations for instance have been used to describe systems at the scale of atoms [4], molecules [5], cells [6], non-human [7], human populations [8] or whole ecosystems [9]. Multi-agent systems (MAS) are another tool that can be used to describe such systems [10]. Thus, the complexity of a system is not inherent, but rather depends on the desired level of detail to be resolved.

Some properties of a system however cannot be explained without considering the individual elements that comprise it, their interactions, and their environment. These are called emergent properties. According to Goldstein [11], emergent aspects are characterised by different qualities:

- They are not deducible from lower level components.
- They are integrated wholes that maintain themselves over time.
- They characterise a macro-level property of the system.
- They are dynamical constructs that arise as the complex system evolves.
- They are "ostensive", which is to say that emerging patterns can be recognised as such by an observer.

Like systems, emergent patterns can be found in many disciplines (see Figure 1):

- A snowflake is the result of the sum of the local interactions of water molecules. It has properties of fractal patterns and central symmetry that are not observed at the level of the molecule. Its structure stays coherent over time. It arises as a complex system [12].
- More generally, crystallisation is a "process in which an ordered phase emerges from a disorder state" [13]. Crystals are composed of a limited number of building blocks that arranged in simple ways. The macrocrystal may have properties that cannot be found at the level of these building blocks, such as birefringence, optical activity, pyroelectricity or piezoelectricity [14].
- Termite cathedrals are structures several orders of magnitude the size of the insects that build them. Their structure presents a high degree of regularity, with different rooms that appear to present specific functions [15]. No individual termite would be capable of building such a structure, nor has a construction
plan: cathedrals emerge from the repeated interactions of simple behaviours, such as moving soil pellets and following pheromone trails.
- Cells need to coordinate their action to form a tissue. It has been shown that this coordination can occur when cells modify their behaviour (in terms of death, migration or division) in response to interaction with their local environment, in particular the mechanical pressure from other cells [16]. This process can lead to the creation of tissue and its preservation in a healthy form, by healing it at a steady rate.
- Similarly, mechanical stress plays a role in the growth pattern of leaves, causing the initiation of the growth of new organs in the leaf along a fractal structure [17]. The organisation of the cells is such that their mechanical forces cause new shapes to emerge [18].
- The brain is a complex system of neurons and glial cells: each neuron is connected on average to 7000 others. Neurons in the human brain fire as a response to the signal they receive from the others, which gives rise to all the cognitive functions observable at the level of the individual [19].
- A lonely ant would drown in a large quantity of water. Yet as a group, ant colonies are able to form water-repellent "rafts", allowing them to survive until they reach the ground [20]. These rafts do not require any additional material and are self-assembling: despite having no knowledge of the organisation of the group except at the local scale, ants are able to organise in a structure that traps small pockets of air, making the group as a whole buoyant.
- Fish in schools align their position with that of their neighbours along precise rules [21]. Thus they form structures that are robust to attacks, as they can quickly regather along the same formation, always moving together as a group [22].
- Many species of birds fly in flocks; flocks of starlings can count thousands of individuals flying in a synchronised manner. An individual starling does not know its required position in the flock; instead, the regular structure is achieved through local interactions, with each bird adjusting its velocity towards the average of that of its neighbours [23]. The resulting structure offers several
advantages to the birds, in terms of aerodynamics, communication and navigation [24].
- Humans can react to subtle social cues, which can lead to all sorts of emergent behaviours in a group. One example of such behaviour is the wave, in which rows of people stand up, lift their arms and sit down in the pattern of a wave. This pattern is often observed in stadiums ; the largest examples can gather more than hundred thousand people [25]. Waves are triggered by the simultaneous excitation of a small group of people, and propagate at a regular pace (of 20 seats per second, or $12 \mathrm{~m} / \mathrm{s}$ ) along a stable pattern [26].
- Although cities can be built according to a top-down approach, the interactions of the different neighbourhoods that constitute it also contribute to its growth. Typically, individual preferences in the choice of neighbours can result in the emergence of segregated neighbourhoods [27], even if every individual prefers to live in a diverse neighbourhood [28].
- In economy, the aggregation of individual transactions can lead to selfreinforcing mechanisms, which can generate bubbles or crashes [29].


Figure 1: snowflakes [12], crystals [13], termite cathedrals [15], cell colonies [16], tree leaves [18], neurons [19], ant colonies [20], fish schools [21], bird flocks [23], human group behaviours [26], cities [27] and the stock market [29] have all been described as having emergent properties. All pictures were selected from Wikimedia Commons [30]

### 1.2. Systems in biology: the group

In the case of living organisms, individuals seek to maximise their own survival and reproduction, which results in changes in gene frequencies in populations (evolution) over time in response to ecological and social conditions [31]. Where individuals can maximise their survival and reproduction via social interaction and collective patterns of behaviour, these will be selected for. One general strategy has proven so effective for survival that it has been selected for in very diverse species: being part of a group.

Groups can provide different kinds of advantages to an individual. For organisms that reproduce through sexual reproduction, the group provides an easy access to mates. Different kinds of strategies have emerged to optimise this access. Some species of animals form durable monogamous pairs [32]. Others reserve mating to one or few individuals. In some mammal species, the existence of a dominance hierarchy ensures a higher reproductive success for the alpha (male and female) [33], [34] and individuals compete for such roles. In eusocial insects, all eggs are produced by the queen [35]; the variability in phenotypes, all associated with specific functions, sharpens the division of labour, which is thought to increase the fitness of the colony [36]. Although the death of the sole reproducing female can throw the colony into disarray if a new queen is not raised soon [37], the presence of the hive ensures the queen's security in most circumstances. Hence, although the queen is a single point of failure for the system, her high level of protection makes eusociality an effective strategy to reduce stressors [38].

Groups also offer protection from predators. Although a large gathering of prey will attract predators, the probability for the individual to get eaten is lower (dilution effect [39]). According to the selfish herd theory [40], individuals in a group of potential preys will try to remain in the centre of the cluster, as the edges are the region most likely to be preyed upon. The larger the group, the smaller the edge area will be in proportion to the group. This however comes with a trade-off, as individuals at the centre of the group might be less successful in their foraging [41]. Group motion also tends to confuse predators: the large number of preys can make it difficult for a predator to focus on a single one [42]. Whereas a single individual would have to be vigilant regarding the entire environment, a group can watch over their environment together. Individual vigilance is allowed to decrease, thus costing fewer resources on the individual. It also
adds redundancy to the system, since even if several members of the group were not to notice the approach of a predator, one individual who does is enough to warn the entire group [43]. Vervet monkeys in particular have developed a complex alarm call system, with different cries depending on the predator, allowing the other group members to select the most appropriate escape strategy [44]. Finally, a large enough group might be able to discourage potential predators from attacking in the first place, a strategy called mobbing [45].

The group is also a way to give every member a better access to food [46]. In a patchy environment, a lot of resources can be spent in locating a patch of food, but an individual might not be able to eat it on their own. Being part of the group means that several individuals can scout for new sources of food, and then benefit from the discovery of a patch, whether they are the one that discovered it or not. A group might also be able to hunt larger preys than a solitary animal could (cooperative hunting) [47].

The group can make locomotion simpler. Schools of fish [48] and flocks of birds [49] organise themselves in a way that minimises the effort for most of the individuals in the group, by facilitating the flow of water or air. Being part of a group also makes it possible to overcome geographical obstacles that would be insurmountable for the individual. For instance, ants are capable of building bridges and rafts using the bodies of everyone in the colony [20] (see Figure 1).

Finally, being in a group means that an individual can become specialised in the skills that are the most useful to the group [50]: group size has been shown to promote specialisation [51]. Social animals might take different roles within the group based on their personality for instance [52]. These advantages are not restricted to non-human animals, and similar advantages arising from diversity have been observed in humans [53], [54], although negative effects of diversity have also been reported for human groups [55]. Chapter 4 delves deeper into the influence of personality on group strategy. In cell populations, specialisation is thought to be the process that has led to the transition from colonies of unicellular organisms to multicellular organisms [56].

Those advantages come with trade-offs. If the size of the group is larger than the amount of resources available, competition within the group forces it to split [57].

Similarly, in an environment where the number of predators is too large, the safety in number strategy can breakdown and living in a group becomes a risk to draw predators [58], [59]. For this reason, there is an optimum group size that minimises costs and maximises benefits for individuals. Disease is a third factor that can limit the size of groups: as the group grows larger, the number of contacts between members increases, so more members can become infected [60]. In the case of sexually transmitted diseases in particular, because large groups have more nodes in their sexual networks, pathogens can spread faster than they can between small communities [61]. When the size of the group becomes far greater than that number, the collective will split, whereas two small groups might choose to join to gain strength in numbers [62].

Variation in individual behaviour has an impact on group level patterns: the resulting system can present complex emergent properties, which are known as "collective behaviour" [63]. There is no direct influence of the modification of behaviour of an individual on the group, yet the behaviour of the group cannot be properly understood without studying the individual behaviours that lead to the emergence of its structure. Understanding collective behaviour requires a precise description of the group behaviour as well as the behaviour of its members.

### 1.3. Data collection methods

Gathering data on all members of the group and all parts of the system can be difficult [64], [65]. The most common method is simple direct observation [64]. For instance in the early $20^{\text {th }}$ century, Selous [66] observed various species of birds over 30 years in order to understand the synchronicity of their flight. Other researchers have also attempted to gather data on flock flying over longer timescale and in larger environments by following the flocks, be it by flying a sailplane [67] or by driving a car [68].

Modern day methods to study collective behaviour of birds and other organisms tend to rely on trajectories being captured through photograph or video recording. The study of collective motion requires most of the group to be visible at all time, so a fixed
recording device would only work if the organisms are expected to remain inside the field of view for a sufficiently long time to observe the collective behaviour, without being occluded. Examples include studies of flocks of birds against a clear sky over a few seconds [69], herds of buffalo migrating through an empty background as observed from far above [70], fish inside a tank [71]-[73], or small organisms such as insects [74], [75], or cells under a microscope [76]. In the case of humans, ethical considerations must be taken into account before capturing their image, especially if they are observed in a public environment in which they could not be asked for consent beforehand.

Photographs can be useful when the movement is very slow [77]. It is notably used to observe cell cultures [78], [79]. The frequency of capture is important, as too low a frequency could result in overlooking important biological interactions or processes and the reactions to pharmacological perturbations. Also, when individual organisms are physically very similar and move in a coordinated motion at relatively close distances, it can become intensely difficult to accurately and robustly identify each individual within the system or larger population. If the researcher is only interested in the geometric and dynamic structure of the group though, such instances of label swapping could be acceptable [80]. Film recording can be considered as the limit of the technique, in which the frequency is equal to the framerate. Even at such a high sampling frequency, label swapping can still be a problem, in particular in cases in which some organisms are occluded, (by another organism for instance) [81].

### 1.4. Tracking

Before computer tracking became an option, researchers were limited to qualitative information on the collective patterns, or to smaller data sets gathered manually. The difficulty of analysis could be reduced by lowering the frequency of capture, the number of organisms, and the complexity of the behaviour observed. Progress in automated image analysis has allowed the study of much larger groups. Still, automatic detection is limited to organisms that move slowly compared to the framerate, on an empty background, and are distant enough from one another that each moving object in the
frame can be identified as one organism [82], [83]. In 2008, project StarFlag used stereophotography to analyse of a dense flock of starlings [84]. The data collected followed the trajectory of individual starlings in a 3D environment, within a murmuration that included up to 2600 individuals. These data were used to show that starlings were interacting with the individuals present in their topological range (their 6 or 7 closest neighbours), rather than their geometrical neighbourhood as it was previously thought. Since then, more advanced software have been developed for the automated tracking of various organisms [85]-[88]. I explain in Chapter 3 how I tracked sea bass with such a program, idTracker [81].

A common method to make tracking simpler for the algorithm is to set up the experiment in such a way that the organisms appear very distinct from the background and from each other. The different organisms in the group can be selected distinct enough from one another for the algorithm to tell them apart simply in most circumstances. They can also be made more distinctive by marking them with different patterns [89], or by placing a coloured tag on them [90]. Then, each individual can be followed by having an image processing algorithm search the picture for such pattern or colours. The effect of such marking varies depending on the species being examined. Fish appear to suffer little effect from it. In the case of three-spined sticklebacks, it has been shown that the stress response of the fish to this tagging process is within the normal range of stressful experiences, and thus having the tag should not disturb their normal behaviour [91]. It has also been reported that the swimming patterns of zebra fish is not modified by the presence of a tag [92]. For fowls on the other hand, a minority of marked individuals could become the target of aggression from non-marked birds [93].

Marking is also possible in microbiology, by labelling the cells with fluorescent markers [94], so cells can be followed by focusing on the pixels containing the corresponding colour. This is the technique that will be presented in Chapter 2, using quantum dot nanoparticles as a fluorescent marker; these semiconductor nanoparticles emit a light at a wavelength of 705 nm (red) when irradiated with ultra-violet light. This is particularly useful, as contrary to other organisms that can be studied in the lab, the cell population cannot be assumed to remain constant throughout the experiment. Such fluorescent
markers then get transmitted to daughter cells through generations, allowing to track them too, until the signal gets too diluted to be detectable from the noise [94]. One way to avoid the problem of dilution is to work with cells that have been genetically engineered to emit fluorescence on their own [95].

The location of the organism need not be detected visually: rather than using a tag to help the animal being detected by a camera, the tag can instead be a sensor itself. One example of such technology is Radio Frequency Identification (RFID) systems. Now common in our everyday lives through applications such as passport security [96] or shoplifting prevention [97], the technology has also been used in animal tracking [98]. Examples include such diverse species as salmons, deer, bison, ferrets, bears, tortoises or armadillos [99]. Antennas are regularly located throughout the environment and emit radio signals; the tags contain transponders that reply to the antennas when they receive their signal to transmit via radio waves individualised information. Finally a central unit processes the information gathered by the system: the distance between the RFID tag and the antennas in its vicinity can be estimated from the strength of the signal received by the antennas from the tag, and the position of the tag can be deduced by trilateration [100]. The advantage of the system is the low cost of the tags, making it possible to study large populations; but at the same time, animals can only be detected near the antennas (usually closer than 1 m , although recent systems allow detection up to 90 m [99]), and very close animals might be detected as the same individual, leading to imprecisions in the analysis of dense populations. RFID technology is thus mostly useful in a circumscribed environment. Typically, Kritzler et al. [101] used a system of subcutaneous transponders, antennas and readers to detect the position of mice inside a $1.75 \times 1.75 \times 2.1 \mathrm{~m}^{3}$ semi-natural environment, to collect data on the mice movement, behaviour and dominance relations. Another example by Aguzzi et al. [102] used RFID technology to circumvent the limitation of telemetry in saltwater, and were thus able to monitor the behaviour of Norway lobsters with a precision of 1 cm .

Other technologies rely on the same approach of 'biotelemetry'. Radiotelemetry is mostly used to track land or aerial movement; the size of the tag rarely is a problem, as there are tags so small they have successfully been used to track bees [103]. Acoustic and ultrasonic telemetry on the other hand is more appropriate for marine animals.

Another example is radar; for instance, the movement of cows inside their barn has been tracked within a precision of 25 cm using radar technology with a transmitter fixed at the neckband of the cows [104]. A comparison of the different techniques can be found in [105] (table 1 in particular); the choice mostly depends on the environment (air or water, and size of the area), the number of animals to be tracked, and the precision required.

Humans already use a system to track their movements all over the surface of the world: Global Positioning System (GPS). This is a different telemetric system, in which the antennas are a network of satellites in orbit around the Earth. Use for aquatic animals is possible, but it requires them to surface to generate new fixes, meaning that the general trajectory of the animal can be followed, but not the position on a finer time scale [106]. Similarly, when tracking penguins, Ryan et al. [107] obtained precise data above the surface of the sea, but they observed that the GPS signals are interrupted when the bird dives. In the case of turtles, GPS data were successfully collected when the turtles were swimming up to 4 m deep [108]. GPS are only usable if there are no obstacles between the animal and the sky, which excludes indoor use, or inside a dense forest [109]. Its precision is of about 10 to 15 m for $95 \%$ of readings [110], so GPS should not be used on its own for applications that require more accuracy (although continuous recording and post-processing and selection of satellite data can help bring accuracy closer to $\pm 5$ m [111], see Chapter 4 for further discussion). One final limitation of GPS is that communication with the satellites requires a lot of energy, which imposes a trade-off between the size of the battery, the frequency of readings and the duration of the tracking. Other applications of GPS for biotelemetry include tracking of sheep [111], moose [112], [113], wolves [114], zebras [115], migrating birds [116], [117] or seals [118].

Since it is the communication with the satellite that is so energy intensive, one solution to save energy without losing too much accuracy is to keep recording data, but transmit them less frequently. This approach of storing data inside the tag is called bio-logging [119]. Besides the location of the animal, the sensor can measure diverse variables such as acceleration, temperature, barometric pressure, light or salinity, which can then be used in a variety of way to understand the animal's behaviour [120]. Accelerometry data
on the jaw could be used to detect mastication and thus feeding [121]. The animal's oxygen partial pressure and heart rate are indications of the animal's energy expenditure [122], [123]. Stomach temperature and pH (as measured through ingested sensors) can help understanding the animal's digestion [124], [125]. Sensors can also be used to determine properties of the environment such as depth [126], thus putting the animal's actions into context [127]-[129]. Some measurements can be used as proxies to infer data that would otherwise be hard to access. For instance, the evolution of the buoyancy of a seal can be used as a proxy for the formation of body fat [130].

Because of the richness of the data that can be inferred from a collection of sensors, it would be tempting to add any sensors that could possibly gather any useful information. But besides the cost of the sensors and the complexity of constructing a coherent picture from disparate measurements, the limitation of the battery must be considered, and increasing the battery life would impact another aspect of the tag: weight. So as not to hinder the animal's movement, the tag should not represent more than $5 \%$ of the animal's mass [131] ( $3 \%$ for flying birds [132]), and be positioned on the animal in a way that does not perturb movement; in particular, questions of hydro- and aerodynamics should be considered for swimming and flying animals respectively [133] One particular example of bio-logger is the daily diary [134], a multi-sensor archival tag which consists of a triaxial accelerometer, a triaxial magnetometer, a barometric pressure sensor, two thermometers, two light sensors and a GPS. The GPS is used to provide periodic recalibrations of the position, in order to cancel the cumulative error from the magnetometer and accelerometer; with 8 samples per seconds from the accelerometer and the magnetometer, and a normal GPS use, the battery normally lasts for 80 days. The data gathered from the sensors can then be used to infer the animal's location and movement, its behaviour, its energy expenditure and the environmental conditions. Daily diaries have been used on a multitude of species, including sharks, turtles, crocodiles, penguins, albatross, vultures, emus, badgers, sea lions, seals, cheetahs, dogs and horses [134].

The data gathered from the bio-logger can either be collected regularly via biotelemetry (either by satellite or by radiotelemetry; the latter requires following the animal from a distance in order to be regularly within range of emission [135]), or by recovering the
bio-logger. This can either mean recapturing the animal [134], or using a drop-off mechanism that releases the tag once its memory is full or its battery is depleted [136].

There are also non-visual methods of tracking that do not rely upon fitting the organism with sensors. For instance, Makris et al. [137] were able to monitor fish populations over an area representing thousands of square kilometres using ocean acoustic waveguide propagation. In all of these methods, the presence of noise means that the data must be cleansed before drawing conclusions on the fine details within the data set. Finally, in a different approach, the environment itself can be employed as a sensor [138]. This study was about the processes that lead to stampedes in a large religious event; the area was too large to be covered through video surveillance, and the number of people ( 30 millions) prevented the use of individual sensors. Instead, collective motion data were gathered through portable mats with pressure sensors, making it possible to collect data in real time on the flow of pedestrians at any time or place during the festival.

### 1.5. Importance of the environment

One of the factors influencing the choice of the tracking method is the environment. Tracking based on video recording for instance requires the organisms to remain within the field of view of the camera for the entire duration of the experiment. One way to ensure of that is to limit the scope of the environment. Observing the organisms within a controlled environment presents several advantages:

- Researchers can have the organisms move within an environment that is optimised for detectability. Many animals have evolved so that they would blend within their environment [139], so automated detection would not work when observing such animals within their natural environment. Fish for instance are often observed in a bright featureless tank, with little influence on their behaviour [140]-[143]; such a design is employed in Chapter 3.
- Some behaviours are environment-dependent, so in the wild, researchers have to wait until the animals move towards an environment in which the desired behaviour can be observed. In the laboratory on the other hand, the environment
can be designed to promote the behaviour, increasing the likelihood of it being observed.
- In the wild, animals will most often be observed in their natural habitat. To understand the influence of the environment on the animal's behaviour, it can be useful to observe the animal in a completely different environment. Typically, cuttlefish are able to effectively camouflage within their complex environment by changing their appearance. In order to understand what visual aspects of the environment guide these transformations, Barbosa et al. [144] placed cuttlefish in a checkered environment, something unlikely to happen spontaneously in the wild. The behaviour of the cuttlefish allowed them to deduce that cuttlefish responds to stimuli both in the horizontal and vertical plane.
- Laboratory experiments also allow researchers to reproduce environmental conditions that are inaccessible directly. Such experiments can help us understand whether microorganisms can survive on other planets [145], understand life in the Archean Eon [146], or even the future. In Chapter 3, I show one method of studying whether animal behaviour will change in the next hundred years because of ocean acidification.
- In order to control for individual traits, animals observed in the wild can be captured, tested for those traits, tagged and released [147]. If instead of being released, they are placed in a controlled environment, it is possible to select specimens according to those traits and measure their influence on behaviour [148]. Such traits can include sex [149], [150], personality [151]-[153], or dominance [154], [155].
- To compare the behaviour of two individuals, they must be observed in the same conditions, which requires presenting them with the same stimuli. This can be a problem when the stimulus under study is the behaviour of conspecifics (other subjects of the same species), as even if the animal exhibits the expected behaviour for the focal animal to respond to, this behaviour will never be exactly the same for all individuals. This difficulty can be overcome in the laboratory by using robots. For example, Frohnwieser et al. [156] used a robotic bearded dragon so that they could observe the reaction of the subjects (living bearded dragons) to something they would consider as a conspecific, that would move in
exactly the same way in every instance. Robotic animals can be really useful for the study of collective behaviour, as a robot can be programmed to follow the rules we believe to be followed by the animals in the group, allowing researchers to observe whether biological animals interact with it in the same manner. This approach has been successfully applied to cockroaches [157] and fish [158]. In a similar manner virtual reality can be used to overcome most of the difficulties presented by lab experiments, providing an open yet controlled environments for the animals to explore. This approach has been applied to mice, drosophila and zebrafish [159].

Yet laboratory experiments do present some drawbacks compared to observations in the wild:

- As the animal is not observed in its natural habitat, there is no guarantee that its behaviour will be similar. Captive animals may present some differences compared to wild ones; Calisi et al. [160] remarked that providing a captive animal with an abundance of resources may mask some trade-offs that normally occur in nature. The behaviour of an animal may change depending on its environment [161]-[163]; it is thus essential to provide an environment that is as close to the normal habitat of the animal [164], or to ensure it provides the same level of stimulation [101]. If not, it is important to control that the studied species does not exhibit major differences of behaviour depending on the environment.
- The environmental factors that influence the animals' behaviour are not always well-known, so they might not have been taken into account when designing the artificial environment. For instance, it is by providing fish schools with patches of different light intensity that Berdahl et al. [165] discovered their property of emergent sensing.
- Some environments are extremely complex and hard to model accurately. Such is the case for cells for which the environment is the human body. Although progress has been made towards accurate models of human biology [166], predicting the effect of a drug on a cell population is still a problem that requires
an iterative process towards more precise models, from Petri dishes to cell tissues, to animal bodies, to human bodies [167].
- Because of the spatial requirements to host a large population and the difficulties in tracking very large populations, experimental set-ups usually involve a smaller population than the groups that are observed in the wild. Even a large scale experiment like the one conducted by Buhl et al. [75], which studied a plague of 120 locusts in an arena 80 cm in diameter is still extremely small compared to the swarms of billions of individuals that can cover more than a $1000 \mathrm{~km}^{2}$ [168].
- Some experiments would be unethical. Typically, one form of human collective movements has been extensively researched: stampedes [169]-[172]. This research is normally done by studying footage recorded during such events. Experiments have been conducted to test the hypotheses that came from those studies, but only on a smaller scale, in order to limit the danger for the subjects [173], [174].
- Some important behaviours require specific conditions that would be difficult to reproduce in the lab. One example would be the attack of a predator. Yet even such behaviours have been successfully modelled for some species [175], [176].
- Some behaviours, by their very nature, require to be studied on a large scale, be it spatial (migrations for instance) or temporal [177] (such as evolutionary processes). Lab experiments are not well-suited for either.

Lab experiments are an ideal way to complement observations in the wild, but their results need to be compared with one another to ensure that the behaviour observed in the lab matches what happens in the wild, and to control the validity of the explanation given for behaviour in the wild [178]. It is sometimes possible to use the methods of the laboratory in the wild, thus allowing for comparisons between the natural and the artificial environment. This is what was done by Ward et al. [179] when they recorded the structure of a wild shoal of sticklebacks. Two cameras were set up in a river, one at the bottom filming up to record the shoal on a horizontal plane, and one next to the river bank just below the surface to record on the vertical plane orthogonal to the flow of the river. Outside of the ideal conditions of the laboratory, they had to do the tracking by
hand. Ward et al. reported behaviours similar to those that had been observed in the lab by other researchers, in terms of lateral position of the fish, speed and group polarization [179]. Reciprocally, by bringing greater realism to the design of experimental environments, experimental researchers can provide a better understanding of the mechanisms of collective behaviours and the reasons that might have led to their evolution [178].

### 1.6. Computer models

A third tool can help us to complete our understanding of collective behaviour: computer simulations. Multi-agent systems (MAS) have proven invaluable when modelling systems in which many units interact together. They have found applications in various domains, including engineering, ecology, biology, economics and psychology [180]. In a MAS, agents can be programmed with goals as well as simple rules that dictate how they interact with their environment and each other. That environment could be a continuous geometrical space, a grid, or something more abstract. For instance in agent-based computational economics, an agent could be characterised by its information about firms, its information about consumers, its address book and its recorded communications, but no geometrical coordinates [181]. In models of land use, the environment can be characterised by which subset of the land each farmer agent can act upon, and the movement of the farmer to any of these places does not need to be implemented [182]; thus the geography of the environment is implemented indirectly.

Grid-based models represent space discretely and provide an interesting compromise when the spatial relationship between agents is important but their precise dynamics is not. Segovia-Juarez et al. [183] for instance describe a model of infection of alveolar lung tissue by bacteria in which the environment is a lattice whose squares represent an area of $20 \mu \mathrm{~m} \times 20 \mu \mathrm{~m}$. Each of these microcompartments can contain at once: one macrophage and one T cell (the agents of the model, which can move to different microcompartments), chemokine and bacteria. At each discrete time step (one time step representing 6 s in this example), bacteria will replicate, the macrophage will have a certain probability of eating some of the bacteria in its microcompartment or be infected
by them; infected macrophages will release more chemokine in surrounding squares, the level of which will decide whether a new macrophage will appear. T cells may or may not attack infected macrophages depending on their number in the squares surrounding the macrophage, which may spread the bacteria in the surrounding squares; finally a chronically infected macrophage will burst on its own, here again spreading bacteria in the surrounding squares. Although the discrete nature of the model is a simplification that makes it inappropriate to represent the movement of the cells, at a larger scale, it provides an accurate simulation of the formation of a granuloma [183].

A specific class of MAS aims to model the dynamics of collective motion: selfpropelled particles models. Most models can be categorised as either Eulerian or Lagrangian; Eulerian models describe the group as a continuous surface or volume with local change in density [184]-[190]. Such models are suitable in cases where the hypothesis of a continuous group holds, and if the interactions rules are not too complex to be modelled by differential equations. In Lagrangian models on the other hand, organisms are represented as particles interacting with their neighbours [23], [191], [200], [192]-[199]. At any time, a particle's position and speed can be deduced from the speed and position of all particles in the system at the previous time point, $t$, and the mathematical rules that control their interactions. As a consequence, in order to determine the state of a system at a further time point, $t+\Delta t$, knowing the state of the system at time $t$, it is necessary to propagate the system over the interval $\Delta t$. Since every additional particle adds the possibility of interacting with all other particles in the system, this approach becomes very computationally demanding as the size of the system increases, complexity increasing at least as a function of the square of the number of agents [201].

Since particles' behaviour in a Lagrangian model depends on their interactions with their neighbours, the differences between such models lies in the nature of those interactions, as well as the definition of the neighbourhood: it can either include all particles closer than a certain distance, or all of the particle's $N$ nearest neighbours, including the particle itself or not, depending on the model. In the boids model for example [201], particles move at a constant speed. When their distance to another particle becomes lower than the repulsion threshold, they will move away from it.

Otherwise, a particle will be attracted to its neighbours and will try to align its direction to theirs. These simple rules are sufficient to produce different group behaviours, such as swarming, parallel motion and milling, which can be found in nature [24], [202], [203]. Many models have been created based on the boids formalism [204]-[207].

Computer models not only aim at reproducing natural group behaviours, but also at understanding how these behaviours emerge from individual behaviours. But the number of models illustrates the difficulty of the issue: different models may generate similar natural-looking behaviours. Vicsek et al. [193] for instance showed that attraction and repulsion were not necessary for a self-propelled particle model to generate swarming behaviour or particles alignment. Different models may induce similar group behaviours, which means that the mere presence of patterns similar to those observed in nature or in the laboratory does not prove that the model is an accurate description of the way the organisms really interact. This problem is illustrated in Chapter 2, with 3 different models of cell growth having the same explanatory power. Hence, the validity of a model must be tested through a precise comparison with the actual behaviour of the organism, both on the scale of the individual and the group. The behaviour of the particle must be biologically plausible; for instance, its perceptions should be coherent with the sensory abilities and biological constraints of the organism. This observation has led for instance to the development of models that give the particles a blind angle (a field of view is defined for the particles, and the particles that are in the blind angle do not modify the particle's trajectory even if they are close) [22]. This approach can be extended by having the particle reproduce the same behaviour as its biological counterpart in terms of speed distribution or interaction with the environment for instance. Then, the behaviour of the simulated group should be compared to the behaviour of a similar biological group (see Chapter 3). Different local and global variables have been used to dictate the behaviour of the agent and validate the global behaviour of the model (see Table 1). Models tend to be implemented based on local rules and validated based on global behaviour, although some global variables such as the number of agents tend to be used to initialize the model. Models also allow access to variables that cannot be easily measured from watching the organisms,
allowing researchers to control their effect. For instance, field of view [74] or the pressure exerted between agents [208].

Table 1: Most common variables used to describe collective motion at the local and global scale, for different types of species.

|  |  | cells | insects | fish | birds | humans |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Individual variables | Agent's size |  |  |  |  | [169], [209] |
|  | Agent's speed | $\begin{aligned} & {[210],} \\ & {[211]} \end{aligned}$ | $\begin{aligned} & {[74], \quad[75],} \\ & {[212]} \end{aligned}$ | [21], [65] | [213] | $\begin{aligned} & {[169], \quad[208],} \\ & {[209]} \end{aligned}$ |
|  | Acceleration |  | [74] |  |  | [169], [208] |
|  | Angular speed |  | [74], [212] | [21], [65] | [214] |  |
|  | Angle relative to the group | [210] | [75] |  | [214] |  |
|  | Wall following/ avoidance |  |  | [21], [65] |  | [169], [208] |
|  | Centrality |  | [74] | [21] | [214] |  |
| Global variables | Number | [210] | [75], [215] | [21] |  |  |
|  | Group shape |  |  | [21] | [214] |  |
|  | Distance to the Nth nearest neighbour / local density | [211] | [74], [75] | [22] | $\begin{aligned} & \text { [82], } \\ & {[213],} \\ & {[214]} \end{aligned}$ | [209] |
|  | Polarization |  | [75], [212] | [21], [65] | $\begin{aligned} & \text { [214], } \\ & {[216]} \end{aligned}$ | [208] |
|  | Group Speed |  | [75], [215] |  |  |  |
|  | Cohesion |  |  | [21] | [216] |  |

### 1.7. Chapters organisation and objectives

In this thesis, I study different organisms that exist in large groups in nature. Using semi-automated detection methods, I track their movements among smaller groups, as they are observed in controlled environments under different conditions. I analyse the systematic errors produced by these metrologies and demonstrate how they can be corrected; all errors do not need to be corrected to deduce the actual model of the organism's behaviour. From the movements of all the organisms, I deduce different descriptors of individual and global behaviours. Statistical analysis of these variables then allows me to state what impact the changes in the environment may have on the organisms, in the experimental setting. I then present simulated models in which the organisms behave in ways that match the data gathered pertaining to individual behaviour. Through these models, I test the applicability of the results to larger populations or longer time scales (see Table 2).

In Chapter 2, this approach is applied to dental pulp progenitor (DPP) stem cells and their division process. If a population of cells is injected with a drug, it is important to know how the drug will propagate through the tissue as the cells divide, which requires following the lineage of the cells. The propagation of the drug can be followed with the use of fluorescent quantum dots (QD). Through flow cytometry, it is possible to know the distribution of QD among the population at different times, but not to follow their propagation through the lineage. Another method is to directly observe the cell divisions through microscopy, but this can only be done at a much smaller scale. Moreover, whether a cell will divide depends on its environment and the density of other mesenchymal cells around it. This experiment combines the large data sets of flow cytometry with the local behaviours observed through microscopy. I explain the compromises associated with the parameters of recording for the microscopy, and the different methods I attempted to follow the cells throughout the recordings, as well as the reasons for their imprecision. I then describe the process used to establish computer models that would fall within the margin of error of the data and match the data from the flow cytometer.

In Chapter 3, I study juvenile sea bass. Sea bass usually form shoals that go from a few dozens of individuals to many thousands [217], and this shoaling behaviour could be
disturbed by ocean acidification (OA). In this case, it is not possible to observe the behaviour of the organism in a natural environment, as said environment (the acidified ocean, as it is expected to be 100 years from now) does not exist yet: experimental observation and simulation are the only ways to determine whether OA could have an impact on sea bass shoaling behaviour in the distant future. I conclude that neither the individual nor the collective behaviour of sea bass are impacted by OA in the experimental setting, and I discuss the limits of the approach regarding other aspects of sea bass behaviour: the model can only extrapolate from the behaviours that can be observed in an experimental setting, so we cannot conclude over the impact that ocean acidification could have on other sea bass behaviours.

In Chapter 4, I study human foraging behaviour. Humans are used to working in groups of different sizes and various compositions, and the experiment focused on the impact of the size of the groups and the traits of the individuals, such as their personality, on the strategy of the subjects, the organisation of the group and its overall efficiency. I conclude that subjects draw information from others' behaviour, that the size of the group determines the best strategy, and that personality does not impact cooperation on this task. I discuss the possibility of social amplification in the choice of foraging patches, as well as the possibility of modelling the experiment by a multi-armed bandit system (MAB). With a MAB model, it would be possible to quantify the importance of information gathering in the foraging task and to observe the impact of different strategies.

Chapter 5 summarises the conclusions from the previous chapters, and expands on the possible applications of the interaction of studies in the wild, lab studies, and simulated models for our understanding of collective behaviour, suggesting a framework for such interdisciplinary studies.

Table 2: Main aspects of the systems studied in the following chapters, and of the methods used to study them.

| Individual | Cell | Sea bass | Human |
| :---: | :---: | :---: | :---: |
| Interaction | Growth requires a particular density | Form shoals | Draw information from each other and follow each other |
| Tracking instrument | Different methods of automatic tracking were attempted, none were suited to DPP | idTracker | GPS |
| Time scale | One recording every 30 minutes | 24 images per second | One data point per second |
| Environment | Petri dish (rather than tissue) | Arena | Artificial setting in which the subjects were following instructions |
| Sources of error in the data | Different sources of noise, which could not all be corrected | Low video quality | Noise in the GPS, some data were not usable |
| Individual variables | IMT, QD, speed | Speed, time not moving, thigmotaxis, exploration speed | Speed, score, switching frequency... |
| Group variables | Density | Number of neighbours, distance to nearest neighbour | Number of neighbours, distance to nearest neighbour |
| Simulation | Three variants of a growth model | No-interaction model | One-armed bandit models |
| Perturbation | Phototoxicity | Normal water and ocean acidification | Patches running out |

# CHAPTER 2 - DOUBLE PLATFORM ANALYSIS OF MESENCHYMAL STEM CELLS GROWTH 

### 2.1. Introduction

### 2.1.1. Stem cells

Stem cells are cells that have the capacity to self-renew and the ability to generate differentiated cells [218]. Most somatic cells can divide a limited number of times when grown in vitro (doubling up to 80 times) before they become senescent. Stem cells on the other hand display an unlimited proliferative capacity in vitro, and are able to maintain their cell line throughout the life of the animal in vivo [218].

Stem cells also have the ability to generate differentiated cells. They can go through asymmetric division, in which case one daughter will be a stem cell identical to the mother, and the other will be a differentiated cell; or they can divide symmetrically, in which case both daughter cells are identical, either both stem cells or both differentiated [219]. Stem cells have different levels of potency, which determine the number of cell types they can differentiate into. The potency of the cell depends on its source: embryonic or adult stem cells.

Embryonic stem cells are any kind of stem cells found at the embryonic stage, but are usually considered to be the cells from the inner cell mass of a blastocyst (a precursor to an embryo four to five days after fertilisation); a blastocyst contains about 100 cells [220]. Early embryonic stem cells (from blastomeres and blastocysts) are the only cells that can be qualified as totipotent, cells that can produce an entire viable organism [221]. The umbilical cord blood and the amniotic fluid are also an important source of stem cells early in life, that can be used to reconstitute bone marrow as part of treatment for various blood cancers [222].

Pluripotent stem cells are cells that are able to differentiate into cells of any of the three germ layers (endoderm, mesoderm and ectoderm); the most potent cells in this category can differentiate into any of the hundreds of cell types present in the human body. At
the end of the potency spectrum are unipotent stem cells, which have the property of self-renewal but can only differentiate into one type of cell.

Because of their properties of self-renewal and differentiation, stem cells, and pluripotent stem cells in particular, have the potential to help cure many diseases, such as diabetes, neurodegenerative diseases or heart diseases [223]. There is ongoing research on the use of stem cells to grow organs [224]-[227], which would remove the need for transplants and minimise the risk of transplant rejection. Stem cells could also be used to design precise models of the human body, making it possible to design drugs adapted to every individual [228].

Pluripotent stem cells can also be generated from adult, non-stem cells. These induced pluripotent stem cells can be obtained by transducing genes associated with stem cell behaviour by the injection of a vector. Different vectors have been used successfully to achieve this result, such as retroviral systems [229], plasmids [230] or adenoviruses [231]. Another method bypasses the use of a transduction vector by using molecules that mimic the effect of the transcription factors activating the stem cell behaviour [232]. However these methods present some risk of causing tumours [233] or have a low efficiency [234], and are thus not yet an effective alternative to naturally pluripotent stem cells.

Aside from embryonic stem cells and the cells in the umbilical cord blood, pluripotent stem cells have been found in the blood, the bone marrow, the adipose tissue and the dental pulp [235]. One example of pluripotent stem cells found in the dental pulp is called dental pulp progenitor stem cells (DPPSC, or DPP), which are notably involved in the dentine regeneration process [236]. Because they are located in the dental pulp, this makes them a useful source of adult stem cells, as the dental pulp can be extracted from discarded teeth. DPP can also be extracted from primary teeth [237]. In their natural environment, they differentiate into odontoblasts, the cells that create the dentin within the tooth [238]. But they can also differentiate into other types of cells, such as bone cells [239], muscle cells and brain tissue [240]. DPP have been shown to be useful in repairing neuronal damage [241].

Many of the possible applications of stem cells rely on their ability to differentiate, so understanding the mechanism that underlies this differentiation could help control this
process. The repartition of the mother cell's endosomes between its daughters has been linked to the differentiation process: by distributing the regulators of self-renewal asymmetrically before the division process, a cell can ensure that only one of its daughter maintains the mother's property of self-renewal [242]. It has also been suggested that a cell can distribute its damaged proteins asymmetrically so that they all become part of one cell, and the other one will remain healthy and maintain the lineage [243], [244].

Different properties of the cells can be extracted through direct observation via microscopy, such as motility, morphology and proliferation, as well as the division process itself [245]-[247]. It is also possible, by injecting the cells with fluorescent markers, to follow the content of a cell throughout its lineage [248]. Thus live cell imaging can give us an insight of what is happening at the local scale [249]. But understanding the behaviour of the whole cell population would require the study of a number of cells much greater than what can be observed via live-cell imaging [94]. Such large amounts of cells are better suited for a different technology: flow cytometry. The next paragraphs explain the concepts, applications and limits of fluorescence tagging, live cell imaging and flow cytometry.

### 2.1.2. Methods of observation

### 2.1.2.1 Fluorescence tagging

Fluorescent substances, or fluorophores, absorb radiations from a certain band of wavelengths and then emit radiations in a different one [250]. Contrary to phosphorescence, there is virtually no delay between the absorption and emission [251]. When a photon interact with an electron of a fluorescent molecule, the electron gets moved to a higher energy state, then returns to its ground state by the emission of a photon [252]. Although there might be some overlap between the excitation spectrum and the emission spectrum of a fluorescent molecule, without additional energy, the emitted photon must have a longer wavelength than the one absorbed. The difference between the most absorbed wavelength and the most emitted is called Stokes' shift (see Figure 2).


Figure 2: The absorption and emission spectra of Rhodamine 6G, a fluorescent molecule. The graph highlights the Stokes' shift between them, which is $\sim 25 \mathrm{~nm}$. Source: Wikimedia Foundation [30].

The ratio of the number of photons emitted to the number of photons received is called the quantum yield; it can reach values of almost 1 (the molecule emits a photon for every photon received) for rhodamine 101 [253], [254], although the quantum yield can decrease over time, a phenomenon called fluorescent quenching. The fluorophore can also lose its ability to fluoresce by exposure to light; this is called photobleaching. The likelihood of photobleaching depends on the molecule, some fluorophores start bleaching after emitting a few photons only, while others can remain fluorescent for millions of cycles before bleaching [255].

Those properties of fluorophores make them good biomarkers. Hulka et al. [256] defines biomarkers as "cellular, biochemical, or molecular alterations that are measurable in biological media, such as human tissues, cells or fluids". The quality of an imaging biomarker depends on it being highly detectable with a process that returns a low rate of false positive, and on how well it binds to the observation target compared to any other object in the area of observation. If the sample or the organism is lit in a spectrum that the organism's molecules do not scatter, then only the fluorophores will appear on the image. Thus, injecting fluorophores that bind to the observation target will highlight the location of those targets or their quantity within a cell or a tissue
sample when they are observed with an instrument such as a fluorescence microscope or a flow cytometer (see below).

One method for fluorescence tagging relies on fluorescent proteins. Such proteins occur naturally in diverse species, including fish species (eels, seahorses, anglerfish, pufferfish...), jellyfish, plankton [257], crustaceans [258], spiders [259], butterflies [260], birds [261], flowers [262] and corals [263]. This diversity is explained by the various evolutionary advantages fluorescence provides. For corals, fluorescence is thought to be a mean to attract animal symbionts [264] and to enhance the photosynthesis reaction in algae symbionts [263] (although arguments have been advanced against this explanation: the light provided by the fluorescent coral would be too low in certain wavelengths to help with photosynthesis, and some fluorescent corals do not have vegetal symbionts [265]). They also protect the coral against oxidation [266] and excessive sunlight [267]. As corals are fluorescent, the animal species that lives in their midst gain a camouflage advantage from being fluorescent too. In some species, fluorescence can be a way of luring preys or attracting mates [257], [261]. Fluorescence having evolved independently in different species, these species express it through different genes, coding for different proteins and fluorophores, each with their own properties in terms of excitation and emission spectra [265].

In 1962, Shimomura et al. [268] isolated such a protein called GFP (for green fluorescent protein) from a species of jellyfish, Aequorea Victoria. The gene coding this protein was isolated 30 years later [269]. This paved the way for the synthesis of numerous variants of GFP and to the discovery of many other natural fluorescent proteins. Those fluorescent proteins can absorb and emit light at wavelengths all over the visible spectrum (see Figure 3); examples can be found in [265], with their excitation and emission peaks, as well as their quantum yield. Some fluorescent proteins introduced new properties such as photoactivation (their fluorescence properties can be changed by irradiating them with light under a specific wavelength, intensity and duration [270]).


Figure 3: Petri dish seeded with bacterial colonies that fluoresce in various colours, so that they represent a beach scene in San Diego. The proteins used for the fluorescence are variants of GFP and of the red-fluorescent coral protein drFP583 (which was originally found in discosoma sea anemones): BFP, mTFP1, Emerald, Citrine, mOrange, mApple, mCherry and mGrape. Artwork by Nathan Shaner, photography by Paul Steinbach, created in the lab of Roger Tsien in 2006.

The genes responsible for these fluorescent proteins can be inserted into many proteins without altering their function, making them fluorescent variants of those proteins. This technique allows in particular the creation of fluorescent antibodies [271], which will not only attach themselves to a particular protein or virus but also make them fluorescent. Observing tissues with the right wavelength, this technique reveals a map of the antibody's target. Advantages of this method compared to radioisotopes (which used to be the most common labelling method [272]) is that fluorescent proteins are safe, and
they do not decay. They are however sensitive to photobleaching, which limits their use for long term observation. Because the genes will keep producing the fluorescent protein, the signal does not get diluted within the sample. The diversity of fluorescent proteins means that different targets can be highlighted at the same time, with no possible confusion between them even when they occupy the same position (see Figure 4).


Figure 4: Endothelial cells observed under fluorescence microscopy. Because different fluorescent proteins have been used, the microtubules (marked green by an antibody bound to FITC) can clearly be distinguished from the nuclei (blue, stained with DAPI) and the actin filaments (marked red with a protein bound to TRITC). Source: Wikimedia Commons [30].

Some molecules in the human body, such as pyridine and flavine nucleotides, make the cells fluorescent in blue and green wavelengths respectively [273], [274]. This
phenomenon of autofluorescence can be useful when the purpose of the study is to observe those molecules [275], [276], but autofluorescence is generally a hindrance for fluorescence studies. One way around this problem is to use photobleaching in order to quench the autofluorescence before labelling the tissue with fluorescent molecules. Another method is to observe the sample in a wavelength that does not belong to the autofluorescent molecules emission spectra. When the autofluorescence signal cannot be removed, the signal from the biomarker needs to be more intense in order to be detected. As the signal gets diluted, detecting the target can become harder, although the biomarker can be tracked for a longer time if the presence of autofluorescence is accounted for in the model [277].

Another type of fluorophore that can be attached to a protein is the semiconductor nanoparticle known as the quantum dot (QD) [278]. QD are small semiconductor particles (for example, cadmium selenide QD, which are the most commonly used type, range in size from 1 to 25 nm [279]-[281]) that can emit light at specific frequencies when excited by electricity or light; the precise frequency depends on the shape and size of the QD [282]. They have applications in quantum computing [283], photovoltaic devices [284], the design of light emitting diodes [285] and new types of TV and mobile phone screens [286], and in biology [287].

The most common types of QD contain toxic heavy metals, such as cadmium or lead [288], [289]. Under UV light, those nanoparticles can dissolve, in a process known as photolysis (photons breaking covalent bonds). When the nanoparticles are in a biological sample, photolysis releases the heavy metals within the sample. They can then react with the oxygen atoms in the cells' molecules, which can lead to DNA damage, changes in cell motility, cytotoxicity, apoptosis and cancer [290]. This process can be limited by the use of a non-toxic semiconductor coating, such as zinc sulphide [291]-[294]. Beside isolating the core, the coating protects it from photobleaching, and enhances the photoluminescence of the QD [295]-[297].

The QD is made soluble in aqueous solutions by adding an organic coating to them. In order to be suited for biological applications, the QD must be linked to biomolecules [287] (see Figure 5). This is done in such a way that the conjugate keeps the properties of the biological compound: the QD is a biologically non-functional probe that is meant
to have minimal impact on the local environment. Conversely, the conjugation should not interfere with the fluorescent properties of the QD. Finally, the QD should only get attached to the intended biomolecule, so that the fluorescent signal becomes a proxy for the presence of the molecule; the resulting nanoparticle can then enter a cell through endocytosis [94], transfection, or peptide-mediated delivery [287]. Methods exist to minimise the risk of unintended binding during the conjugation phase of the design of the QD [298], [299]. This toxicity problem does mean however that QD need to be recycled in a specific way [300], which could be a problem for their potential widespread applications [301]. In order to avoid the problems caused by the use of heavy-metals, alternatives are being developed that use peptides as the semiconductor [302].


Figure 5: Structure of a quantum dot nanoparticle. The core is a semiconductor nano-crystal, which is covered by a shell-coating. The organic coating makes the nanoparticle soluble in aqueous solution. Biologically, the QD has the properties of the biomolecules attached to it (here, biotin). Image taken from [303].

Compared to fluorescent proteins, QD emit light in a narrower and symmetric spectrum. They are up to 20 times brighter than fluorescent proteins, and have a longer lifetime, as they are up to 100 times less susceptible to photobleaching [304]. The properties of QD (emission wavelength, narrowness of the emission profile, signal strength) all depend on the morphology of the QD, which can be precisely controlled during its synthesis: size and composition of the semiconductor core, composition of the shell and surface coating. Hence, it is possible to synthesise QD that can absorb and emit in any narrow waveband between ultraviolet and near-infrared, as long as the excitation wavelength is shorter than the emission wavelength [305], [306]. It is for instance possible to design
different types of nanoparticles that would be excited by the same wavelengths, but emit at different wavelengths from each other and that would be attracted to different biomolecules. Lighting a sample seeded with both nanoparticle populations would then reveal both types of biomolecules at once in different colours, which not only simplifies the imaging process but also mitigates the problem of phototoxicity, since the sample needs to be lit only once [307]. This method can be used to visualise either the molecules the fluorescent particles target or different circulation systems at once [308].

Both fluorescent proteins and QD are used in biology by modifying a biomolecule so it becomes fluorescent. Thus, observing a cell or a sample at the proper wavelength reveals the amount of these molecules present in the object under observation, as well as their distribution. In particular, it makes it possible to follow the circulation of the molecule throughout the body. This has applications to the understanding of embryogenesis [309] and cancer metastasis [310] for instance. Fluorescent proteins can be used to label and track whole organisms, cells, organelles, DNA/RNA, down to a protein. By binding to DNA, fluorescent proteins can indicate at what stage a cell is within its cell cycle [311]; knowing the proportion of cells at any stage of their cycle can give researchers information on their metabolism.

Fluorescence imaging has been used in 3-D environment [312]. Moreover, using three types of QD each designed to target a different kind of human tumour, Cui et al. [313] were able to use fluorescence imaging in vivo. Different parts of the animal became fluorescent and bright enough to be seen through the skin. The tumours were not present on the injection sites only, which highlights this technique's potential to provide a new method to detect the progression of tumours. Yet, the problem of nanotoxicity remains an obstacle to using this technique for therapeutic purposes.

Some fluorescent molecules are said to be impermeant: they cannot cross the membranes of cells. Thus, their presence in a cell can be considered as an indicator that the membrane is compromised and that the cell has gone through necrosis. Presence of impermeant dyes in mitochondria would however be a sign of apoptosis, so using impermeant dyes that can target mitochondria and other cytoplasmic biomolecules can allow researchers to determine whether a cell died through necrosis or apoptosis [314]. It is however possible to permeabilise a cell using acetone, which makes it possible to
use impermeant dyes to label the inside of cells. A similar method is fixation, which can be done with paraformaldehyde [315]; this can be a mean to stabilize a cell.

Since fluorophores are activated by specific wavelengths, and emit light at specific wavelengths, it is possible to use the emission from one type of fluorophore to activate another. This method is called fluorescence (or Förster) resonance energy transfer (FRET) [316]. Since fluorophores are not bright enough to excite another molecule beyond a close proximity, the colour corresponding to the emission wavelength of the second fluorophore is indicative of a close proximity between the two types of fluorophores. FRET can thus be used to measure the interactions between two fluorescent molecules. However, it is also possible to design a protein that carries both fluorophores of a FRET pair. In this case, the colour of the second fluorophore is indicative of a change of shape in the molecule [317].

Besides imaging, applications of fluorescence to biology include surgery and gene sequencing. In fluorescence image-guided surgery, fluorescence can be used to label a tumour, making it easier for the surgeon to know precisely where to operate [318]. One method of DNA sequencing called Sanger sequencing relies on fluorescent molecules to target chain terminators in the DNA, making it easier to detect the end of a DNA fragment.

### 2.1.2.2 Microscopy

Live cell imaging consists in observing a cell population through time-lapse microscopy in order to quantify how interesting cell properties of the wider cellular population evolve through time [319]. Cells are grown in multi-well plates and snapshots of the wells are taken at regular intervals. Contrary to static imaging, the cells need to be kept healthy throughout the recording. This includes maintaining a constant pH and temperature. Temperatures of $20^{\circ} \mathrm{C}$ to $22^{\circ} \mathrm{C}$ could interrupt the cell cycle, and between $23^{\circ} \mathrm{C}$ and $35^{\circ} \mathrm{C}$, the temperature-induced stress can change the timing of the mitosis [320]. For this reason, imaging platforms designed for live cell imaging usually contain an incubation chamber, which maintains the environmental conditions in terms of temperature, pH , humidity and atmosphere composition.

It is also important that the cells be observed within enough media to allow their growth over the time of the recording. In order to reduce the risk of infection by bacteria or fungi, antibiotics are often added to the medium [321]. The medium must include the nine amino acids that cannot be synthesized by vertebrate animal cells as well as those normally synthesized by specialised cells in the body, and glutamine, as a source of nitrogen. The other components of cell medium include vitamins, salts, fatty acids, glucose and blood serum [322].

A more difficult problem is that of phototoxicity. For the observation to be successful, the cells must be clearly lit, but any amount of photons is damaging to cells [323]. This compromise must be considered in the experimental design, with regard to wavelength, light intensity or exposition time, and different imaging methods differ widely in those regards. In fluorescence microscopy in particular, light can cause the alteration of the fluorophore molecules, a phenomenon called photobleaching: the more light is projected onto the cells, the harder it will be to detect the fluorescent molecules over time [324], [325].

Because of the limitation of the field of view of the microscope, cells are typically observed in small environments, the volume of wells usually being between $100 \mu \mathrm{~L}$ and $500 \mu \mathrm{~L}$ and filled with $5 \mu \mathrm{~L}$ to $50 \mu \mathrm{~L}$ of media [326]. Since snapshots are taken at regular intervals, multiple wells of the plate are usually observed simultaneously, with different fields of view (FOV) of different wells being recorded at all times. This allows the observation of different instances of the same experimental conditions, as well as variations in conditions including a control. Thus, the imaging platform must be able to come back at regular intervals to the exact same position, and vibrations within the room should be kept at a minimum. The different images are then often processed algorithmically to extract different properties of the images and the cell phenotypes they underlie [249]. When such algorithmic solutions are possible, data can be extracted for up to a few thousand cells [327]. Some microscopes can however capture images at a high resolution in a large FOV (typically, $6 \times 6 \mathrm{~mm}$ [328]); using such a microscope, data could be extracted for hundreds of thousands of cells in one FOV.

In other words, although live cell imaging produces a large amount of data, only a small part of it typically gets analysed by the biologists, the image being reduced from a large
amount of pixels to a few visual properties [329]. For instance, detecting areas of high contrast can highlight boundaries between cells, allowing researcher to estimate the size of the population. Areas of high colour gradient are a good proxy for cell coverage, since the gradient is close to zero in the uniform background [330]. The study of cell motility requires more frequent observations than a study of the state of the system at different times, so for equal perturbation of the biology, a compromise between quality of the image and frequency of acquisition would be resolved in favour of the latter [329].

For this reason, different methods of observation have been introduced, depending on the specificities of the cell population and the research question, in order to optimise the recording process for the detection of these visual properties. Light microscopes use several lenses to magnify a specimen. The total magnification is the product of the magnification of the individual lenses; this is the scale ratio between the image and the specimen under observation. By adding lenses, the image can be magnified to any degree, but will not necessarily be usable: magnifying a blurry image will only return a larger blurry image. For this reason, the most important property of a microscope is not its magnification, but rather its resolution, i.e. its ability to discriminate between two closely located objects. This number is a function of the wavelength of the light, the refractive index of the medium between the specimen and the objective lens, and the half angle of the cone of light entering the objective lens. These factors make the maximum resolution of a light microscope $0.2 \mu \mathrm{~m}$ [331]. This is larger than the smallest cells [332], and close to the size of an endosome [333]. For comparison, the size of a DPP cell ranges from $20 \mu \mathrm{~m}$ to $200 \mu \mathrm{~m}$ in diameter [334].

In a bright field microscope, white light is projected from below the sample, which is observed from above. The light passes through a condenser lens which focuses the light onto the specimen to be observed. The image is magnified by an objective lens and reaches either a camera or an eye-piece (see Figure 6a). This method has low contrast, as cells are translucent. For this reason, cells observed through bright field microscopy require staining, which can interfere with cell behaviour.

Since cells are translucent in the absence of a colouring dye, there is little variation in amplitude between the rays that hit or do not hit the cells. The refraction index and the
thickness, though, might vary between the different parts of a cell and well areas that do not contain cells. For this reason, even though the light waves that hit the specimen are plane waves with uniform phases, the refracted wave presents small variations in phase depending on the index and the thickness of the material through which the wave went: the wave is phase-modulated.

Phase contrast microscopy [335] relies on the same principle as bright-field microscopy, except for the presence of an annular diaphragm and a phase plate (see Figure 6b). Light shone from a lamp passes through an annular diaphragm that only lets a thin cylinder through, which is condensed as a cone by the condenser lens. As with the bright field microscope, the specimen is set at the focal point of this cone. The light that goes through the specimen is slightly refracted to a variable degree depending on the refractive index and the thickness of the specimen, whereas the light that does not go through the specimen (direct light) is not refracted. The refraction changes the angle with which the light hits the objective lens. Above this lens is a phase plate, which slightly refracts the direct light, and refracts the light that went through the specimen by a larger angle. Those different lights go through the projection lens where they are focused together onto the image plane. Thus, the different lights are out of phase with each other, and the amplitude of the resulting ray of light depends of the initial refraction by the specimen. In summary, phase contrast microscopy allows to translate phase differences into amplitude differences, making them visible without requiring the use of a staining agent.

In fluorescence microscopy (Figure 6c), the light that goes through the collector lens then goes through an excitation filter which only lets through a narrow waveband, such as the excitation wavelength of the compound (some models ensure a limited waveband by using a laser as the light source). This short wavelength light is then reflected by a dichroic mirror through the objective lens and onto the specimen. The fluorescent compounds in the specimen then emit light at the fluorescence wavelength through the objective lens and the dichroic mirror (as the fluorescence wavelength is higher than the mirror threshold), which does not reflect this wavelength. Thus, only the light emitted by the fluorescent compound goes through the projection lens to be focused on the image plane. Fluorescence microscopy requires a brighter light than bright field microscopy, which makes the problems of phototoxicity greater. For this reason, a
compromise must be found between the intensity of the light, the frequency of sampling and the duration of the observation. More generally, the problem of phototoxicity must be avoided by optimising the system to get the best signal to noise ratio, while minimising the total light cast over the specimen over time (see [336] for a review).

Some cells naturally contain molecules that become fluorescent under certain types of radiations [275]. This makes fluorescence microscopy the ideal method of observation if those are the cells that are to be observed, but this autofluorescence can be a hindrance if they are not, or if other types of cells emit fluorescence at the same wavelength. If the fluorescence or excitation wavelength of the different compounds do not cover most of the visible spectrum, this problem can be circumvented by using different wavelengths, and injecting the cells with fluorescent labels for a different wavelength if necessary [337]. Artificial labels might be much brighter than the autofluorescence, in which case the image analysis is possible, but requires some image processing to remove the background fluorescence.

Because of the similarity of the three technologies, imaging platforms often offer configurations for bright-field, phase-contrast and fluorescence microscopy.


Figure 6: Principle of a bright field microscope (a), a phase-contrast microscope (b) and a fluorescence microscope (c). Image from [331].

### 2.1.2.3 Flow cytometry

Cytometry is the analysis of the physical and chemical properties of cells [314]. Flow cytometry is a technique that allows doing such analysis of individual cells, processing high numbers of cells in a short time. The cells from a sample are organised into a line that intersects orthogonally with a laser; through the scattering of the laser, important information can be inferred about the cells. Different methods exist to push the sample toward the flow chamber, such as pressure differential, syringe pumps and peristaltic pump systems (see Figure 7). In all methods, the idea is to push a constant amount of fluids through the flow chamber. It is thus important that the cells are in suspension through the sample: adherent cells would not be sent with the sample fluid to the flow chamber. Those cells need to be detached from the cell culture dish wall by using trypsin [338] or ethylenediaminetetraacetic acid (EDTA) [339], [340].


Figure 7: Three delivery methods for flow cytometry. a: pressure differential: air pressure pushes the sample toward the cytometer. $\mathbf{b}$ : the sample is pushed by syringes. $\mathbf{c}$ : the sample is pumped toward the cytometer by a peristaltic pump. Source: [341].

The flow chamber is the part of the system where the cells get positioned in a line. This is done through a technique called hydrodynamic focusing (see Figure 8). The sample is injected into the centre of a different "sheath fluid" (usually saline [342]) which is moving faster; this focuses the sample into a narrow core stream ( $50-150 \mu \mathrm{~m}$ in
diameter [343]). The Reynolds number $R e$ for an object going through a fluid is defined by Equation 1 :

$$
\begin{equation*}
R e=\frac{\rho \times s \times L}{\mu} \tag{Equation 1}
\end{equation*}
$$

$\rho$ is the density of the fluid, $s$ the speed of the sample, $L$ the diameter of the tube and $\mu$ the dynamic viscosity of the fluid. The flow is said to be laminar when the Reynolds number is less than 2300 [344]; this means that the fluids flow in parallel layers with no disruption between them. For a given flow cytometer and sheath fluid, $\rho, L$ and $\mu$ are all fixed: the only parameter that can ensure that the flow is laminar is the speed of the sample, which depends on the differential pressure between the core stream and the sheath stream [314]. Acoustic-assisted hydrodynamic focusing is an alternative to hydrodynamic focusing in which the sample is focused by sound waves. This minimises the impact of pressure on the cell alignment [345].


Figure 8: a: Flow chamber without hydrodynamic focusing; the cells move randomly through the tube, and their number at the measurement point cannot be predicted. b: Flow chamber with hydrodynamic focusing; the core stream is narrow enough that only one cell can pass through the measurement point at a time. Source: [342].

The measurement point (or interrogation point) is the point where the laser intersects the core stream. A laser (light amplification by stimulated emission of radiation) is a device that excites atoms through stimulated emission, which means that the photons emitted by the atoms share the same properties in terms of wavelength, direction, phase and polarization. Because lasers are focused, they are brighter than spontaneous emission sources. Furthermore, the energy output of the laser is constant, so any difference in the way the light is scattered at the measurement point is a consequence of the optical properties of the object going through the measurement point, rather than the laser.

The principle of the laser works at any wavelength, and existing lasers cover a range that goes from the X-rays [346] to the microwaves [347]; the wavelengths used in flow cytometry range from the ultraviolet to the infrared [314]. Objects whose dimensions are close to the laser wavelength scatter the light in a stochastic manner. The choice of laser is thus particularly important for the analysis of the smallest objects [314].

The power of the cross section of a laser is distributed along a Gaussian [348], so it does not have clearly defined edges. The diameter of the lasers used in flow cytometry can be considered to be approximately $1-2 \mathrm{~mm}$; this is much wider than the size of a cell, which is why the laser needs to be focused through spherical lenses to a diameter of 60 $\mu \mathrm{m}$ [343]. Shapiro [314] showed that a typical flow cytometry laser hitting a $10 \mu \mathrm{~m}$ bacterium would only illuminate it with $10 \%$ of its power; that number becomes $1 \%$ for a $1 \mu \mathrm{~m}$ bacterium, and $0.002 \%$ for a $0.1 \mu \mathrm{~m}$ virus.

The laser is scattered differently depending on whether a cell is at the measurement point. The forward scatter detector (FSC) is located on the trajectory of the laser behind the measurement point. To ensure that only the scattered light is received by the FSC, an obscuration bar is placed in the direct path of the laser. The dimensions and the placement of the obscuration bar (or beam blocker) determine the minimum scatter angle that can be detected. The maximum angle can be decided by the operator through the use of an iris [314]. The range of the angles that define the FSC usually goes from $0.5^{\circ}$ to $2^{\circ}$ [349]. The collection optics of the FSC is shown in Figure 9.


Figure 9: Collection optics for forward scatter signal. a: side view. b: top view. Source: [314]

A cell scatters the light of the laser more than the saline that contains it, so the bigger the cell, the more scattering. The height of the event is the maximum signal detected, and the width is the time the signal remains above a given threshold; the integral of the signal (or area) gives an idea of the size of the cell [350]. This signal should however not be used as such, as the FSC also depends on the difference of refractive index between the particle and the fluid, the presence of light absorbing substances inside the cell, as well as the design of the measurement system [314]. Typically, dead cells tend to be larger than live cells, but since their outer membrane is compromised, more of the medium is present within the cell, hence a similarity of refracting index between the dead cell and its medium. This leads to the FSC for the cell being lower than that of a live cell, despite the cell being larger: in such circumstances, the FSC would be a better indicator of cell viability than size [351]. The FSC can instead be used as an event detector.

The forward scattered light depends on the refractive index of the particle going through the measurement point, but cells at the measurement point actually scatter light in all directions. The level of scattering depends on the granularity of the particle. The side scatter signal (SSC) is obtained from the light emitted in a direction orthogonal to both the flow and the laser. This light is focused through a lens, and then reaches a dichroic mirror. The mirror reflects wavelengths close to that of the laser, which correspond to the scattered signal. This light then goes through a filter that selects the wavelength of
the laser, and then through a photoreceptor, such as a photomultiplier tube [314]. This is the first part of the SSC (SSC-A).

Some light goes through the dichroic mirror. Since lasers are monochromatic, this light does not come from the laser. It is the fluorescent signal; it can be autofluorescence, or it can come from fluorescent markers with which the cell was labelled. Conforming to Stokes' law, the wavelengths of the fluorescent signal are longer than that of the laser. Behind the first dichroic mirror is a series of dichroic mirrors each of them letting a shorter waveband through and reflecting the lowest wavelengths they receive, each to a filter and a photodetector. Because of the filtering, each photoreceptor measures the fluorescence in a different wavelength, which usually corresponds to a different fluorescent marker. The system that collects the FSC and the SSC is presented in Figure 10.


Figure 10: Simplified model of a flow cytometer. Through hydrodynamic focusing, cells arrive one by one at the measurement point, where they are illuminated by the laser. The direct light from the laser is blocked by the obscuration bar. The forward scattered light is filtered before reaching the forward scatter detector. Part of the cell fluorescence and the side scattered light is focused by a lens and then goes through a series of dichroic mirrors. Each mirror reflects part of the light toward a filter and a photosensor. Adapted from [352].

Although FSC is not a reliable way to determine the size of a cell, cell size can be measured accurately in most flow cytometers. This can be done with a Coulter counter [353]. The technique relies on the difference of conductivity between cells and the saline solution (which is why the medium cannot be water for this application [314]). Although the intracellular fluid is an ionic solution and thus conductive, the membrane is made of lipids, which makes the cell resistive. In a Coulter counter, two chambers filled with saline are joined by a small orifice. One chamber contains the anode and the other the cathode of a dipole. When the orifice is filled with saline, there is no difference of electric potential between the poles. When a cell goes through the orifice however, the volume of conductive saline decreases and the resistance of the orifice increases, which can be measured as a difference in tension, which is a good proxy for the size of the cell (better than the FSC).

Most flow cytometers allows selecting the data that needs to be collected through gating [343]. The researcher can delineate regions in the multidimensional space defined by the different measurements made by the flow cytometer: FSC, SSC and the various fluorescence channels. Those regions define the properties that describe a type of particles. For instance, cells of a certain size (as estimated by their forward scatter), that emit a certain amount of green fluorescence. Those regions can be redefined throughout the experiment, as particles are being analysed and the different scatter plots describing the experiment are being drawn, so that they correctly describe the different clusters of data in those scatter plots. Often, one region describes the debris: those are the particles that are much smaller than expected (or rather, their FSC is lower) and have low fluorescence and side-scattering. Another region of interest is that of the doublets: despite the use of trypsin or EDTA, two cells might still be attached to each other. Those doublets can be identified by their FSC. Debris and doublets can be simply removed from the analysis by setting up a gate, which defines the different regions that should be included in the analysis. Gates are defined by logical operations on regions: combinations of AND, OR and NOT. Gating is a simple way to describe the cell population, as the software will indicate in real time the proportion of cells in every region defined by the experimenter, even as they redefine the boundaries of the regions. In most applications, the fluids and the particles are sent to waste after leaving the measurement point. In some flow cytometers though, particles can be sorted according
to their forward and side scatter, a method called fluorescence-activated cell sorting (FACS) [354]. The idea of FACS is that the stream is broken into droplets that contain one particle or less. An electric charge is applied to those droplets based on what was measured at the measurement point, and the droplets are sorted based on that electric charge.

In more detail, vibrations from a piezoelectric or electromagnetic transducer make the nozzle vibrate. This breaks the stream into droplets at predictable intervals: droplets always break at the same point in the stream, which is called "the breakoff point". It is usually located 3 mm below the measurement point, and the space between two consecutive droplets is around $250-600 \mu \mathrm{~m}$ [314]; the actual dimensions depend on the speed and the diameter of the stream, as well as the frequency of the vibration. The diameter of a drop is close to that of the nozzle, usually $70 \mu \mathrm{~m}$ (because of the risk of doublets, the nozzle should be three times the diameter of the cell to minimise the risk of clogging [355]. For particles to be properly sorted, one drop should never contain more than one particle).

Voltage is applied to the stream at the breakoff point by an electrical charging ring. This charge depends on the gated region the particle belongs to, based on the signal that was received by the different photosensors when the particle was at the measurement point. It is based on this voltage that the particle is sorted; depending on the device, up to six magnitudes of charge can be distinguished [356]. Because the surrounding droplets can receive the same charge, it is recommended that there is no more than 1 particle per 5 drops [314], [355]. The particle frequency is a function of the concentration of particles in the sample and of the speed of the stream. As a droplet falls, it passes between two deflecting plates that are generating an electric current. The droplet is pulled toward the plate with the opposite charge as a function of its own electric charge; uncharged droplets fall vertically to the waste container.

The performance of the sorting is measured by its yield and purity. The yield is the ratio of the cells collected by the number of cells of the desired type in the sample; it is the complementary of the proportion of false negatives. The purity is the proportion of cells that have been properly sorted; it is the complementary of the proportion of false positives. As with all kinds of classification, a compromise must be found between
maximising the yield or the purity, since the simplest way to maximise the purity is to sort for one cell, and the simplest way to maximise the yield is not to sort at all. The best compromise depends on the application. If the desired type of cell is very rare within the sample, the absolute yield will be very low; if strict selection criteria are applied to this sample, it is possible that no cells will be harvested. So in this case, it might be better to resolve the compromise in favour of the yield.

A classic application of cell sorting is the sorting of blood cells by type (for instance, to collect red blood cells) [357], [358]. Another use of cell sorting is for sperm: since Y chromosomes are shorter than X chromosomes, sperm stained with DNA marker will be more fluorescent if the sperm contains an X chromosome than a Y . This distinction makes it possible to sort sperm by sexual chromosome, thus deciding the sex of a fetus [359]. This technique is widely used in farming [360], but restricted in the case of human use [361]. Finally, cell sorting makes it possible to select a population of cells based on certain characteristics, and grow them. This makes it possible to determine the inheritability of the characteristic [362], and if so, to engineer a new population specially selected for that characteristic [363], [364]. Using populations that have been specifically chosen for a trait ensures that all the cells are in the same condition and have the same properties. Thus, the state of the population after an experimental treatment does not depend on variations within the initial cell population [365].

Flow cytometry also makes it possible to analyse assays of different cells populations at once, through a method called cell barcoding [366]. The idea is that fluorescent particles can be used, not to highlight an intrinsic structure or property of the cell, but rather to denote its origin. As cell cultures are prepared under different conditions and with the various fluorescent molecules required to observe the cell characteristics under study through flow cytometry, other fluorescent molecules are added at various concentrations as a barcode for the cell. Then, all cultures can be analysed as one sample. Just by observing the channels associated with the barcode, the origin of each individual cell can be known, as well as the treatment it received. If three different concentrations of three different dyes are used for the barcode for example, the number of treatments that can be discriminated is $3^{3}=27$. As flow cytometry experiment can take several hours [367], cell bar-coding can make the most complex experiments viable timewise.

Besides cell sorting, flow cytometry has found diverse applications in medicine and cell biology. It can be used to diagnose cancer [368], [369] and to monitor the progression of an HIV infection [370]. As explained earlier, because of the difference of refractive index between dead and viable cells, flow cytometry is a useful tool for viability studies [371], [372]. Since the amount of DNA changes through the cell cycle, it is possible to use flow cytometry and a DNA stain to determine the proportion of cells at different stages of their cell cycle [311], [373], [374]. This can give researchers valuable information on the cell cycle [375] and how different treatments can upset it [376].

### 2.1.2.4 Cross-platform analysis

Microscopy can give information for a small area only. At a density such that 200 cells are visible in the field of view (as we have with the microscope used for this chapter), 100 fields of view would need to be observed to capture images of 20000 cells (a number of observations frequently chosen in a flow cytometry observations [377][382]). Since it takes around 5 seconds to record a field of view through bright field microscopy and 10 seconds with fluorescence microscopy, it would take 25 minutes to gather pictures of this many fields of view. Thus, the aggregation of those fields of view would be enough to collect data on as many cells as with flow cytometry, with a frequency superior to 2 per hour, which would be sufficient to track the cells. But such a large amount of data could not be processed manually [383]; processing of large amounts of cells would only be possible for cells that lend themselves to automated detection: cells with regular shape and size, with clear boundaries and low speed. Furthermore, as cells are observed in a group rather than individually as they are with a flow cytometer, cells might occult one another. This makes the tracking more difficult (especially for automated systems), and it also reduces the observable fluorescence. Finally, adding a secondary observation method increases the robustness of the results, and makes it possible to control for the effect of phototoxicity.

Flow cytometry makes it possible to collect data from a high number of cells, but each data collection must be set up manually, which is really time consuming and not convenient for frequent observations. Moreover, the observation disturbs the system:
even if the observed cells are kept and undamaged by the process, their local arrangement prior the flow cytometry experiment is lost. This is particularly true for adherent cells, which must be trypsinised before being used in a flow cytometry experiment. Thus even if all cells from a sample were to be observed at every collection time, it would not be possible to link them to other cells in terms of lineage. Even though their level of fluorescence could give an idea of the number of cell divisions in their lineage since the time of loading, this number would only be an average and would not be reliable for individual cells.

Previous research in osteosarcoma cells has shown the potential of combining microscopy and flow cytometry. Summers et al. [384] used electron microscopy to gain information on individual nanoparticle clusters. These data demonstrated that the fluorescence was highly variable across cell vesicles, but that the difference of fluorescence across cells was mostly due to different numbers of vesicles per cell. This result could not have been shown from flow cytometry alone, as flow cytometry can only reveal the total amount of fluorescence for the whole cell. These data helped calibrating the dose of nanoparticles required to load a cell for flow cytometry experiments, accounting for the heterogeneity within the cell colony.

Microscopy was also used to study the influence of the state of the cell within its cycle on nanoparticle loading [311]. Since nanoparticle loading is a prerequisite to flow cytometry, these data could not have been gathered without the use of microscopy. On the other hand, flow cytometry allowed the collection of data sets larger by an order of magnitude compared to what could have been gathered from microscopy alone.

In a different cross-platform approach [383], mitotic events were collected through microscopy, revealing information on lineage (generation), the life of the cell (age of the cell and intermitotic time) and the behaviour of the population as a whole (cell growth). Flow cytometry gives more data on the cell cycle, and a Coulter counter confirmed the cell growth (which suggests the same cell behaviour in both experiments). Those data were used together to design a computer model of the behaviour of the cell colony. This highlighted previously unknown synchronisation patterns in the cycle of different cells.

### 2.1.3. Research objectives

In this study, I evaluate the possibilities of the combined use of microscopy (bright field and fluorescence) and flow cytometry for DPP. In particular, I am interested in the properties of their division process, such as how their fluorescence (used as a proxy for their endosomes) is divided at mitosis. I also consider the cell cycle, and the cells' intermitotic time in particular. These data are then used as the basis for computer models that can connect the two collection methods and establish some continuity between the times of flow cytometry data collection. Previous flow cytometry research has shown that although the average fluorescence per cell decreases throughout the experiment, some cells retain a high fluorescence throughout, suggesting that they do not divide. Using microscopy, I find out more about their properties, such as whether there are other differences in behaviour with the other cells, in particular regarding their motility, and how their proportion evolves through time. Impacting the cell cycle of the surrounding cells is one way in which the group might influence the individual.

### 2.2. Methods

### 2.2.1. Cell harvesting and culture

Dental pulp tissue was obtained from the impacted wisdom teeth of patients undergoing dental treatment at the School of Dentistry, Cardiff University, following ethical approval and patient consent. The tissues were minced and treated with $4 \mu \mathrm{~g} / \mathrm{mL}$ of a collagenase/dispase solution. The tissues were left in that solution for 1 hour, at $37^{\circ} \mathrm{C}$. Once the pulps were digested, they were passed through a $70 \mu \mathrm{~m}$ cell strainer, resulting in a single cell suspension [385]. Those cells were then seeded onto fibronectin-coated plates: the differential adhesion of dental pulp progenitor stem cells (DPP) to fibronectin allowed the isolation from the other types of cells contained in the suspension [386]. After 20 minutes, the cells that were still in suspension were removed, leaving only the DPP attached to the fibronectin. The cells were then grown in growth medium (DPPSC medium; alpha-MEM supplemented with $20 \%$ fetal bovine serum) and preserved for further experimental use. Some of the cells were cultured at a density of $4 \times 10^{4}$
cells $/ \mathrm{cm}^{2}$ and incubated under standard culture conditions $\left(5 \% \mathrm{CO}_{2}, 37^{\circ} \mathrm{C}\right.$, humidified HeraCell incubator).

After 24 hours, the cells were loaded with 4nM Qtracker705® (Thermofisher [387]). They are $\mathrm{CdTe} / \mathrm{ZnS}$ quantum dots ( QD ): one dot is constituted of a core of cadmium telluride and a shell of zinc sulphide, which is surrounded by a coating of arginine-rich peptides (Arg)9. This is a peptide comprising of 9 arginine residues. This coating makes it possible for the nanoparticles to penetrate the cell through endocytosis [388]. The cells were incubated at $37^{\circ} \mathrm{C}$ for 1 hour with the QD, as per the manufacturer's instructions [387]. They were then washed twice with fresh warm culture medium, so that no QD remained outside the cells. As DPP are adherent cells, they were trypsinised after 24 hours. Cells were counted using a haemocytometer and suspended in enough media to be at a density of $4 \times 10^{4} / \mathrm{cm}^{2}$. They were then transferred into a 24 well plate (IBIDI - Thistle Scientific), with 1 mL of media per well, to be observed through microscopy. In parallel, a $25 \mathrm{~cm}^{2}$ culture flask was filled with the same solution for those cells to be observed through flow cytometry.

### 2.2.2. Data acquisition

At $t=24 \mathrm{~h}, 48 \mathrm{~h}, 72 \mathrm{~h}$ and 144 h (with $\mathrm{t}=0$ corresponding to the time of seeding), cells were harvested from the flask with trypsin and resuspended in fresh medium. The test tube was placed in a FACSCalibur flow cytometer (Becton Dickinson Inc., Cowley, UK). Using CELLQuest software, the flow cytometer was programmed to observe those cells, with gates set up to only collect particles that emit fluorescence, so as to avoid counting debris. The flow cytometer was programmed to collect 25000 observations, so that at least 20000 cells would be analysed on each observation. The lowest number of observations was 19763 measurement of red fluorescence, which was obtained for the latest acquisition. All three others were higher than 20000 observations.

The cell culture plate was placed in a Cell-IQ imaging platform. Cell-IQ is a fully integrated, automated cell culture and analysis system: environmental conditions (temperature, pH and humidity) can be controlled throughout the observation, and the time of recording and the position of the lens can be pre-set. The platform can thus
automatically record observations without any environmental perturbation, apart from the illumination at the moment of observation. The cells were otherwise kept in complete darkness. Because the position of the light sources and of the microscope can be set precisely throughout the experiment, it is possible to observe different points in the well plate in succession and come back to those exact coordinates at regular intervals. This means that the imaging platform makes it possible to follow cells population in different wells concomitantly, and even in contiguous fields of view in one well.

Preliminary tests showed that the duration of a mitosis in DPP could be as short as 30 minutes. As mitoses are central to the observations I needed to make in this study, the frequency of observation had to be at most that. This frequency proved appropriate to determine cell movement between two consecutive recordings. As DPP are very mobile and can change shape, identifying objects in two images as consecutive instances of the same cell is not a trivial task, especially in dense environments. A frequency of one recording every 30 minutes made it possible to follow the cells in most cases, at least when tracking them manually. This frequency is a compromise between the precision of the tracking and the necessity to minimise the effects of phototoxicity over time.

The observation was done using a $\times 20$ objective lens, and the data were recorded at a scale of $0.145 \mu \mathrm{~m}$ per pixel (all FOV represent an area of $200 \mu \mathrm{~m} \times 150 \mu \mathrm{~m}$ ). Long term tracking is not possible using one field of view (FOV) only: the cell would be likely to leave the FOV before the end of its cycle. The precision of the Cell-IQ imaging platform made it possible to record contiguous FOV, allowing tracking cells for a longer time. All tracking was done in wells for which a square of $3 \times 3$ FOV had been recorded. Moving the observation device and lighting the FOV for a bright field image recording took around 5.3 seconds every time; lighting the FOV at the excitation wavelength of the QD and recording the subsequent fluorescence took about 9.2 seconds. Thus, taking a pair of images of one FOV took 14.5 seconds, and in 30 minutes, it was possible to record 108 different FOV. The arrangement of the FOV can be found in Figure 11. The most important consideration was to minimise the time between any pair of adjacent FOV, so that recording of the same well can be approximated as simultaneous. With this microscope trajectory, no adjacent recordings are separated by more than 73 seconds, which is not enough for any cell to have moved
between the two recordings. It is thus possible to observe the total fluorescence of a cell even when it is located across two fields of view: no part of the cell would be missing or counted twice along the FOV boundary.


Figure 11: Representation of the microscope trajectory in the observation of the well plate (not to scale). Every circle represents a well, and every square represents a FOV; the green arrow represents the path followed by the microscope every 30 minutes. The time required to move from a FOV to the next is negligible compared to the total recording time of a FOV, so regardless of the distance between them, every observation is separated by 14.5 seconds. Hereafter, individual wells will be designated by their row and column (letter and figure respectively).

The recording was carried over a period of 120.5 hours, returning 242 recordings of each FOV, both in bright field and fluorescence microscopy. All of these images were recorded as .tif files. Images in this format can be interpreted by Matlab as matrices, making it possible to carry operations over them to extract relevant features.

### 2.2.3. Data processing

### 2.2.3.1 Noise removal

The images from the bright-field microscopy fields of view could be stitched together by simple juxtaposition: I concatenated the nine matrices representing the 9 FOV of a well, and obtained a large matrix that can be read as an image of all recordings of the well at a given time (as all recordings of a well in one 30-minutes cycle were made within 2 minutes of one another, and contiguous FOV in particular were never recorded more than 75 seconds from one another, all FOV images captured within a 30 -minute cycle can be considered as synchronous. See Figure 11). The resulting picture represents the whole $600 \mu \mathrm{~m} \times 450 \mu \mathrm{~m}$ region of the well that was recorded, with no point within that rectangle missing or being represented twice, as can be seen from the continuity of features across the boundaries of FOV within the resulting stitched image. This process did however leave lines of demarcation in some adjacent views. The reason is that the surface of the background differed depending on the picture, and as a consequence, the white balance was not always done in the same way. The differences are negligible in most pictures, and the result of the juxtaposition was considered good enough for the purpose of the experiment (see Figure 12a).


Figure 12: Result of a simple stitching for the same images, as viewed in bright field microscopy (a) or fluorescence imaging microscopy (b). Stitching marks are barely visible in the bright field view, whereas they are very clear in the fluorescence view, due to the nonhomogeneous lighting

Clear discontinuities appeared in the fluorescence imaging views however. The cells were lit from a single point located above the top left quarter of the image. Hence, the area photographed was not homogeneously lit (see Figure 12b); this is particularly problematic at the places where the images are stitched, as it creates important discontinuities [389].

Such chromatic aberrations (or vignetting), are common in fluorescence microscopy [390], and different methods have been proposed to remove them. The image captured is composed of the biological object, a homogeneous background noise signal that varies over time, and a constant illumination that is coordinate specific [391], [392]; the goal of vignetting correction is thus to isolate the biological object from the raw image by removing the background noise and the illumination. Most methods correct the illumination by estimating the vignetting function, which can be obtained by averaging several images [391], [393] or from captures of an empty field of view [394]-[397]. Other more complex methods are able to estimate the vignetting function without relying on other images, by assumptions on the shape of the function [398]-[400]. Leong et al. [401] suggest a similar method that can be implemented with common imaging software: by convolving the image with a Gaussian function, high frequency information is removed, leaving only the difference in illumination. Then the aberration can be removed by blending the obtained filter with the original image with an image editing software. This method does not require capturing an image of the background.

I first used a similar method to estimate the illumination function through an adaptive thresholding method [402]. This method proved to be insufficient to completely remove the vignetting effect (see Figure 13): I rejected this method.


Figure 13: a: False colour representation of a fluorescent image recording. $\mathbf{b}$ : The same image, after adaptive thresholding (same colour scale); although there is less contrast between the centre of the aberration and the edges, the background is still not uniformly lit.

Since only the endosomes containing QD can be detected by the fluorescence imaging process, and those represent a small minority of the image's pixels, averaging methods proved to be more effective. I removed the temporal variations in background noise by normalising all images, and then averaged across all normalised fluorescence recordings. The resulting image is the shape of the aberration (see Figure 14). This process revealed other systematic imperfections in the pictures, which were removed with the aberration by subtracting the average image from all captures.


Figure 14: The average fluorescence of all the images collected. Most of the fluorescence is due to the spherical aberration. This image also highlights some flaws of the lens that can be detected in all fluorescence pictures (colorbar in normalised fluorescence units).

This process makes the noise uniform. This background noise is much lower than the fluorescence of the quantum dots, but the QD are so highly concentrated that their influence on the total fluorescence of the cell would be negligible compared to that of the background noise. Therefore, if those images were to be used, cell fluorescence would appear to be proportional, not to the amount of quantum dots they contain, but to the size of the cells. It is thus important to remove this noise.

Considering the high contrast between the QD signal and the background noise, a thresholding method appeared to be the best solution. The histogram of fluorescence (Figure 15) shows that the range of the noise only covers low values of fluorescence. In order to determine where to set the threshold between signal and noise, I observed the images with the highest amounts of higher fluorescence pixels, and controlled the amount of fluorescence in areas that did not contain cells. This process showed that these areas only contained pixels below 3 fluorescence units, so all pixels below that threshold were set to 0 .


Figure 15: Distribution of fluorescence across the image. The turquoise line represents the median over all times and FOV, showing that for most recordings, the vast majority of pixels are below 4 fluorescence units. The grey lines represent the minimum and maximum across all recordings, showing that the distribution of fluorescence is consistent between pictures

After the noise removal, $4.95 \%$ of pixels on average contained fluorescence. The results of the noise removal can be seen in Figure 16.


Figure 16: $\mathbf{a}$ : Fluorescence image before any processing; $\mathbf{b}$ : The same image without the noise; c: the corresponding bright field image. The fluorescence matches the location of the cells. In particular, the highest concentrations of QD clusters match the positions of two cells undergoing mitosis.

### 2.2.3.2 Automated tracking

Many different algorithms have been developed for automated cell-tracking [329], [403]. Some algorithms have been optimised for particular cell types [404], others for particular methods of observation [405]. Most methods process the recordings in two steps: cell segmentation (identifying which parts of the video frames are cells) and cell association (linking the different cells temporally). In some cases (in particular, when a fluorescent marker is being used), a simple thresholding algorithm is enough: all pixels that are above a certain value are considered to belong to a cell. If no fluorescent markers are used, this method requires the cells to be well separated from each other; their intensity must be clearly distinct from the background; there must not be too much debris in the FOV; the lighting conditions must be constant and uniform. A more robust method consists in presenting the algorithm with images of cells, so that it searches the FOV for similar objects. This method works well when the cells are consistent in shape and texture [406]. Rather than using directly an image of the cell, it is also possible to have the algorithm learn from features of the images of cells, such as intensity, texture and gradients.

Once all cells have been located in the recording, the cell-tracking algorithm needs to link them temporally. The simplest way to do so is to connect each cell to the cell spatially closest in the previous and next frame. This solution is however not appropriate for large and dense cell populations, or for fast-moving cells. The algorithm is sometimes improved by adding some measure of similarity in terms of intensity, area, perimeter, orientation or estimated position. One specificity of cell-tracking compared to tracking of other subjects is that cells divide, and after division, both daughters will be equally likely to be associated with the mother cell. Some algorithms are able to handle mitosis by using topological methods [407], [408]. Automated tracking of the DPPSC was attempted utilising the Cell-IQ Analyser in-built software [409].

I first tested the software with a population of human osteosarcoma cancer cells (U-2 OS, ATCC HTB-96) [410]. The membrane of those cells tends to be more clearly visible under a bright field microscope compared to the DPP stem cells, and their nuclei tend to stand out more from the rest of the cells. Those characteristics were thought to make them simpler to detect algorithmically. Image samples were extracted from the
recording of U-2 OS cells, and the software attempted to find matches in the different pictures. Different types of errors appeared: the background was detected several times, many cells were detected twice or more and some groups of cells were detected as one cell only (see Figure 17). Nevertheless, this was considered as sufficient for a preliminary tracking, to be corrected manually afterward.


Figure 17: An example of automated detection of U-2 OS cells with Cell-IQ Analyzer that illustrates the different error types. According to Cell-IQ, each red dot represents one cell; yet some cells have several dots or none, and some of the dots are on the background.

I then attempted to use the same software to detect the DPP stem cells. Sample images of cells were gathered and classified depending on their shape: round (mitotic), triangular, long, large. Tracking was considered to be too hard a problem to be solved automatically, as the cells would completely change their shape; so one cell could not simply be labelled as the next iteration of the cell of the same class closest to its position in the previous picture. In particular, toward the end of the recording, the density of cells was such that two cells would often be closer than the distance one cell could cross between two captures, so the cell observed at time $t+1$ closest to the previous position of a given cell $C$ at time $t$ would not necessarily be cell $C$. Finally, this project was particularly interested in mitoses, and a simple classifier would not be able to handle
one cell becoming two. Nevertheless, if the cells were already properly detected, this would tremendously accelerate a manual tracking. Unfortunately, the DPPSC presented too many detection errors for the software to be any help in the tracking (see Figure 18), and the idea of automated detection had to be abandoned for this project.


Figure 18: An attempt at cell detection by Cell-IQ analyser. Every symbol (triangles, and coloured disks) represents the location of a cell of a given type, according to the algorithm. As in the previous example, some cells are detected more than once, and sometimes as different classes. Some cells are not detected at all, whereas the background is. The dark bands on the edges are areas that cannot be tracked. The reason for this is that the algorithm compares thumbnail images from the greater image with the sample cells from the learning base. The dark bands are all the pixels in which thumbnails the size of the largest samples cannot be fitted. Cell-IQ Analyser is able to circumvent this limitation by using the neighbouring fields of view [409], so this limitation is only relevant for the sides with no neighbouring field of view.

### 2.2.3.3 Manual tracking

The tracking was entirely done in Matlab (R2014a, 8.3.0.532). In order to gain information on the cell cycle (and the inter-mitotic time in particular), most of the trackings were started from mitosis. The first step was to collect the coordinates and time of all the mitotic cells. I went through all the fields of view from wells D2 and D4 (in the bright field images. See Figure 11), as nine FOV were recorded for both of them. I collected the coordinates of all cells that appeared to be mitotic, i.e. ball-shaped, which are the cells that appear circular and brighter in bright field microscopy [411]. A cell
could be ball-shaped for a long time before undergoing mitosis. Some reversed to their previous shape, and others remained ball-shaped until the end of the recording or until they left the field of view. There were also some cases of cells apparently undergoing mitosis between two frames, but those instances were rare as the time between frames ( 30 minutes) had been chosen based on the preliminary observations, during which no cell underwent mitosis in less than 30 minutes. However, the high density in the later part of the recording made it possible that some mitoses could have been missed.

Tracking was conducted at the scale of one FOV, rather than relying on the stitched image: accurate tracking requires observing details that do not appear on a stitched image as observed on a computer screen (or such details are displayed at a scale that is too small to be easily detected). Yet, tracking a cell across FOV presents difficulties of its own: it is easier to follow a moving object over a fixed background, so changing FOV would make tracking harder for the human eye [412]. To overcome this difficulty, I created composite FOV based on the 9 captured by the phase contrast microscope (see Figure 19, similar FOV were obtained for the fluorescence microscope).


Figure 19: Examples of composite FOV. In green, two halves of two different FOV are assembled into a composite FOV. In red, a new FOV is made of four quarters of original images. These composite FOV are useful to track cells when they are close to the edge or the corner of one of the original microscopy images.

For every potential mitosis, I observed the cell from before the time it took a mitotic shape. I drew an outline of the shape, and obtained its QD load by adding the fluorescence of all pixels within the outline in the fluorescence view. It is important to collect these data before the cell changes shape, as mitotic cells are thicker, and this means that the endosomes they contain are more likely to hide one another. Because all endosomes in the mother cell are transmitted to either of the daughters [413], the amount of fluorescence coming from a cell should not change throughout the cell cycle [414]. But the problem of occultation means that the fluorescence observed in a cell is not necessarily proportional to the amount of quantum dots it contains, even after removing the noise. For the same reason, the fluorescence of the daughter cells is collected about 5 hours after the mitosis, or whenever the cell is sufficiently isolated from the others that occultation is minimal. Originally, the shape of the cell was also noted in order to analyse the influence of this factor over the speed of the cell and its inter-mitotic time (IMT), but the morphology of the cells were too unstable to exploit this information. Example of a cell before and after mitosis can be seen in Figure 20.


Figure 20: A cell and its daughters before, during and after mitosis (resp. a, b, c for the phase contrast views and $\mathbf{d}, \mathbf{e}, \mathbf{f}$ for the fluorescence microscopy view). At the time of mitosis, the cell occupies a smaller volume, and its QD clusters are more concentrated (note the difference in colour scale). A consequence of this is that QD clusters might occult one another.

The manual tracking was simplified as much as possible by writing a program in Matlab. For every potential mitosis, the program finds the (composite) FOV whose centre the cell is closest to. Every cell in the frame that has already been tracked is marked, so as to avoid confusion during tracking (in fact, some cells that entered the area of recording, rather than being born inside it, were tracked from their entrance point, solely to remove ambiguities regarding other cells). If it is known, the trajectory of the cell over the three previous frames is drawn on the image (see Figure 21). When the user clicks on the new position of the cells, its coordinates are stored and the program moves on to the next frame (unless the user indicates that the cell has been lost or has reached mitosis. It is also possible to change the order of the frames, to rewatch the trajectory or to correct a mistake for instance). The program always displays captures from the same (composite) FOV, so as not to change the background. Thus, if the cell crosses the edge of that FOV, it is possible to interrupt the tracking and restart the program from the same point, but on the FOV that is the most appropriate to follow a cell at these coordinates. When the program is tracking a cell that has already been tracked partially, the last three coordinates are shown, so following the movement is easier.


Figure 21: Example of FOV being tracked. The grey cross is over the current position of the cell being tracked; its previous three positions are shown by its trajectory in red. The cells marked in green have already been tracked and thus cannot be the current position of the cell.

I followed the mother cell until mitosis, and tracked each of the daughters in succession. In every frame, I collected the coordinates of the cell. Whenever it was visible, those coordinates were chosen as the position of the nucleus. Otherwise, I selected the approximate centre of the cell. When the cell divides, the tracking is over, and the cell is considered as "dividing". When I was not able to follow the cell up until its division, it was marked as "lost". This could happen either because it entered a dense area in which the tracking became ambiguous, or because the cell left the area of recording. Finally, whenever I managed to track a cell until the end of the recording without observing it divide, the cell was marked as "non-dividing".

It is of course impossible to say for sure whether a cell is non-dividing, since observing that it did not divide within 120 hours does not completely exclude the possibility that it might have divided later. Previous observations had shown that the IMT of DPP stem cells was around 17 hours, but some cells finally divided more than 50 hours after the previous mitosis. Thus a cell born late in the recording would have been unlikely to divide, not because it was non-dividing, but because it would not have had the time. To avoid this bias, I prioritised the cells born during the first half of the recording.

### 2.3. Results

### 2.3.1. Results of the tracking

321 cells were tracked in total. Of these, 116 were mitotic, 88 were non-dividing, and 117 were lost during the tracking process. More tracking was done by colleagues on those cells, bringing the total of cells for which we know the length of the cell cycle to 415 (see Figure 22).


Figure 22: Distribution of inter-mitotic times

Cell fluorescence was observed for 547 triplets of mother and daughter cells. The QD split ratio could be defined in two ways: either as the ratio of either daughter's fluorescence to that of the mother, or as the ratio of either daughter's fluorescence to the fluorescence of both daughters. In theory, both definitions should be equivalent, since all QD clusters are transmitted from the mother to the daughters, so the fluorescence of the mother is in theory equal to the sum of the daughters' [415]. In practice, there are different problems that cause discrepancies between those values. First, as explained before, the lighting is not homogeneous throughout the FOV. So pixels near the centre of the light could become saturated, and as a consequence, removing the background fluorescence might have removed most of the signal, especially after removing the remaining background noise. Out of 547 mother cells analysed, 30 had no fluorescence detected within the cell, so any fluorescence observed in the daughters would make the QD split ratio infinity, according to the first definition.

Another source of errors for the measurement of QD is occultation: the more QD clusters a cell contains, the more likely it is that endosomes that contain QD will be covered by others that will hide their fluorescence. Thus, occultation is more likely to be
a problem for parent cells than for their daughter, which on average only contain half of their QD clusters. This effect can be observed by measuring the total fluorescence in the field of view (see Figure 23): at the beginning of the experiment, the total observed fluorescence increases, as QD clusters spread out between cells. The subsequent decrease corresponds to cells leaving the field of view and the increase in cell density leading to overlapping cells.


Figure 23: Total fluorescence observed throughout the nine fields of view observed for four wells. The fluorescence has been standardised so all wells have a null temporal average fluorescence. In all wells, the fluorescence increases quickly for the first 17 hours; this is probably due to cell divisions in the first 40 hours post-loading decreasing the average number of QD clusters per cell, thus limiting the risk of occultation. An artefact can be seen for all wells at $\mathrm{t}=113 \mathrm{~h}$, which probably corresponds to perturbation in the room at that time. The black line is a linear fit of the average standardised fluorescence between $\mathrm{t}=17 \mathrm{~h}$ and $\mathrm{t}=113 \mathrm{~h}$; the total fluorescence slightly decreases over time ( 0.014 standardised QD units per hour, mean squared error $=0.007$ ). This decrease is slow enough that QD fluorescence can be considered constant at the scale of a mitosis between those times (the total decrease of fluorescence over that period is of $13.8 \%$ ).

Although the data are noisy, the noise should affect both daughters in the same proportion. Hence, the QD split ratio distribution, as measured from the daughters, should reflect the reality to some extent, provided that outliers are removed from the data sets. I considered as outliers the groups of cells for which either the mother or one
of the daughters does not contain any fluorescence, as well as the daughters that contain far more fluorescence than their mother. Considering the imprecision of the measurements, I consider as outliers the mitoses for which the sum of the daughters' fluorescence appears to be three times the fluorescence of their parent. With the outliers removed, the median fluorescence of the parent cells is slightly below twice the median fluorescence of daughter cells, as expected (see Figure 24a). The distribution of QD split ratios can be found in Figure 24b.


Figure 24: a: Distribution of the fluorescence of the parent cells compared to their daughters. Out of the 547 mitotic events observed, 435 showed fluorescence in all three cells, with the daughters representing less than three times the amount of fluorescence of the parent. The distribution of the parents' fluorescence tends to be twice that of the daughters. Figure $\mathbf{b}$ : Distribution of the split ratios, defined as the ratio of the fluorescence of the most fluorescent daughter by the sum of the fluorescence of both daughters.

### 2.3.2. Results on cell motility

An analysis of the trajectory of the cells was originally planned, but it appeared impossible to gather a non-biased data set. The trajectory of a cell is characterised by its diffusion coefficient $D$ [416], which represents the distance travelled over time from the cell's origin point, and is defined by Equation 2:

$$
\begin{equation*}
D(t)=\frac{(x(t+0.5)-x(t))^{2}+(y(t+0.5)-y(t))^{2}}{0.5 L^{2}} \tag{Equation 2}
\end{equation*}
$$

where $x(t)$ and $y(t)$ are the coordinate of the cell at time $t(t$ being in hours), and $L$ is the scale ratio: $0.14 \mu \mathrm{~m} /$ pixel. Cells were found to move within the picture by $4.4 \mu \mathrm{~m}$ every hour on average, meaning that any cell born in the centre of the field of view constantly moving at that average speed in the same direction would have left the area of recording in 28 hours at most. Hence the non-dividing cells that are still observable after 100 hours would be those with a diffusion coefficient lower than average. And indeed, we observe that non-dividing cells have a lower diffusion coefficient than dividing cells ( $\boldsymbol{p}<0.001$, see Figure 25).

But contrary to my hypothesis, the "lost" population also has a lower diffusion coefficient than the mitotic cells ( $\boldsymbol{p}<0.001$ ). The "lost" population is composed of mitotic and non-mitotic cells that left the field of view before they could be identified as either. Whereas the non-dividing cell population is biased toward low diffusion coefficients, if there is any bias within the "lost" population, it is toward high diffusion coefficients.

Yet, the mitotic population has a significantly higher diffusion coefficient than those relatively fast-moving cells. Since it is not possible to make a comparison between nonbiased samples from this data set, it is not possible to draw definitive conclusions from those data. But they suggest that dividing cells are indeed faster than non-dividing ones, and that it is the dividing cells among the "lost" population that draw the distribution toward higher diffusion coefficients for the "lost" population. This suggests the need for more research, with a larger region of observation, so that even the fastest cells would not have time to leave before they could be definitely labelled as dividing or nondividing. If the difference of speed between mitotic and non-dividing cells were to be
confirmed, this would provide a new way for researchers to predict the behaviour of stem cells.


Figure 25: Distribution of diffusion coefficients of the DPP. The "lost" population contains both mitotic and non-dividing cells in unknown proportions.

### 2.4. Models

### 2.4.1. Microscopy data-driven model

The flow cytometry data explain how QD clusters are distributed between large numbers of cells at different times, and through the microscopy data, it is possible to deduce how a cell's QD clusters are divided between both daughters during mitosis in a small population. If both populations behave in similar ways, it should be possible to combine the data from both experiments to explain how the cell population grows and how its QD clusters spread between the cells.

The general idea of the cell proliferation model is that a population of $N$ cells is initialised with a certain amount of fluorescence (following the distribution obtained from the flow cytometry data). A proportion of them will not divide and the rest will divide after the end of their inter-mitotic times, IMT being drawn from the distribution observed in the microscopy data. When cells divide, their QD clusters are shared between their daughters, following the split ratios observed with the fluorescence
microscope. In this model, fluorescence is proportional to the amount of QD contained within the cell. The new cells appear at the next time point after the division, so as not to be present at the same time as their mother. If the new cells are to divide, they also receive an IMT, drawn from the distribution (Figure 26). The simulation is run for 120 h (with a time step of 0.5 h ), after which the distribution of QD at $24 \mathrm{~h}, 48 \mathrm{~h}$ and 120 h are compared to those observed through flow cytometry.

This comparison is done using Hellinger distance [417], [418]. For two distributions $P$ and $Q$ represented as histograms with $p_{i}$ and $q_{i}$ the numbers of elements of $P$ and $Q$ present in bin $i$, the Hellinger distance $h$ between $P$ and $Q$ for this representation can be computed with :

$$
\begin{equation*}
h(P, Q)=\frac{1}{\sqrt{2}} \times \sqrt{\sum_{i=1}^{k}\left(\sqrt{p_{i}}-\sqrt{q_{i}}\right)^{2}} \tag{Equation 3}
\end{equation*}
$$

$k$ represents the number of bins in both histograms. Both histograms are represented with the same bin values. In the case of the fluorescence distribution, because of the large range of values, it is the logarithms of the fluorescence intensity of the cells that were represented in the histograms (the Hellinger distance was computed based on the linear histograms of the logarithmic data).

In an initial version of the model, IMT values were directly drawn from the values observed. This led to the over-representation of some values. For instance, out of the 415 cells observed, the two longest cell cycles were 58 and 53 hours, each observed once. For some distributions of non-dividing cells, if the model starts with 20000 cells, 250000 cells could go through mitosis by the end of the simulation, 120 h later. As a consequence, by drawing values directly from the distribution, about 600 cells would go through mitosis 53 hours after their creation, as many would divide 58 hours after their creation, but no cell would go through mitosis 55 hours after their creation. This highlights the need to draw the values for the IMT from a continuous distribution fitted to the data, rather than from the data set itself. Because of how skewed the distribution is, and the importance of extreme values in the model (lineages with high IMT will go through few divisions and keep an important part of their initial fluorescence), the fitting model that was chosen is the Generalised Extreme Values distribution (GEV) [419], for which the equation of the probability density function is:

$$
\begin{equation*}
f(x)=\frac{1}{\sigma_{I M T}} \times \exp \left(-\left(1+k \frac{x-\mu_{I M T}}{\sigma_{I M T}}\right)^{-\frac{1}{k}}\right) \times\left(1+k \frac{x-\mu_{I M T}}{\sigma_{I M T}}\right)^{-1-\frac{1}{k}} \tag{Equation 4}
\end{equation*}
$$

The resulting distribution can be seen in Figure 26. Similarly, the split ratio was modelled as the normal distribution $f(x)=\frac{1}{\sqrt{2 \pi \sigma^{2}}} \times \exp \left(-\frac{(x-\mu)^{2}}{2 \sigma^{2}}\right)$, represented in Figure 27.


Figure 26: Probability mass function of the inter-mitotic times (IMT) collected by microscopy (blue bars); the red line represents the type I extreme value distribution associated with it. The parameters are $\mu_{I M T}=17.23$ hours; $\sigma_{I M T}=4.65$ hours and $k=0.087$ (mean square distance with the frequency histogram: $9.01 \times 10^{-5}$ ). The medians of the data set and the distribution are 19 h .


Figure 27: Distribution of the proportion of QD obtained by each daughter in a mitosis. Because the distribution includes both daughters for each mitosis, this distribution is symmetrical. The red curve is the normal distribution that can be fitted to these data. The split ratio is the right half of this distribution; it can be modelled as half of the normal distribution of mean $\mu=0.5$ and standard deviation $\sigma=0.19$ (mean square distance with the frequency histogram: $5.34 \times$ $10^{-5}$ ).

The lowest IMT observed in the microscopy data is 9 h . This means that if cells are initialised at $t=0$, no mitosis will occur for the first few hours. This is not what we observe in the recording, because the cells present at $t=0$ were not all created at that time. For this reason, the cells must be created before the beginning of the simulation. The IMT of the initial cells are drawn from the IMT distribution, and the age of the cells is decided so that they have not reached their predicted time of division before $t=0$. In practice, this is done by defining their creation time $t_{c}$, with $t_{c}$ inferior to the opposite of their IMT (the $N$ cells present at the beginning of the simulation must have been created before $t=0$, and cannot have divided before that time either).

The non-dividing fraction (NDF) is the ratio of cells created that will not divide. This means that a cell created at time $t$ with a potential inter-mitotic time imt has a probability of going through mitosis equal to $1-N D F(t+i m t)$. This function cannot be obtained from direct observation. Among the cells tracked, 116 were mitotic and 88 were classified as non-dividing. But as explained in the paragraph on tracking, most of
the non-dividing cells created early in the recording were likely to have left the field of view by the end of the recording, and were thus not counted: these numbers do not represent the actual NDF. In actuality, the probability of a cell to divide is influenced by the density of cells around it [420].

Because the cells present a rough surface against a smooth background, the area of the well covered by the cells can be considered to be the parts of the images with a high gradient. The progression of the cell coverage can be seen in Figure 28.


Figure 28: Evolution of the cell coverage of 9 the FOV of a well throughout the experiment. The turquoise line represents the median area covered by cells at any time across all FOV. The grey lines represent the minimum and maximum coverage of a FOV at any time.

Because the cell proliferation model does not consider the geometry of population, it is not possible to know the local density around a cell from the simulation, and the covered area does not give enough information to design the model. For this reason, I used the number of cells as a proxy for the cell density in order to characterise the NDF function. I counted the number of cells visible in one well every ten hours; these data
show that the growth of the DPP population can be modelled as a second degree polynomial (see Figure 29):

$$
N(t)=-3.31 \times 10^{-2} \times t^{2}+10.1 \times t+538
$$



Figure 29: Growth rate of the cells observed through microscopy. The cells were counted in the same well every 10 hours. The red stars represent the growth rate from $t=0$, which is the number of cells at any time divided by the number of cells at $t=0$. The turquoise line represents the growth as a quadratic function; the grey area is the $95 \%$ confidence interval. The expansion of the colony slows down through time, and has almost stopped by the end of the experiment.

In order to simplify, let us assume that all divisions are synchronised and all the cells that divide do so at a regular interval $I M T_{m} ; I M T_{m}=19 \mathrm{~h}$ is the median of the IMT distribution (see Figure 26). Under these assumptions, the number of cells can be expressed as $N(i)$, where $i$ is the number of cell cycles since the beginning of the experiment. The number of cells created at cycle $i$ is $N(i)-N(i-1)$. During this cycle, 1

- $N D F(i)$ will divide (with $N D F(i)$ being the fraction of the new cells that will never go through mitosis). From this we deduce Equation 6:

$$
\begin{equation*}
N D F(i)=1-\frac{N(i+1)-N(i)}{2 \times(N(i)-N(i-1))} \tag{Equation 6}
\end{equation*}
$$

Equation 6 is based on unrealistic simplifications, and it is not clear at this point whether this is a realistic model of the actual probability a given cell has to divide. The validity of the hypothesis thus needs to be confirmed by comparing the growth of the cell population in the model with what was observed in the microscopy data. Indeed, although other data and hypotheses are present in the model to control the distribution of the QD clusters, the cell growth only depends on the distribution of IMT (which comes from direct observations of the microscopy recording, and is consistent with previous experiments) and the model for the NDF. Again, as the NDF cannot be directly observed, it is not necessarily possible to determine the actual function, but the NDF has to be consistent with the cell growth that was observed through microscopy.

Replacing in Equation 6 the numbers of cells by those estimated from the microscopy data (see Figure 29), we can then deduce the NDF at those times, and fit it as quadratic function: $N D F(t)=2.67 \times 10^{-5} \times t^{2}-2.07 \times 10^{-3} \times t+0.61$ (see Figure 30).

The fitting is based on values computed between 19 hours and 114 hours, and extrapolation beyond this interval is not reliable. The choice of fitting function is only meant to represent the NDF for the duration of the experiment, and should not be extrapolated beyond 120 hours. In particular, this quadratic function would become higher than 1 at 166 hours, which would not make sense considering that, being the fraction of cells that do not divide, it is bounded by 0 and 1 by definition.

A linear function would also return a good fit (as can be seen from the low leading coefficient compared to the others: this definition of the NDF function is close to a linear function at this scale), and a quadratic function could be considered as overfitting. But since it is not known whether the cells grew in the same manner in both experiments and that the data are compatible, it is possible that the model might require some tweaking, in particular for the parameters of the NDF function. The choice of a quadratic function leaves more degrees of freedom in case such modification is
required, yet does not make the exploration space too large, as it would be if the NDF were to be modelled as a more complex function (such as a polynomial of higher degree).


Figure 30: Fraction of the cells that do not go through mitosis. The red stars represent the NDF as deduced from Equation 6, assuming the cell population grew at the speed described by Equation 5 (see Figure 29) and that the mitotic cells divided every 19 hours. The turquoise line is the NDF quadratic function fitted to these data points. The grey area represents the $95 \%$ confidence interval associated with this fitting.

Using these distributions of the IMT, the split ratios and the NDF, we can run the microscopy data driven model (MDDM). A population of $N=20000$ cells is initialised with their potential IMT drawn from the IMT GEV distribution (Figure 26) and a corresponding creation time. Cells are initialised with fluorescence values drawn from the data collected from the flow cytometry experiment at 24 h post QD labelling. Once a time equal to the potential IMT has passed since the creation of the cell, a random number is drawn between 0 and 1 . If this number is higher than the NDF at the time the cell is supposed to divide, then no division occurs for this cell. It is considered as non-
dividing and remains until the end of the simulation without dividing. Otherwise, it divides and its QD clusters are shared between its daughters along a split ratio drawn from the normal split ratio distribution (Figure 27). Both new cells receive a potential IMT, representing the time at which it will be determined whether the cell goes through mitosis. The mother is considered to have existed from its creation to its mitosis. Cells whose time of potential mitosis is later than $\mathrm{t}=120 \mathrm{~h}$ are considered non-dividing. The simulation ends when all remaining cells are labelled as non-dividing. The algorithm is described in Figure 31.


Figure 31: Flowchart describing the cell proliferation model. The text in bold indicates there is randomness at that step. There are five random processes:

- The drawing of $\operatorname{IMT}$ (cell) from the IMT distribution (see Figure 26). This number is the time after which a cell will divide, if it is chosen to divide;
- The selection of a random number rand to decide whether a cell will divide;
- The selection of a creation time $T_{c}$ for the initial cells;
- The drawing of $Q D($ cell) from the distribution of fluorescence observed at 24 hours in the flow cytometry experiment;
- The drawing of $r$ from the QD split ratio distribution (see Figure 27)

The distribution of fluorescence at times $\mathrm{t}=0 \mathrm{~h}, \mathrm{t}=24 \mathrm{~h}, \mathrm{t}=72 \mathrm{~h}$ and $\mathrm{t}=120 \mathrm{~h}$ are then compared with the data collected from the flow cytometry experiments at those times. This comparison is made by computing the Hellinger distance between 15000 cells samples for each distribution (the flow cytometry recordings were set for 25000 cells each time, but the actual number of events recorded ranged between 19763 and 21729. 15000 cells drawn at random are a representative sample for every data set). On average, two samples of 15000 cells drawn from a same distribution of 20000 cells were found to have a Hellinger distance of $4.75 \%$ with each other on average (std=0.003); I consider that a distance inferior to $10 \%$ means that the model is accurate.

The lowest measurement of fluorescence that can be provided through flow cytometry is of 1 fluorescence unit: cells that do not contain enough QD clusters to reach this threshold are rejected from the data set. Similarly, the cells in the simulation that emit less than 1 unit of fluorescence are withdrawn from the simulation. The resulting distribution of fluorescence per cell at those times can be seen in Figure 32b.

The cell growth in the MDDM matches the growth observed through microscopy (see Figure 32c). As explained previously, this does not necessarily mean that the nondividing fraction evolves according to the quadratic function presented in Figure 30, but it confirms that the simplifying hypotheses from which this function was deduced are acceptable: the growth model is consistent with the microscopy data.

The QD distributions are however very different from the data collected in the flow cytometry experiment (Figure 32a). This suggests that cells behave differently depending on whether they are observed through bright field and fluorescence microscopy or not. To understand this difference quantitatively, we need to find a model of cell behaviour that matches the flow cytometry data.


Figure 32: Distribution of fluorescence in the flow cytometry experiment (a) and in the MDDM (b). Only the initial distributions match (Hellinger distance: 5.1\%). The cells tend to keep more QD clusters in the model than they do in the experiment, which suggests that the cells that were observed through flow cytometry divided more often. c: The growth of cells in the microscopy experiment (stars) and in 30 iterations of the model (in grey; the turquoise line represents the median of those simulations): the cell growth in the model does match that of the cells from the microscopy experiment, indicating that the cell proliferation model is a good description of the growth of the cells observed through microscopy.

In order to fit the MDDM to the flow cytometry data, I implemented a differential evolution algorithm (DEA) [421]. Differential evolution is a method that finds an optimal solution to a problem by iteratively improving a group of candidate solutions, which are called agents in differential evolution literature (this optimal solution might just be a local optimum though).

A number $N_{p}$ of agents (possible solutions) was created. Each agent was described by four parameters, which were the dimensions of the search space: the three coefficients $p_{1}, p_{2}$ and $p_{3}$ of the polynomial that described the $\operatorname{NDF}\left(N D F(t)=p_{1} \times t^{2}+p_{2} \times t+\right.$ $p_{3}, p_{1}>0$ ), and the standard deviation $\sigma$ of the normal distribution that described the split ratio (since of QD clusters go to one daughter or the other, the average of the distribution is necessarily $\mu=0.5$ ). Following the recommendations of Storn [422], the number of agents was chosen to be five times the dimension of the problem: $N_{p}=20$.

The three parameters that described the NDF were all drawn from uniform distributions. Since the NDF represents the probability of a cell not to divide, the search was focused to quadratic functions bounded by 0 and 1 : each parameter was sampled from an interval such that there were possible values for which the quadratic defined by the polynomial $\operatorname{NDF}(t)$ would remain between 0 and 1 . For instance, if $p_{3}>1, N D F(0)>$ 1 regardless of the values of $p_{1}$ and $p_{2}$, so no quadratic function with $p_{3}>1$ would remain between 0 and 1 over the interval [ $0 \mathrm{~h}, 120 \mathrm{~h}$ ]. For similar reasons, by limiting the possible values of $p_{1}$ between 0 and $3 \times 10^{-4}$, $p_{2}$ between $-4 \times 10^{-2}$ and $4 \times$ $10^{-2}$, and $p_{3}$ between 0 and 1 , I ensured that all possible polynomials would be included in the search space. The initial values for parameters $p_{1}, p_{2}, p_{3}$ of the $N_{p}$ agents were chosen so that the NDF would remain between 0 and 1 .

The fourth parameter was the exponential of a uniform distribution between -5 and 0 . I considered these numbers to provide an accurate description of the search space, as below $\sigma=e^{-5}$, the normal distribution of split ratios could be considered as analogous to a constant split ratio of 0.5 , and above $\sigma=1$, the distribution of split ratios could be considered as uniform.

The MDDM was then run for all $N_{p}$ agents. The maximum of the Hellinger distances between the four flow cytometry data sets $(t=0 \mathrm{~h}, t=24 \mathrm{~h}, t=48 \mathrm{~h}$ and $t=120 \mathrm{~h})$ and
the cell distribution obtained by the model for those times would be considered as the cost function $h$, which defines the quality of the model.

The idea of a DEA is that for each agent (candidate solution) $x$, three different agents $a$, $b$ and $c$ are randomly chosen from the $N_{p}$ candidate solutions. A new agent $y$ is created based on these four agents by changing the value of some of the parameters $x_{i}$ of $x$ to:

$$
\begin{equation*}
y_{i}=a_{i}+F \times\left(b_{i}-c_{i}\right) \tag{Equation 7}
\end{equation*}
$$

with $a_{i}, b_{i}$ and $c_{i}$ being the $i^{\text {th }}$ parameters of agents $a, b$ and $c$ respectively (in the MDDM, this can correspond to the computation of a new value of $\sigma$ based on the value associated with those three candidate solutions for example), and $F$ being a constant parameter of the DEA which controls the amplification of the differential variation [421]. To choose the parameters that should be changed, a uniformly distributed random number $\rho_{i}$ is drawn between 0 and 1 for each dimension. Furthermore, the number $\rho_{i}$ for one dimension chosen at random is set to 0 , to make sure that at least one parameter gets changed. For all $\rho_{i}$ below a constant number $C R$, the corresponding dimension of $x$ is changed according to Equation 7. If the NDF function associated with $y$ is outside of the [0,1] range, different agents $a, b$ and $c$ are chosen among the $N_{p}$ candidate solutions until the NDF remains within that range.

The MDDM is then run for this new candidate solution $y$. If the cost function associated to agent $y$ is lower than for agent $x$, agent $y$ replaces agent $x$ in the population of $N_{p}$ candidate solutions. This process gets repeated until the first three parameters describe a NDF function that remains between 0 and 1 for the duration of the simulation. Following the guidelines of Price et al. [423], I chose $F=0.9$ and $C R=0.9$.

The process described in the previous paragraphs is executed until one of the end criteria is reached [424]:

- Exhaustion-based criterion: the cell model is run 2000 times for each agent;
- Reference criterion: the cost function for all agents gets below 0.1;
- Distribution-based criterion: the agents become so close to each other than no progress can be expected. To determine whether it is the case, the standard deviation of the
parameters for all agents is divided by the size of the search space. Whenever the average of those numbers gets below 0.1 , the program stops.

This algorithm was run 5 times to determine whether it would always converge on the same model, each time returning 20 agents: 20 models that are able to return distributions of fluorescence that closely match those observed experimentally through flow cytometry. The cost function for those 100 models ranged from $9.46 \%$ to $9.95 \%$. The results can be seen in Figure 33.


Figure 33: Final distributions of each of the four parameters, for each of the five iterations. For each graph, the y axis represents the entire search space for the parameter, so a narrow box plot means that the DEA algorithm converged for the parameter, whereas a long box plot means that a large range of values can lead to a good fitness. $\sigma$ converged on the same narrow sets of values in all iterations $(\mathrm{std}=0.009)$.

Parameter $\sigma$ shows a clear convergence around 0.184 , very close to the distribution observed in the microscopy experiment, for which the standard deviation of the normal distribution of split ratios was 0.19 (see Figure 27). However, the NDF parameters do not show such a clear convergence. In particular, the largest difference between two models of the NDF obtained by the DEA is 0.28 , obtained toward the end of the simulation (see Figure 34). This is explained by the fact that the cells that are created toward the end of the simulation are unlikely to have time to reach mitosis, so their status as dividing or non-dividing cells is irrelevant to the distribution of QD, and as a consequence, to the fitness of the model.


Figure 34: Models of the NDF that best match the different distributions of fluorescence from the flow cytometry experiment. The MDDM returned 100 agents that match the flow data well, each of these solutions having an associated quadratic NDF function; all of these functions are within the grey area. The turquoise line represents the median of those models at all time points. The blue line represents the NDF model deduced directly from the microscopy observations. The models that match the flow cytometry data the best have far more dividing cells than the model directly deduced from the microscopy data.

In the DEA, agents were updated only if their cost function was lower than the cost function obtained for the previous agent. All iterations of the DEA ended when the reference criterion was reached, meaning all 100 agents achieved a cost function below 0.1. This does not mean that those agents would always return a cost function this low, as the cell proliferation model includes a part of randomness. In order to control the real performance of these agents, I ran the cell growth model 30 times for each. The cost functions were extremely consistent for each agents and across agents, with an average cost function $h=13.3 \%$, and a standard deviation between all those cost functions of 0.002 . All these agents can thus be considered as equivalent. The fluorescence distribution obtained for one of them across all 30 iterations can be found in Figure 35.


Figure 35: Evolution of the distribution of fluorescence, as observed experimentally through flow cytometry, and according to the MDDM when run 30 times with one of the agents selected by the DEA. In all graphs, the grey area represents the QD distributions obtained over all 30 iterations of the model for the agent, and the turquoise line is the median of these distributions. a compares the distributions of fluorescence at $t=24$ hours, $\mathbf{b}$ compares the distributions at $\mathrm{t}=48$ hours and $\mathbf{c}$ compares them at $\mathrm{t}=120$ hours. The distribution at $\mathrm{t}=0$ is included in all graphs for comparison; $\mathrm{t}=0$ is the time of the first observation of the cells through flow cytometry, 24 h after loading the QD (the data from the model for this time are not included, as they are directly drawn from the same data set). All runs were initialised with the distribution of fluorescence observed at $\mathrm{t}=0$ and used the same NDF function and split ratio distribution.

### 2.4.2. Flow cytometry data-driven model

The shape of the NDF function in the MDDM was derived from the growth curve for the cell population, which was obtained by counting the cells visible in the microscopy experiment. The cell growth can also be deduced from the flow cytometry data. Because the cells collected from the flask to be observed through flow cytometry are negligible in numbers compared to the entire flask population, the total amount $Q D_{0}$ of QD clusters in the flask can be considered as constant throughout the experiment. Thus, the number of cells in the flask at time $t$ is $N(t)=\frac{Q D_{0}}{c(t)}$, with $N(t)$ being the number of cells in the flask at time $t$ and $c(t)$ the average fluorescence per cell. Since the cells observed in the flow cytometer are a random sample from the flask, the average amount of fluorescence per cell is the same in this sample and in the flask, which gives us Equation 8:

$$
N(t)=Q D_{0} \times \frac{N_{\text {flow }}(t)}{\sum_{i}^{N_{\text {flow }}(t)} Q D_{i}^{\text {flow }}(t)}
$$

In this equation, $N_{\text {flow }}(t)$ is the number of cells observed in the flow cytometer at time $t$, and $Q D_{i}^{\text {flow }}(t)$ is the amount of fluorescence measured in cell $i$ of the population observed in the flow cytometer. Equation 8 gives us the growth curve for the cells in the flask, which can be modelled by the following polynomial of degree 2: $g(t)=8.26 \times$ $10^{-4} \times t^{2}+1.13 \times 10^{-2} \times t+1.1$ (see Figure $36 ; g(t)=1$ means as many cells are present as at the beginning of the experiment).

In contrast with the MDDM, we do not have enough data here to reliably model the system as a quadratic: it is important to insist that the times at which the flow cytometry data were collected were not chosen for this purpose. It is thus not clear whether a quadratic function is the best way to describe the cell growth. With so little data, it is essential to keep models as simple as possible in order to avoid overfitting. Here, we have three data points close together and almost aligned, and a fourth one far from the others which does not match this pattern of linear growth: any function that can be fitted through these points will be heavily influenced by noise in this last measurement. Because the error would be too high with a linear model, a quadratic model is the least complex model that matches the data, but that does not mean it is correct. It is however
the best that can be obtained from so little data, and thus the one that will be used from here on, despite the high level of uncertainty. It will be important to confirm results with more experimental data.


Figure 36: Growth rate $g(t)$ of the cells in the flask. The red stars represent the growth from $t=0$ h, as estimated from the flow cytometry readings, using Equation 8. The turquoise line represents this growth as a quadratic function.

Using these values of growth in Equation 6 gives us the NDF for every cell cycle, from which we deduce the following continuous model of the NDF based on those assumptions:

$$
N D F(t)=\frac{p_{1} \times t+p_{2}}{t+q}
$$

with $p_{1}=0.51, p_{2}=-9.7$ and $q=2.0$ (see Figure 37 ). Running the cell growth model with this NDF function returns on average a cost $h=11.3 \%\left(s t d=2.7 \times 10^{-3}\right)$, which indicates that the model is consistent.


Figure 37: Model of the fraction of cells that do not go through mitosis (turquoise line) with associated $95 \%$ confidence interval (in grey). The red stars represent the NDF as deduced from Equation 6, assuming the cell population grew at the speed described in Figure 36 and the mitotic cells divided every 19 hours. Under these hypotheses, the NDF appears to follow the rational function described in Equation 9.

The MDDM assumed that the NDF was a quadratic function of time, so the DEA would only search for the quadratic function that would best match the flow cytometry data. This new NDF however is a rational function: it is thus necessary to run the DEA with
this model of the NDF (henceforth flow-cytometry-data driven model, or FCDDM). The DEA was thus run as described for the MDDM, with 20 agents.

Using the default values of $p_{1}, p_{2}$ and $q$, the resulting function is negative for the first 19 hours (see Figure 37), which is not logically possible. The choice of a rational function as a model for the NDF was guided by the shape of the data and the necessity to keep the model as simple as possible as a consequence of the low number of measurement (since a more complex model would be overfitting). But as with the MDDM, in which the choice of a quadratic function for the NDF function could not logically be extrapolated beyond a certain time (as such functions diverge toward infinity), it is clear that the model would not be suitable for times at which the NDF would be below 0 or above 1 . But since the parameters deduced from the observations themselves go into the forbidden area, the search space cannot be restricted to values that would maintain the NDF between 0 and 1 as I did in the MDDM. Thus, the agents were initialised between 0 and twice the default value of the parameter, but were allowed to go beyond that range. NDF values below 0 were set to 0 , and those above 1 were set to 1 .

Initial runs of the FCDDM showed a quick convergence for the standard deviation of the split ratio, towards the same range of values as the previous model. For this reason, the split ratio distribution was fixed to the distribution observed in the microscopy data, with $\sigma=0.19$. The DEA was run five times with agents defined by the three parameters of the rational NDF function; each time, the only criterion of convergence was the reference criterion, which states that all agents should reach a cost score below 0.1. The final distributions of agents can be found in Figure 38. Examples of the evolution of the non-dividing fraction over time, according to the DEA, can be found in Figure 39.


Figure 38: Final distributions of each of the three parameters of the FCDDM, for each of the five iterations. Although each agent fits the flow cytometry data well ( $h<0.1$ ), the agents show large variations.


Figure 39: Three examples of NDF model obtained from the FCDDM, and the model directly obtained from the flow cytometry data. The values of parameters $p_{1}, p_{2}$ and $q$ for agents 1,2 and 3 are respectively $\{19.3,-1734,-4.43\},\{1.13,-59.1,-3.37\}$ and $\{2.52,-188,-3.77\}$. Although the NDF is defined as a rational function, because it is bounded by 0 and 1 , the shape of the function could be better understood as a sigmoid, or even a step function for agent 1 .

For any negative value of $q_{1}$, (which is the case for all the models obtained from the DEA), as well as for values of $p_{1}$ and $p_{2}$ for which $p_{2}<-p_{1}$, the quadratic function that describes the NDF is undefined at the beginning of the simulation, and negative for a few hours later. Over this period, the NDF can be considered as a constant equal to 0 . Similarly, some values of the parameters lead to the rational function getting higher than 1. As a consequence, the shape of the NDF should be considered as a sigmoid rather than a rational function.

Despite the heterogeneity of the agents found with the FCDDM, the actual performance of these agents (the average cost function across 30 iterations of the growth model using these agents' parameters) were consistent for each agent and across agents, with an average cost function $h=11.6 \%$, and a standard deviation between all those cost functions of 0.001 . Therefore, as with the MDDM, all these agents can be considered as equivalent regarding the distributions of fluorescence they return. The distributions obtained for one of these agents across all 30 iterations can be found in Figure 40.


Figure 40: Evolution of the distribution of fluorescence, as observed experimentally through flow cytometry, and according to the FCDDM when run 30 times with one of the agents selected by the DEA. In all graphs, the grey area represents the QD distributions obtained over all 30 iterations of the model for the agent, and the turquoise line is the median of these distributions. a compares the distributions of fluorescence at $t=24$ hours, $\mathbf{b}$ compares the distributions at $t=48$ hours and $\mathbf{c}$ compares them at $\mathrm{t}=120$ hours. The distribution at $\mathrm{t}=0$ is included in all graphs for comparison. All runs were initialised with the distribution of fluorescence observed at $\mathrm{t}=0$ and used the same NDF function and split ratio distribution.

### 2.4.3. Third model

Although the FCDDM was searching for representations of the NDF that were rational functions, the parameters described by some of the agents obtained through the DEA describe rational functions that cross the $[0,1]$ interval for a very short amount of time. Since the NDF must be entirely within the [0, 1] interval, the NDF was set to 0 for times at which the rational function is negative, and to 1 for times at which the rational function is above 1 . As a consequence, functions that start below 0 and end above 1 after a very short transition could be better approximated as step functions rather than rational functions.

This was notably the case for the agents obtained from the FCDDM after more than 100 generations. Running the DEA for 10 hours without any other end criterion confirmed this tendency: the agents achieved a better fitness, with costs function ranging between $8.81 \%$ and $9.07 \%$ (because of the randomness of the cell proliferation model, this does not necessarily mean that they are better representations of the experimental cell growth). For all of these agents the NDF consisted of two phases. In the first phase, all cells would divide; this phase would last for the first 90 to 97 hours. In the second phase, no cells would divide. This phase would last until the end of the simulation. The transition between both phases lasted less than an hour for all agents. This result suggests a third model for the NDF, as a shifted unit step function.

This model depends on one parameter only, the time of the transition. Therefore, instead of using the DEA, I ran the model 30 times for all possible transition times. The results can be seen in Figure 41. For all stepping times between 86 hours and 96 hours, the cost functions returned are similar to those obtained with the MDDM and the FCDDM, with $h=11.6 \%$, std $=7.3 \times 10^{-4}$. The fluorescence distributions that result from such NDF functions across 30 iterations can be found in Figure 42.


Figure 41: Cost function obtained when the NDF is defined as a step function. The abscissa represents the time of stepping. The grey area represents the cost function for all iterations using a given time of stepping, the turquoise line is the median of these values. When the stepping time is between 86 hours and 96 hours a step function model returns a cost function similar to those of the agents obtained using the DEA for the MDDM and the FCDDM.


Figure 42: Evolution of the distribution of fluorescence, as observed experimentally through flow cytometry, and according to the step model when run 30 times with one of the agents selected by the DEA. In all graphs, the grey area represents the QD distributions obtained over all 30 iterations of the model for the agent, and the turquoise line is the median of these distributions. a compares the distributions of fluorescence at $\mathrm{t}=24$ hours, $\mathbf{b}$ compares the distributions at $\mathrm{t}=48$ hours and $\mathbf{c}$ compares them at $\mathrm{t}=120$ hours. The distribution at $\mathrm{t}=0$ is included in all graphs for comparison. All runs were initialised with the distribution of fluorescence observed at $\mathrm{t}=0$ and used the same NDF function and split ratio distribution.

### 2.5. Discussion

In this study, I observed a population of dental pulp progenitors (DPP) which had been seeded with fluorescent nanoparticles, using three different techniques: bright field microscopy, fluorescence microscopy and flow cytometry. The cells observed through both imaging techniques returned a data set of 54 fields of view (FOV) over 120.5 hours, of which two sets of $3 \times 3$ FOV were used to track the cells. I used those images to extract the trajectories of 321 cells and obtained the intermitotic time (IMT) of 415 . In addition, I observed 435 mitotic events and determined how the QD intensity of the parent cell was redistributed to its daughter cells. I also gathered data about colony growth; I deduced the number of cells present in each well and the surface coverage by these cells. I used these data to implement three in-silico models of DPP growth, intrinsically embedding the redistribution of QD signal via mitosis as a key element.

The microscopy observations and the different models highlighted a proportion of cells that do not divide throughout the observation. As most of the cells were tracked from an early time in the experiment, those cells can indeed be characterised as "non-dividing", since if a mitosis were to occur after the end of the experiment, the IMT for that cycle would be much higher than those observed throughout the experiment. It is thus unlikely that those cells would have divided, had the experiment run longer.

Among the cells that did divide, the IMT could be described as a Generalised Extreme Value model, with location, scale and shape parameters respectively $\mu=17.23$ hours, $\sigma$ $=4.65$ hours and $k=0.087$. This long tailed distribution confirms the observations of Gronthos et al. [425] who described the clonal expansion of DPP using Poisson distribution statistics.

A consequence of the non-dividing cells was that even considering groups of $3 \times 3$ FOV was not enough to follow a cell throughout its life: the longer a cell lives, the more likely it is to leave the observable part of the well. By definition, non-dividing cells have a longer life than dividing ones, which could introduce a bias in the speed of the cells that were tracked. For this reason, the results of the observations of motility were inconclusive. Still, comparison between dividing and those cells for which it was not known whether they would divide or not, suggested that on average dividing cells move faster than the non-dividing cells.

I observed that in a mitosis, the fluorescence of the mother cell is not divided equally between the daughters. The data collected from the fluorescence microscopy showed a clear normal distribution, whose standard deviation was confirmed by all computer models. Although quantum dots (QD) are not distributed uniformly across all available endosomes, due to random uptake during the labelling phase [94], differences in cell fluorescence is better explained by the number of endosomes per cell than by difference in the number of QD per cluster. This means that this normal distribution of the fluorescence division ratios represents the way QD-labelled endosomes are redistributed from the parent to the daughter cells.

Endosome distribution has been found to be asymmetric in different types of cells. Bergeland et al. [413] observed a "stochastic, yet ordered" division process, and Dunster et al. [426] described a random but biased dilution of nanoparticles carrier endosomes. Although there seems to be a consensus on the asymmetric nature of endosomal division during mitosis, different models have been used to describe this asymmetry. For instance, Errington et al. [311] relied on a constant split ratio of $0.74 / 0.26$ for their model. Other models use binomial distributions, considering the attribution of one endosome to one daughter or the other as a stochastic process with a binary outcome [94]. The normal distribution presented here can be seen as a continuous view of that discrete stochastic process. However, where Summers et al.'s model relies on a mean probability of 0.52-0.72 [94], the model presented in this chapter indicates that the mean of the distribution can be 0.50 , as long as a non-dividing fraction accounts for cells maintaining a large amount of nanoparticles throughout the experiment.

Asymmetric division processes have been observed for the distribution of different organelles, such as centrosomes [427], lysosomes and Golgi apparatus [426]. These processes play a role in the differentiation of stem cells [243], [244], [428], and the different ways of dividing organelles could be strategies for optimising the survival of the cell's lineage [243], [429]. Asymmetric division could also be the cause of tumour in stem cells [430].

An important result of this study is that the cell growth observed through microscopy could not account for the results observed through flow cytometry. Assuming the
growth of the colony could be described by similar equations, the parameters of the equations describing the microscopy data and the flow cytometry data would need to be very different. The cells observed through microscopy divided much more slowly than those observed through flow cytometry.

The most probable cause for this discrepancy is that the light from the microscope (whether the bright field or the fluorescent light) is damaging to the cells and thus modifies their behaviour. In a previous cross-platform experiment involving U2-OS cells, Khan et al. [383] observed similar behaviours in both platforms. In that study, microscopy observations were conducted twice as frequently as here, but only relied on the bright field channel, which is less toxic for the cells than the fluorescent channel. In the study presented in this chapter, the fluorescence microscopy provided valuable information regarding the endosome division process. Since I have shown that this information could be obtained through flow cytometry alone, it might prove useful for further research to observe the cells in bright field microscopy only.

Because of the difference in growth between the microscopy cells and the flow cytometry cells, no unified model could account for the results observed with both platforms. The variables obtained from the microscopy data had to be greatly modified through the use of an evolutionary algorithm in order to match the distribution of fluorescence obtained by flow cytometry. This evolved model however no longer matched the cell growth observed through microscopy. Another model was designed based solely on the data obtained by flow cytometry. The models differed in the way the non-dividing fraction of cells (NDF) evolved over time, (i) as a quadratic function or (ii) as a rational function. In both cases, different values of the parameters describing the NDF had the same power to explain the distribution of the QD fluorescence per cell that was observed through flow cytometry. In the case of the MDDM, the differences in parameters between the successful variants of the models were not meaningful, as they impacted the NDF curve for times at which the cells were unlikely to divide anyway. The FCDDM however showed a wide variation among the successful models.

A third, simpler model was implemented as a variant of the FCDDM, in which the NDF was modelled as a step function, all cells dividing before the step and no cells dividing
afterwards. Here, I found an optimum position of the step, which resided in the interval between 86 and 96 hours in order to explain the flow cytometry data.

Because all three models have the same explanatory power, more data would be required to understand better the way the NDF evolves over time, in particular during the 86-96 hours interval. All three models do show however that a fraction of the cells does not divide, and that this fraction increases over time.

Of course, cells from this colony have been grown for much longer than 96 hours, and there is nothing special about this number. Time should instead be understood as a proxy for the cell density, which is an important factor in the growth of stem cells [420], [431], [432]. In conclusion, those results show that the cell density impacts the behaviour of the cells, or in other words, that individual cell behaviour is impacted by the structure of the group.

### 2.6. Conclusion

Cells are by nature a collective system, as most cells are able to give birth to a colony. This description is more relevant for stem cells, which can differentiate into different cell types, all interacting according to their own rules. In this chapter, I tried to understand how one type of cells, dental pulp progenitor stem cells (DPP), behave and interact with each other. I obtained recordings of cell colonies, and attempted to track them automatically. This process highlighted the current limits of automated tracking, as the irregular shapes of the cells make them difficult to distinguish from: one another, groups of them, the background and debris.

Moreover, once cells have been detected, a higher sampling frequency would be required in order to connect images of one cell through time. This would damage the cells, and is another constraint to the choice of the sampling frequency, besides the capacity of the recording device and the duration of behaviours of interest. In order to obtain relevant results, it is essential that the observation disturbs the system a little as possible. In this case, the sampling frequency was determined by the length of mitotic events: in order to observe most mitoses, the sampling frequency could not have been lower than 2 images per hour. But this frequency is already too high, as the level of
phototoxicity already had an impact on cell behaviour. It is important to note that this problem is mostly due to fluorescence imaging: other similar experiments have shown that with bright field microscopy only, a higher sampling frequency could have been chosen without damaging the cells [383]. The ability to adjust the sampling frequency is another advantage of cross-platform recordings.

All the observations were done in-vitro. In the case of the flow cytometry, the cells were observed outside of their environment, even isolated from the other cells at the moment of observation at the measurement point. It is interesting to note that even in these conditions, relevant information can be deduced about the group behaviour: the impact of the group can be seen even when the cell is observed in isolation. The microscopy observations were conducted on microlayers, but it would also be possible to observe the progression of cells within tissues [385], or even in-vivo [313]. Those techniques would shed new knowledge on how DPP behave in their natural environments on a macroscale, but in-vitro imaging was required to understand the behaviour of the cell on the level of the individual.

Tracking and counting was done manually, and thus there was no source of error to correct. Fluorescence observations on the other hand introduce different sources of error. One is autofluorescence; it can be dealt with by observing the cells at a different wavelength from their emission waveband. Another source of error was the background fluorescence. Although I attempted to remove it in a way that would maximise the signal to noise ratio, the definition of the threshold does have an impact on the results. For this reason, the fluorescence split ratio was initially treated as an indicator of how endosomes were divided between daughters at mitosis. However, the simulated models confirmed those observations very precisely, thus validating the measurement despite the noise.

The cross-platform approach made it possible to collect data on the local and global scale. Local information included intermitotic time, cell moving speed, and the fluorescence of a cell. Global information was the number of cells and the distribution of QD fluorescence per cell obtained by flow cytometry. In the model, the only explicit interaction is the mitosis: when considering the population of all the cells present at any point in the experiment, a mitosis can be seen as the interaction between the mother cell
and its daughters, and the split ratio is a property of that interaction. These data were the basis upon which different computer models were designed. As I mentioned previously, the fact that those different models converge regarding the split ratio, and in particular converge on the same values as the direct observations, is a strong indicator of the validity of that aspect of the model. The three models also confirmed the existence of a non-dividing fraction (NDF) of cells, which increases over time. All three models have the same explanatory power regarding the flow cytometry observations, even though all three models of the NDF differ. This makes it impossible to determine how the NDF evolves using this data set alone. However, they provide clues as to which specific times and aspects of the population should be more closely observed in future experiments. It is only through a continuous exchange between the computer models and the biological observations that our models of biological behaviour can approach the behaviours themselves.

# CHAPTER 3 - IMPACT OF OCEAN ACIDIFICATION ON <br> THE SHOALING BEHAVIOUR OF EUROPEAN SEA BASS 

### 3.1. Introduction

### 3.1.1. A changing ocean

### 3.1.1.1 The carbon cycle

Carbon (C) accounts for $0.045 \%$ of the chemical composition of the Earth [433], and $0.05 \%$ of its mass, or $3.2 \times 10^{21} \mathrm{~kg}$ [434]. Most of it is trapped deep within the Earth's mantle, and only $7 \times 10^{19} \mathrm{~kg}$ can be found in the upper mantle [435]. In the Earth's crust, carbon is mostly found in sedimentary rock, in organic ( $1.2-1.56 \times 10^{19} \mathrm{~kg}$ [436], [437]) and most importantly carbonate form ( $5.4-6.5 \times 10^{19} \mathrm{~kg}$ [436], [438]). The rest of crustal carbon is stored as basaltic, granitic and metamorphic rocks ( $1.33 \times$ $\left.10^{19} \mathrm{~kg}\right)$ and as fossil fuels ( $4 \times 10^{15} \mathrm{~kg}$ [439]). $3.9 \times 10^{16} \mathrm{~kg} \mathrm{C}$ are dissolved in the ocean, $98 \%$ of which as carbonate molecules. The atmosphere contains $6.4 \times 10^{14} \mathrm{~kg}$ C as carbon dioxide $\left(\mathrm{CO}_{2}\right)$, as well as organic molecules like methane. Finally, carbon is essential to life and the mass of the carbon in living things (on land and in water) represents $3-6 \times 10^{14} \mathrm{~kg}$ [440].

The deep carbon cycle [441] describes the flux of carbon that involves the mantle and the crust [440]. The lithosphere can inject carbon into the atmosphere or the ocean notably through the fluxes of carbon dioxide released during volcanic activity ( 30 Mt C per year [442]) or as gasses trapped in the crust get released. Carbon contained deep within the crust can also reach the surface as it erodes: 43 to 97 Mt C become exposed due to erosion every year [443], [444]. Reciprocally, 70 Mt C return to the mantle every year through the subduction of the lithosphere [445]. The calcium carbonate $\left(\mathrm{CaCO}_{3}\right)$ stored in the shells of marine calcifying organisms can form limestone through sedimentation. The sedimentation and burial of terrestrial organisms can result, under high heat and pressure, in the formation of kerogen (fossil fuels) [446].

The deep carbon cycle is very important when considering large timescales, as it is through this process that the carbon present in the different reservoirs in the surface can be exchanged with some of the carbon present in the mantle. All these processes are however very slow or involve negligible quantities of carbon compared to the exchanges between the different surface reservoirs: the deep carbon cycle accounts for 1 to 10 Mt C exchanged per year, and it would take about one billion years for all the carbon currently contained in the mantle to cycle through [447]. Thus the next few paragraphs will focus on the exchanges between those faster moving surface reservoirs: fossil fuels, dissolved carbon molecules in the ocean, $\mathrm{CO}_{2}$ in the atmosphere and living beings.

Living biomass contains $300000-600000 \mathrm{Mt} \mathrm{C}$ [440], most of which is plants: the total biomass of plants contains $560000-615000 \mathrm{Mt} \mathrm{C}$ [448], [449]. The total exchanges of carbon between plants and the atmosphere result in carbon from the atmosphere being sequestered by the plants. Plants sequester 60000 Mt C every year [450]. That carbon is eventually released as the plant decays, burns or is eaten. Yet because the delay between the uptake and the release (particularly in the case of trees), plants can be a useful control mechanism for the carbon cycle [451].

The uptake process is called photosynthesis [452]. Under the presence of light energy, the following reaction occurs in autotrophic organisms:

$$
\begin{equation*}
2 \mathrm{H}_{2} \mathrm{~A}+\mathrm{CO}_{2} \longrightarrow 2 \mathrm{~A}+\mathrm{CH}_{2} \mathrm{O}+\mathrm{H}_{2} \mathrm{O} \tag{Equation 10}
\end{equation*}
$$

$A$ being the oxidising agent. In the case of plants, this general equation takes the following form:

$$
\begin{equation*}
6 \mathrm{H}_{2} \mathrm{O}+6 \mathrm{CO}_{2} \longrightarrow \mathrm{C}_{6} \mathrm{H}_{12} \mathrm{O}_{6}+6 \mathrm{O}_{2} \tag{Equation 11}
\end{equation*}
$$

The process of photosynthesis allows plant to store the energy from sunlight as sugar, and produces dioxygen $\left(\mathrm{O}_{2}\right)$. The quantum yield efficiency of photosynthesis by plants (sunlight absorbed by chlorophyll) is 0.081 mol of $\mathrm{CO}_{2}$ per mole of photons absorbed, when environment is optimal [453]. For each 30 g of carbohydrate produced in this reaction, 469 J is converted from sunlight to chemical energy tied up in sugars [454].

The $\mathrm{O}_{2}$ produced by the reaction can then be consumed by all aerobic organisms through aerobic respiration:

$$
\begin{equation*}
\mathrm{C}_{6} \mathrm{H}_{12} \mathrm{O}_{6}+6 \mathrm{O}_{2} \longrightarrow 6 \mathrm{H}_{2} \mathrm{O}+6 \mathrm{CO}_{2} \tag{Equation 12}
\end{equation*}
$$

About half of the carbon fixated by photosynthesis (Equation 11) is respired by plants (Equation 12), which is why the gross amount of carbon captured by the plant through photosynthesis is often twice the measured rate of carbon uptake [455]. More $\mathrm{O}_{2}$ is produced in ecosystems presenting high amounts of growth, such as young forests. Thus, more carbon is sequestered in spring and summer during daytime than in winter and at night, when plants rely on aerobic respiration rather than photosynthesis.

In the case of sea organisms, cellular respiration results in the increase in concentration of $\mathrm{CO}_{2}$ in the sea. For land organisms, the $\mathrm{CO}_{2}$ is released in the atmosphere, from which it might dissolve into the sea according to the following reactions:

$$
\begin{gathered}
\mathrm{CO}_{2(\text { atmos })} \rightleftarrows \mathrm{CO}_{2(\mathrm{aq})}+\mathrm{H}_{2} \mathrm{O} \rightleftarrows \mathrm{H}_{2} \mathrm{CO}_{3} \\
\mathrm{H}_{2} \mathrm{CO}_{3} \rightleftarrows \mathrm{H}^{+} \mathrm{HCO}_{3}^{-} \\
\mathrm{HCO}_{3}^{-} \\
\rightleftarrows \mathrm{H}^{+} \mathrm{CO}_{3}^{2-}
\end{gathered}
$$

Equation 13
Equation 14
Equation 15

The $\mathrm{CO}_{2}$ in the atmosphere dissolves into carbonic acid (Equation 13) which dissociates into bicarbonate (Equation 14), or into carbonate ions (Equation 15). Those two reactions are acid/base reactions (they rely on the transfer of a proton), and thus change the acidity of the system. According to Le Châtelier's principle [456], as more $\mathrm{CO}_{2}$ is added to the ocean, the equilibrium of the system shifts toward increasing the concentration of the product. The possible proportions of the different forms of dissolved inorganic carbon can be found in Figure 43.


Figure 43: Bjerrum plot describing the concentrations of carbonates in the ocean at $25^{\circ} \mathrm{C}$ and 5000 ppm . The concentration of $\mathrm{CO}_{2} / \mathrm{H}_{2} \mathrm{CO}_{3}$ determines the concentration of the other forms, and thus the pH of the ocean. Image taken from Wikimedia Commons [30].

Carbonate ions can react with calcium ions to form calcium carbonate $\mathrm{CaCO}_{3}$; these molecules can organise into crystalline forms such as calcite and aragonite. In calcifying organisms such as molluscs, the protein matrix of seashell is secreted out of the cells, which bind the calcium ions [457]. Thus, a solid shell of calcite or aragonite is created. A similar process can occur to form coral reefs [458]. 5 billion tons of $\mathrm{CaCO}_{3}$ are produced annually. As those calcifying organisms die, their remains can either dissolve into the ocean ( 2 billion tons per year) or aggregate to form sedimentary rocks ( 3 billion tons per year) [459], thus closing the carbon cycle. $20 \%$ of the carbon absorbed from the atmosphere by the ocean sinks into the deep ocean as organic matter [460]. $\mathrm{CO}_{2}$ from the atmosphere that dissolves into the surface layers of the ocean will eventually settle to the bottom after 1000 to 5000 years [461].

### 3.1.1.2 Impact of human activities on the distribution of carbon

As breathing organisms, humans have always had an impact on the $\mathrm{CO}_{2}$ concentration of the atmosphere. The $\mathrm{CO}_{2}$ production of human through breathing is 1 kg per day on average, and depends on the person's activity level [462]. This impact is however
negligible compared to the global production. Human impact on $\mathrm{CO}_{2}$ production became stronger through deforestation and the use of fire, which releases sequestered carbon [463], and through the development of agriculture. Agriculture modifies the emissions of carbon through diverse practices, such as land clearing, the replacement of perennial vegetation by annual crops, and nutrient subsidies in the form of fertilizers [464]. Between 4000 b.c.e. and 1850 c.e., the influence of human activity on the atmospheric concentration of $\mathrm{CO}_{2}$ amounted to a net carbon release of 79000 Mt C from permanent agriculture and 35000 Mt C from non-permanent agriculture [465]. From 1750 to 1850 , land use represented $28000-52000 \mathrm{Mt} \mathrm{C}[466]$.

The impact of humanity on atmospheric carbon emissions tremendously increased after 1850 , through the combustion of fossil fuels. Since the mid- $18^{\text {th }}$ century, 360000 Mt C have been burned, half of it in the last 40 years [467]. In the year 2015 alone, $3.2 \times$ $10^{13} \mathrm{~kg} \mathrm{CO}_{2}$ were released in the atmosphere from the combustion of fossil fuel [468]; as $\mathrm{CO}_{2}$ is 3.67 times heavier than C , this means the year 2015 alone represents $2.4 \%$ of the carbon emitted in the previous 165 years. Oil, coal and gas represent respectively $32 \%, 28 \%$ and $21 \%$ of the energy production and $34 \%, 45 \%$ and $20 \%$ of $\mathrm{CO}_{2}$ emissions from energy production [468]. Fossil fuels became the main source of $\mathrm{CO}_{2}$ emissions around 1920 [469]. $\mathrm{CO}_{2}$ emissions due to fossil fuels have been accelerating, with their growth rate increasing from $1.1 \%$ annually for the years 1990-1999 to $>3 \%$ for 20002004 [470]. Although the growth rate decreased in recent years (1.7\% in 2013, 0.6\% in 2014 and $-0.1 \%$ in 2015) [468], 2017 and 2018 have shown increases in $\mathrm{CO}_{2}$ emissions by $2.0 \%$ [471] and $2.7 \%$ respectively [472].

The $19^{\text {th }}$ century also saw the resurgence of concrete as a construction material [473]. The production of concrete requires the transformation of the $\mathrm{CaCO}_{3}$ in limestone into calcium oxide. This reaction releases the $\mathrm{CO}_{2}$ previously sequestered by the organisms whose skeletal fragments the limestone is made of. This represents $50 \%$ of the $\mathrm{CO}_{2}$ emissions due to concrete. $40 \%$ of the emissions are caused by the combustion of fossil fuels required in the manufacturing process. The remaining $10 \%$ come from the transport of concrete and the electricity otherwise consumed by the concrete industry [474]. Overall, the concrete industry represents $5 \%$ of anthropogenic $\mathrm{CO}_{2}$ emissions. Finally, 15 to 17 million hectares of forest are destroyed every year, which contributes to a release of $1200-3200 \mathrm{Mt} \mathrm{C}$ in the atmosphere [475].

The total anthropogenic $\mathrm{CO}_{2}$ emissions amounted to $10150-10250 \mathrm{Mt} \mathrm{C}$ in the year 2016 [476]. $30 \%$ of it was sequestered by plants on the land, $26 \%$ dissolved into the oceans, and $44 \%$ remained in the atmosphere, resulting in an increase of $4.81 \pm 0.04$ ppm atmospheric CO2 compared to the previous year, and a total of $399.4 \pm 0.1 \mathrm{ppm}$ [476]. This represents an increase of $44 \%$ compared to 280 ppm , the CO2 concentration that was maintained during the 10000 years before the industrial revolution [477]. Such high concentrations have occurred in the distance past (24 million years ago [478]), so the study of the climate from this era can give us an idea of how the geological impact of such a high $\mathrm{CO}_{2}$ level. The impact on current biology is harder to assess though, especially since this high $\mathrm{CO}_{2}$ concentration is coupled with a high rate of increase: current species did not evolve to endure the conditions that we can expect for the year 2100.

It is thus important to understand what these conditions will be, and to study their possible impact on current species. The next part will present the impact anthropogenic $\mathrm{CO}_{2}$ is having on the climate as well as on the ocean, and how those effects are projected to evolve over the century. The rest of this chapter will present the effect of those conditions on a marine species, the European sea bass.

### 3.1.1.3 The environmental consequences of carbon dioxide

The Earth receives an important part of its energy from the sun. The actual amount depends on the angle at which the sun's rays hit the Earth's surface, but the average is $340.2 \pm 0.1$ W. $\mathrm{m}^{-2}$ [479]. Part of it is reflected on (29\%), or absorbed by (21\%) the different layers of the atmosphere. The action of the atmosphere's layers varies with the wavelength of the radiations (see Figure 44); X-rays for instance are completely blocked, and so are most of the ultraviolet radiations. About $55 \%$ of the sun's light reaches the ground, increasing its heat [479]. The Earth's surface radiates this heat back at a wavelength that is a function of the surface's temperature. The effective radiating temperature of the Earth is 255 K [480]; at this temperature, a perfect emitter radiates in an infrared waveband of $4-100 \mu \mathrm{~m}$ (to be compared with the sun, 6000 K , which emits mostly in the $0.2-4 \mu \mathrm{~m}$ range, see Figure 44)


Figure 44: a: Radiations transmitted by the sun (red) and the Earth (blue); b: Total absorption and scattering of the electromagnetic spectrum by the atmosphere; c: Decomposition of the absorption and scattering by its major components: greenhouse gases and Rayleigh scattering. Image taken from Wikimedia Commons [30].

The atmosphere contains gases that are transparent to most radiations emitted by the sun, but absorb infrared radiation (see Figure 44). The heat absorbed is reemitted in all directions, further warming the Earth. This process, the greenhouse effect, ultimately results in a small part of the energy received from the sun remaining on the Earth's surface in the form of a slight increase in temperature, rather than being radiated back to space [479]. Greenhouse gases are very important for life, as without them, the Earth would be much closer to being a perfect emitter, and its average temperature would be $255 \mathrm{~K}, 32 \mathrm{~K}$ colder than its current average temperature of $14^{\circ} \mathrm{C}$ [481]. Water represents
$75 \%$ of the greenhouse effect, with $50 \%$ coming from water vapour, and $25 \%$ from clouds. $\mathrm{CO}_{2}$ represents $20 \%$ of the greenhouse effect. Water does play a role in the increase of the greenhouse effect, but it is as feedback from the pre-existing increase, which is due to the increase of the atmospheric concentration of $\mathrm{CO}_{2}$ [482].

The consequences of the increase of atmospheric $\mathrm{CO}_{2}$ and of the greenhouse effect are varied. They include an increased frequency of extreme weather [483], a higher sea level [484], a modification of the oceans' currents [485], as well as the consequences from such changes on biological systems and human populations. Some of the effects are negative feedback; for instance, plants react to more $\mathrm{CO}_{2}$ in the atmosphere with increased photosynthesis, which decreases the quantity of $\mathrm{CO}_{2}$. There is also positive feedback. The rise in temperature will lead to the thawing of the permafrost, thus releasing the methane trapped therein [486], another greenhouse gas. The melting of glaciers will reduce the albedo of the Earth's surface, increase the total amount of sunlight absorbed by the surface, and thus the temperature [487]. As the atmosphere gets warmer, more water vapour forms, thereby increasing the greenhouse effect [487].

The interaction of these different effects is difficult to predict, and because the main cause of climate change is human activity, all models are contingent on accurately predicting the rate of fossil fuel burning (although the consequences of Climate Change will remain long after humans have stopped using fossil fuel [488], [489]). The average global temperature has already risen by $0.85{ }^{\circ} \mathrm{C}$ since 1880 [490], and most models agree that the temperature increase will probably be between $1.5^{\circ} \mathrm{C}$ and $4.5^{\circ} \mathrm{C}$ by the end of the century. The impact of climate change is also already observable in the oceans, as their surface temperature has been increasing by $0.11 \pm 0.02{ }^{\circ} \mathrm{C}$ per decade over the past 40 years [491]. It is expected to increase by 2 more degrees by 2100 [477], [492].

The greenhouse effect is mitigated by oceans absorbing $30 \%$ of $\mathrm{CO}_{2}$ emissions [493], or $95000-129000 \mathrm{Mt} \mathrm{C}$ absorbed in the past 200 years [494]. If oceans did not act as a sink for $\mathrm{CO}_{2}$, the atmospheric concentration would be approximately 450 ppm instead of 400 ppm , and the consequences of climate change would be even more dire [495], [496]. The rate of absorption seems to have decreased in recent years, as oceans get closer to saturation [476].

This comes at a cost for the ocean though: as $\mathrm{CO}_{2}$ dissolves into the ocean, it forms carbonic acid $\mathrm{H}_{2} \mathrm{CO}_{3}$, thus decreasing the pH of the ocean by $0.0017 \pm 0.0004$ every year [497], [498] (the total decline since the beginning of the industrial era was from 8.179 to 8.104 [499]). $\mathrm{H}_{2} \mathrm{CO}_{3}$ reacts with carbonate ions $\mathrm{CO}_{3}^{2-}$ to form bicarbonate $\mathrm{HCO}_{3}^{-}$. The ocean is saturated with $\mathrm{CO}_{3}^{2-}$, which is essential for the metabolism of calcifying organisms [500]. Today, the upper Arctic Ocean is getting desaturated [501], a situation that will spread to other oceans as they continue to absorb $\mathrm{CO}_{2}$ [502]. In an ocean desaturated from $\mathrm{CO}_{3}^{2-}$, aragonite would dissolve (calcite, less soluble, would dissolve too as the ocean becomes more acidic) [503]. The loss of $\mathrm{CO}_{3}^{2-}$ from the ocean will also limit its ability to absorb $\mathrm{CO}_{2}$, therefore increasing global warming [504].

The biological impact of ocean acidification (OA) greatly varies depending on the species. Corals are very sensitive to the acidity level of the ocean; in fact, the structure of old corals has been used to deduce the acidity level at the time of growth [505]. A more acidic water can decrease their density and the lower concentration in $\mathrm{CO}_{3}^{2-}$ can impede their growth [506]. As the corals get more brittle, they become more at risk of getting damaged by storms [507]. Finally, as growth becomes more difficult, corals need to invest more energy into it rather than reproduction; this reduces the reef's larval output, and further impairs the chances for recolonization. As corals are also very sensitive to thermal stress, most of them might die by the time oceans' pH reaches 7.9 [508], which is predicted to occur before the end of the century [509]. Many species rely on coral reefs for survival, and their death would represent a monetary loss of $\$ 5$ billion, from tourism alone [510].

The response of marine calcifiers to OA is more varied in the animal kingdom. Many calcifying organisms (including crabs) elevate the pH of the water around the site of calcification; this increases the local concentration of carbonate ions and decrease the influence of OA on these organisms [511]. Some organisms also protect their shells with an external organic layer, further increasing their resistance to acidity. Ries et al. [511] observed that these strategies allowed some calcifying species, such as crustaceans, to even increase their calcification response pattern in more acidic water; but low pH has also been found to impair crustacean decision making process [512]. The net calcification of bivalve molluscs on the other hand decreased as pH got lower, and their shells even dissolved completely in the treatment with the highest $\mathrm{CO}_{2}$
concentration (corresponding to $2856 \pm 54 \mathrm{ppm}$ in the atmosphere, or 10 times the preindustrial atmospheric concentration) [511]. Other studies have found that mussels' and oysters' calcification ability would decrease by $25 \%$ and $10 \%$ respectively for the $\mathrm{CO}_{2}$ concentrations predicted for the end of the century [513].

Fish tend to be less impacted by OA [514]-[517], although the lower pH might affect their internal chemistry (intra- and extra-cellular) and reduce their blood's ability to transport oxygen [500]. pH plays a key role in different physiological processes such as enzyme activity and protein functions [518]. The lower pH can present benefits however: Larsen et al. [519] report that it offers protection against copper intoxication in Atlantic cods. Vertebrate cells use different regulatory systems that protect them from small variations [520], so OA does not present the same lethal risk as it does for calcifying organisms. It can however cause changes in behaviour [521]. More specifically, OA can affect fish olfactory perception [522], [523], vision [524], [525] and audition [526], in particular in regard to the detection of predators [527], [528] and preys [529]-[531]. Conversely, it can make fish more anxious [532]. On the other hand, higher temperature with no modification of pH has been found to improve anti-predator behaviour in the guppy [533]. OA can also interfere with neurotransmitter functions [534], which affects behavioural lateralisation [535]. OA can also impact homing ability [523], [536], swimming performances [537], foraging [538] and hunting behaviour [539], as well habitat preferences [540], [541]. OA can also have an impact on social interactions [542], [543], in particular shoaling behaviour: coral reef fish raised in OA conditions exhibit less group cohesion and a more random shoal distribution [544].

By the end of the century the atmospheric concentration in $\mathrm{CO}_{2}$ will be in the range of $985 \pm 97 \mathrm{ppm}$ [477], and climate change and ocean acidification will have profoundly transformed the oceans. Oceans' surface temperatures will be higher by $2-4{ }^{\circ} \mathrm{C}$ on average, and oceans' pH will drop by up to 0.4 units compared to its preindustrial value [509]. That pH perturbation may reach depths deeper than 1000 m [518]. In this chapter, I present the effects these conditions may have on the shoaling behaviour of the European bass, Dicentrarchus Labrax.

### 3.1.2. Sea bass

The European bass (see Figure 45a) is a species of fish that can be found around the coasts of Europe (except for the White and Barents Seas, the Baltic Sea and the Caspian Sea) and the Northern coast of Africa [545] (see Figure 45b). It is found in large quantities in the North Sea and the Irish Sea [546]. In their first year, sea bass tend to be found in the shallow waters of tidal lagoons; young bass spend most of the year near estuaries, or close to the coastline, and move to the deeper waters as they grow. In England and Wales, sea bass usually spend their first four years in estuaries [547]. Fully mature fish can go as deep as 80 m , depending on the location of their food. Sea bass exhibit annual migratory behaviour. Based on tagging experiments, Pickett and Pawson [546] report that sea bass around Britain and Britany tend to migrate to the North from February to June, and to the south from September to December. Those migrations can range from less than 100 km to 800 km for migrations to the Bay of Biscay.


Figure 45: European sea bass (a) and its habitat in red (b). Source: Wikimedia Commons [30]

Although bass are a sea water fish, sea bass of all ages can tolerate fresh water. Chervinski [548] has shown that although a direct transfer from seawater to tap water would result in death, sea bass can actually survive in tap water, provided that the dilution of the sea water is progressive enough. Adults can survive in temperature ranging from $8^{\circ} \mathrm{C}$ to $32^{\circ} \mathrm{C}$ [546]. In autumn, when the water in the estuaries drop below $15^{\circ} \mathrm{C}$, juvenile sea bass can leave for deeper, warmer waters; below $10^{\circ} \mathrm{C}$, they may stop feeding altogether [549]. The optimal temperature for the growth of juvenile sea bass is between $22^{\circ} \mathrm{C}$ and $24^{\circ} \mathrm{C}$ [550]; the effect of temperature on sea bass growth is usually greater than the effect of food availability [549]. More generally, sea bass tend to
concentrate in the warmer areas of their environment. They tend to adapt well to being kept in a transparent tank.

Because of the influence of temperature, the average size of sea bass at a given age depends on the region. A 1 year-old sea bass would be on average 7 cm long around the coasts of Ireland, but 14.5 cm near Morocco. This effect appears to be less strong for older sea bass: at age 7, average sizes vary between 36.8 in Egypt and 52 cm on the East coast of the UK [550], even though Egyptian seas are warmer. A possible explanation for this difference could be the influence of feeding [551]. At all ages, females are on average larger than males, with some specimens reaching sizes of 100 cm long. Female also have a longer life expectancy, the oldest being 30 years old (observed around the UK and Ireland). For instance, Kelley [552] reports that among a sample of 108 females and 62 males caught on the West coast of Wales, 20 females and 3 males were older than 10 years old and longer than 50 cm . The relationship between a sea bass' length and weight is proportional to a cubic function, with the coefficient of proportionality depending on the environmental conditions. In particular, Pickett and Pawson [546] report that this coefficient can vary greatly depending on the month, from $6 \times 10^{-6}$ in November to $33 \times 10^{-6}$ in July. The maximum theoretical weight for a European sea bass is estimated to be 11.75 kg [550].

Metamorphosing larvae do not exhibit shoaling behaviour, but postlarval sea bass do gather around physical barriers in estuaries [553]. Past that stage, sea bass exhibit shoaling behaviour their entire lives, with shoals ranging from a few dozen individuals to many thousands [546]. Shoals as large as 800 m long $\times 150 \mathrm{~m}$ wide $\times 2 \mathrm{~m}$ deep have been observed through sonar technology, which would suggest a number of individuals in the order of 10 million [546].

Globally, sea bass represent a growing market of $\$ 1.5$ billion per year, with 169900 tons produced in 2017 [554]. Although a large part of the global production comes from aquaculture, $10 \%$ of the production comes from fishing [555]. Thus, predicting the influence of future oceanic conditions on this species has important economic applications.

### 3.1.3. Research question

Models of the effect of climate change on ocean temperature predict an increase in temperature in the dwelling areas of the sea bass by $2^{\circ} \mathrm{C}$. Although sea bass are able to handle a great range of temperatures and tend to prefer warmer water [556], the more frequent heat waves and more extreme temperatures might negatively affect their fitness [557]. More precisely, Malavasi et al. [558] have shown that the cohesiveness of sea bass shoals is better at $18^{\circ} \mathrm{C}$ than at $22^{\circ} \mathrm{C}$ or $26^{\circ} \mathrm{C}$, and deduced that the antipredator response was dependent on temperature. The effect of OA will be stronger near the surface, which is where juvenile sea bass live. Sea bass can survive extreme acidification, up to $35 \mathrm{mmHg}(46000 \mu \mathrm{~atm}$ ) [559], however Cecchini et al. [560] found that increased the $\mathrm{CO}_{2}$ concentration of the water affected feed intake of sea bass and affected their blood chemistry in terms of $\mathrm{pH}, \mathrm{pCO}_{2}$ and plasma. Crespel et al. [561] studied the influence of different projected levels of pCO 2 on sea bass, and found little effects of the low acidification projection ( $980 \mu \mathrm{~atm}$ ), and high acidification (1 520 $\mu \mathrm{atm})$ would in fact improve skeletal development.

In this study, I compare the behaviour of sea bass raised under normal ocean water $\left(17^{\circ} \mathrm{C}, 500 \mathrm{ppm} \mathrm{CO} 2\right)$ and the conditions that are expected to occur by $2100\left(19^{\circ} \mathrm{C}\right.$, $1000 \mu \mathrm{~atm} \mathrm{pCO}_{2}$ ). I focus more precisely on behavioural variables that are commonly used as measures of inter-individual differences in behaviour within populations [175], [562] and which could influence shoaling behaviour. Significant differences in fish behaviour across the three experimental conditions would provide evidence for behavioural alterations in the sea bass in response to OA. By contrast, if fish showed statistically similar movements and interacted with their environment and each other in comparable ways, this would indicate behavioural resilience to OA. Given that the laboratory environment could constrain fish behaviour (e.g. the restriction of movement to the test arena), I also present an artificial agent-based model of the experiment, which I compare with the experimental data, making it possible to gauge whether the analysis method is sufficient for detecting subtle changes in fish behaviour.

### 3.2. Material and methods

### 3.2.1. Animals

Fertilised Dicentrarchus Labrax eggs from a mixed spawn (multiple males and females) were purchased from the Écloserie Marine de Gravelines, France. They were first transferred to $12 \times 10 \mathrm{~L}$ incubators inside $12 \times 150 \mathrm{~L}$ experimental tanks (each tank containing three incubators). All incubators were kept at $13^{\circ} \mathrm{C}$ and ambient $\mathrm{pCO}_{2}$. As the bubbling and the agitation could have impacted the behaviour of the fish, the $\mathrm{pCO}_{2}$ was regulated in a different header tank, from which it was pumped to the tanks that contained the fish (see Figure 46). Upon hatching, the larvae were transferred to the experimental tanks and the incubators were removed.


Figure 46: One of the systems used to maintain $\mathrm{pCO}_{2}$ and temperature throughout the growth of the fish. Before hatching, the eggs were kept inside 10 L incubators (three incubators per system). All incubators were kept at $13^{\circ} \mathrm{C}$ and ambient $\mathrm{CO}_{2}$. After hatching, the larvae were transferred to one of the experimental tanks, where the temperature and $\mathrm{pCO}_{2}$ were progressively adjusted to the desired conditions for the system $\left(17^{\circ} \mathrm{C}\right.$ or $19^{\circ} \mathrm{C}$ and $585 \mu \mathrm{~atm}$ or $1000 \mu \mathrm{~atm} \mathrm{pCO}_{2}$ ). A rhythm of 12 hours light / 12 hours dark was kept inside the experimental tanks. RAS = recirculating aquaculture system. Image from [563].

For the "acidic ocean" conditions, the $\mathrm{pCO}_{2}$ was brought to $1000 \mu$ atm over 24 hours. As the fish were breathing $\mathrm{CO}_{2}$ into the water, the $\mathrm{pCO}_{2}$ in the tank was $200 \mu \mathrm{~atm}$
higher than it would be from the dissolution of the atmospheric $\mathrm{CO}_{2}$ alone. The temperatures in the tank were adjusted at a rate of $1^{\circ} \mathrm{C}$ per day, to $17^{\circ} \mathrm{C}$ or $19^{\circ} \mathrm{C}$. The fish observed in this study were either raised in the "ambient" condition ( $585 \mu \mathrm{~atm}$ $\mathrm{pCO}_{2}, 17^{\circ} \mathrm{C}$ ) or the "ocean acidification condition" (OA: $1000 \mu \mathrm{~atm} \mathrm{pCO}_{2}, 19^{\circ} \mathrm{C}$ ). In both conditions, the salinity in the experimental tanks was maintained to $28.17 \pm 0.22$ according to the practical salinity scale [564]. A 12 hour period day/night cycle was maintained in the experimental tanks (median light $=6.5 \mu \mathrm{~mol}$ photon $\mathrm{m}^{-2} \mathrm{~s}^{-1}$, range $=$ $4.2-12.4 \mu \mathrm{~mol}$ photon $\mathrm{m}^{-2} \mathrm{~s}^{-1}$ ). Alkalinity levels were also controlled throughout the growth of the larvae twice a week.

### 3.2.2. Observations

The fish were observed by groups of 10 for 1 hour each, between day 59 and day 68 after hatching; at this age, sea bass are entering the juvenile stage. The observations were conducted on 90 fish from the OA condition (mean length $\pm$ s.d. $=35.35 \pm 4.61$ mm ) and 180 fish were randomly selected from the ambient condition (mean length $\pm$ s.d. $=32.34 \pm 3.12 \mathrm{~mm}$ ). Just before each observation period, 3 groups of 10 sea bass were collected from their home tank with a sweep net, and each group was placed in an observation tank, in one of the three circular Plexiglas 'test arenas' ( 30 cm diameter, 3 cm height) (see Figure 47). The bottom of the arena was a mesh, which allowed water to circulate through in and out the arena and through the tank. Test arenas were separated with opaque plastic partitions and surrounded by a custom-built aluminium frame and white screen (PhotoSEL BK13CW White Screen). A Panasonic HDC-SD60 HD video camera (Panasonic Corporation of North America, Secaucus, NJ, USA) was positioned above each test arena. Four photographer's lights (each with $4 \times 25$ W 240 V 6400 K True Day light bulbs) lit the arenas from the outside, hence dispersing light evenly over the arenas and enabling optimum conditions for video recording.


Figure 47: a: Observation setup. Groups of ten fish are observed in each arena. Each arena is isolated from the others by an opaque plastic partition so the fish are not influenced by their conspecifics in the other arenas. b: Two views of an arena Since the fish swim in shallow water, and the arenas are recorded from above, trajectories can be considered as 2 dimensional.

9 groups of 10 fish from those reared in ambient water were observed in ambient water. The 18 other groups of fish were observed in OA water. Thus 9 groups of 10 fish were observed for each of the following conditions: fish reared and observed in ambient water ("ambient"), fish reared and observed in OA water ("OA"), and fish reared in ambient water and observed in OA water ("ambient $\rightarrow$ OA"). No fish reared in OA water were observed in ambient water, as this condition would not represent any expected ecological conditions. Video recordings were started approximately 60 s after the fish were placed into the arena, so the fish haven't had time to explore the arena, but still have time to recover from the transfer to the arena. Arenas were recorded for 60 min of footage ( 90000 frames at 25 Hz ; representing 90000 data points per fish). In total, 26 hours of footage were recorded, as a processing error occurred in one of the video files for the ambient $\rightarrow \mathrm{OA}$ conditions.

### 3.2.3. Tracking

Experiments were recorded as AVCHD videos, and converted into mp4 using Freemake (v. 4.1.5.4) [565]. As one of the cameras did not work properly in the ambient $\rightarrow \mathrm{OA}$ condition, this gave 261 -hour recordings of the behaviour of sea bass in the three conditions (all the cameras that worked properly produced more than 1 hour of recording. All the recordings were processed entirely, but only 1 hour of data were utilised for the analysis, so all conditions would have similar amounts of data). I then used idTracker [81] to track the movement of the fish in the film. This software extracts a characteristic fingerprint from each animal in a video recording of a group, and then uses these fingerprints to identify every individual throughout the video. In this particular image detection problem, it is important to note that the appearance of the objects can completely change throughout the video, as the fish bend their tails one way or another for instance. By using a large number of frames of reference (typically, more than 1000), idTracker is able to reliably identify and label the object in all the positions it displays in the recording.

For each video, I restricted the analysis to the arena, and selected the minimum size and intensity threshold in a way that the ten fish were detected, and no artefacts were labelled as fish, in ten frames randomly chosen in the video (see Figure 48). I also removed the background, which for some videos required finding a selection of frames in which all fish were moving sufficiently to be detected by the software. idTracker returns the x , y coordinates for each individual fish in each frame (temporal resolution of 25 Hz ); these coordinates tend to designate the centre of the fish. The limited water depth ( 3 cm ; see Figure 47b) ensures accuracy in the tracking in two dimensions and calculation of subsequent behavioural variables (see below), but also restricted fish movement in the vertical plane.


Figure 48: The interface of idTracker. As some of the recordings included partial views of the other arenas, it was necessary to restrict the tracking area to the arena (walls excluded, to avoid tracking the reflections of the fish). The role of the intensity threshold and minimum size is to remove the noise. The number of frames of reference determines the size of the learning base for the algorithm, to decide the nature of the "fingerprint" for every fish. See [81] for details.

For all but one of the videos, I directly used the output of idTracker from the nogap mode, in order to avoid losing data when fish cross each other. Indeed, although the shallow depth of water makes it possible to approximate the fish trajectory as 2 dimensional, fish did sometimes cross one another. In the recording, the moment of crossing appears as a moving object that does not properly matches the fingerprint of either fish. In the nogap mode, idTracker ignores those objects and treats them as the intersection of the fish whose extrapolated position is closest.

For the remaining video, I used idPlayer to control the labels of the fish determined by idTracker. I manually eliminated all artefacts and corrected any change of labels over time. As only coordinates at which moving objects had been identified could be labelled, I then reran idTracker with the corrected data. Finally, I ran idPlayer in the nogap mode on the result of that tracking, so the software would interpolate the correct trajectory of the fish from the data I had input manually and resolve the occlusion events. This last step also allowed me to check that the labels were still coherent.

Normally, the determination of fingerprints by idTracker would prevent any label swapping. idTracker is meant to be used with HD video files, yet the specific format in which the video camera recorded was incompatible with Matlab and thus idTracker. This is the reason why the videos had to be converted to mp4. Unfortunately, this resulted in an important loss of data, making the fish difficult to distinguish by the human eye or the software (see Figure 49). So although idTracker was usually successful at determining the position of all 10 moving objects (fish) in the frame, it was not always able to properly label them, in particular in cases of occlusions. By manually tracking one of the videos, I obtained a "gold standard" to which the other parts of the data set could be compared.


Figure 49: One frame of a video with the associated labels. idTracker associates a different label to every moving object in the frame, up to the number of individuals to detect (here, 10). The objects for which a label can't be determined are given the label 0 . Due to the low quality of the video, the fish present little feature allowing idTracker to detect them reliably.

### 3.2.4. Postprocessing

I observed several types of errors when processing the video with idTracker normally. I also needed to account for the slight differences of position of the arena relative to the
camera (due to vibrations in the room). All the postprocessing of the idTracker data was done in Matlab (version R2014b).

### 3.2.4.1 Scaling

The first step was to compensate for the differences in scale between videos. The sea bass had a tendency to follow the border of their tank. In every video, at least one pair of diametrically opposed points were visited by some of the fish, so the length of the diameter could be easily obtained by measuring the extreme distance between two points. This made it possible to determine the scale of the picture (it was not possible to place all cameras at exactly the same height, resulting in small variations of scale). I normalised the images so that all arena measured 1000 pixels in diameter, hence all the data I processed were at a scale of $0.03 \mathrm{~cm} /$ pixel.

### 3.2.4.2 Error elimination

- When fish swim near the edge of the tank, their reflection sometimes appear on the video and can be detected by idTracker as another fish. Excluding the edge of the arena from the tracking area was not always possible, because since the fish were often swimming very close to the edge, part of them would have been excluded from the processing, making them unrecognisable to the software. These reflections can be easily detected as they draw an arc of circle concentric to the arena (see the data in red in Figure 51 for an example). For the recordings where those were identified, I rescaled the data until all the other points just fit inside the disc of centre ( $0 ; 0$ ) and diameter 1000 pixels ( 29 cm ), then I eliminated all points that lay outside the arena.
- A common source of errors was the label swapping, when a fish is confused with another, or with an artefact. For each frame and for each fish, I calculated the instantaneous displacement $d_{t}$ :

$$
\begin{equation*}
d_{t}=\sqrt{\sum_{i=1}^{2}\left(X_{i}(t)-X_{i}(t-1)\right)^{2}} \tag{Equation 16}
\end{equation*}
$$

with $X_{i}$ being the coordinates of the fish and $t$ designating the current time frame. Those speeds were compared with the maximum possible speed, considered to be the fastest speed detected in the gold standard, or $0.91 \mathrm{~m} / \mathrm{s}$. This speed is coherent with the
maximum speed observed by Claireaux et al. [566] for juvenile sea bass at these temperatures. I also eliminated coordinates as possible artefacts when the position of the fish was not known for more than three consecutive frames.

At this point in the processing, all the types of errors that had been identified had been corrected, except for label swapping: in other words, all data points match the coordinates of a fish, even though its label might change throughout the recording. Examples of errors can be seen in Figure 50.


Figure 50: Example of erroneous data removal: in red, the points removed for being outside of the arena ( 510 points); in cyan, the points removed for excessive speed (8); in purple, isolated points (102). The other 78222 points (in blue) are considered correct. The rest of the 90000 time points had no associated coordinates, explaining the lack of continuity in the trajectory detected by idTracker.

Before the errors were eliminated, $4.7 \%$ of data points were missing, and $77.7 \%$ of frames had at least one fish missing. This processing eliminated the $0.1 \%$ of data points that were considered as erroneous. Because the label swapping errors could not be removed, it is not possible to compare the behaviour of fish from the same arena. Comparisons can still be made between the properties of different groups, as well as between the distributions of individual properties between different groups.

### 3.2.4.3 Interpolation

I then interpolated the sections of missing points that were sufficiently short. For every sequence of missing data shorter than one second, for which I knew the position of the fish for the five frames just before and just after that sequence, I computed a piecewise cubic interpolation to determine the polynomial which approximated the best the missing values, for both coordinates of the fish. This step allowed me to determine 104 143 more data points, giving us a total of 2563633 frames with all fish (1 824637 frames when considering only the 1 hour sample of the recordings that are to be analysed). After processing, $5 \%$ of the data was missing, and $76 \%$ of frames had all fish detected.

The final data set contained more than 31 million data points, $28.7 \%$ of which came from the ambient condition, $27.9 \%$ from the OA condition and $43.4 \%$ for the ambient $\rightarrow$ OA condition (respectively $35.0 \%, 34.6 \%$ and $30.4 \%$ for the reduced data set used in the analysis).


Figure 51: Trajectory of the fish observed in Figure 50, after the data have been processed. Because data is missing for some time points, the trajectory is not continuous.

### 3.2.5. Data extraction

I calculated different variables that described individual and collective movements of the fish. The individual variables calculated were:

- the time spent near the edge (TEd),
- the median over time of the speed of each fish (ms);
- the time spent still (TS),
- the length of time the fish required to explore $10 \%$ of the arena (Exp10),
- the total area explored after an hour $(\operatorname{Exp} T)$.

The group variables were:

- the median distance to the nearest neighbour ( $D N N$ )
- the local group size ( $L G S$ ), which is related to the number of fish close to any given fish.


### 3.2.5.1 Local variables

I considered that a fish stayed still between two frames if its movement was less than 1 pixel of the scaled image, which amounts to a speed of $0.7 \mathrm{~cm} . \mathrm{s}^{-1}$. The speed distributions for each condition can be found in Figure 52


Figure 52: Distribution of speeds for all conditions. a: ambient, $\mathbf{b}$ : $\mathrm{OA}, \mathbf{c}$ : ambient $\rightarrow \mathrm{OA}$. The box plots represent the proportion of time spent moving or still for all fish.

The time spent near the edge describes the proportion of time when the fish is within 3 cm of the edge, which is about the length of the longest fish; this length represents $20 \%$ of the radius of the arena. The tendency to remain near walls, or thigmotaxis, is considered as a sign of shyness or anxiety in different species [567]-[569], including European sea bass [570]-[572]. Figure 53 shows that thigmotaxis is present in all three conditions.



Figure 53: Density plot of the positions of all the fish from each conditions (a: ambient; $\mathbf{b}$ : $\mathrm{OA} ; \mathbf{c}$ : ambient $\rightarrow \mathrm{OA}$ ). The colour bar corresponds to the number of frames in which one of the 90 fish from the corresponding condition is in one pixel. The fish are most commonly near the edge, showing that thigmotaxis happens in all conditions.

To determine the exploration variables, I discretised the arena into a grid whose squares were 1 cm in width (the size of the smallest fish observed); a $100 \%$ exploration means that the fish had entered at least once each of the 788 squares of the grid (see Figure 54).


Figure 54: a: One example of trajectory and the corresponding explored area. The white area represent the surface that is considered to be inside the tank, the grey area represent the fraction of it that has been explored by the fish; $\mathbf{b}$ : Progression of the exploration for the fish across all conditions. The turquoise line represents the median of the areas explored at any time point, the grey lines represent the first and third quartiles of the area observed by all the fish. The total area explored at the end of the recording ranges from $11.7 \%$ to $92.5 \%$.

### 3.2.5.2 Global variables

The global parameters were the distance to the closest fish ( $D N N$ ) and the local group size (LGS). More precisely, since all variables are defined for one fish, $D N N$ is the median of the distance to the fish that is closest to the focal fish in the frame, over all frames (i.e., the closest fish can change between fish. As label swapping is a possibility, it is not possible to determine whether two given fish tend to remain close. Thus, this parameter describes the geometry of the shoal, rather than inter-fish relationships). The distribution of inter-fish distances can be found in Figure 55.


Figure 55: Distribution of inter-fish distances for each condition (a: ambient, $\mathbf{b}$ : OA, $\mathbf{c}$ : ambient $\rightarrow \mathrm{OA}$ ). The red line represents the median distance between all fish in the arena, for all arenas: fish that are closer than this distance are considered as "connected".

I considered that two fish were "connected" to each other if they were closer than the median distance between fish (across all conditions), which was determined to be 6.9 cm over the whole data set (see Figure 56). LGS is defined as the average over time of the median proportion of the other fish in the group that are connected to the focal fish. For instance, if a fish is close to every other fish in all frames, its $L G S$ score will be 9 ; if it is always far from the other fish, it's score will be 0 .


Figure 56: Neighbourhood area of a focal fish (the fish being studied) and the fish connected to it. In this example, four fish are close enough to the focal fish (in green) to be considered as connected to it. If the situation stayed the same for the entire experiment, the connectivity for this focal fish would be $44 \%$. This figure illustrates the fact that being close to the edge limits the area where the other close fish could be.

### 3.3. Data analysis

Since the 7 variables I chose for the analysis were correlated, I conducted a Principal Component Analysis (PCA) [573] to reduce the number of variables. I first normalised them:

$$
\begin{equation*}
X_{i} \rightarrow \frac{X_{i}-\operatorname{mean}(X)}{\operatorname{std}(X)} \tag{Equation 17}
\end{equation*}
$$

for all variables $X, i$ representing each individual fish. I then combined these normalised variables into principal component scores using the Matlab pca function. PCA is a way to compress the data set to a lower dimension space, while extracting the most information from the original data. It simplifies the description of the data set by highlighting the correlations between the different variables. This is achieved by computing new variables that are linear combinations of the original variables. The first principal component is computed in a way that maximises the variance it explains, and each of the following components is computed so as to be orthogonal to all previous components and to maximise the variance. Thus, the first few variables concentrate most of the variance, and the dimension of the problem can be reduced while keeping an accurate representation of the data [574]. Finally, I used a varimax rotation [575] (the rotatefactor function in Matlab) on the resulting factors. This rotation makes the loading factors from the original data set to the new variables closer to either 0 or 1 , making the new factors simpler to interpret as a combination of the original factors.

To test for differences in the behaviour of fish across all three experimental conditions, I fitted different linear mixed models (LMMs) with the first three principal components as the response variables, and the experimental condition ('ambient', 'OA', 'ambient $\rightarrow \mathrm{OA}$ ) as a categorical fixed effect. Linear mixed models are a way to model the response variables as a linear combination of fixed effects and the random effects that need to be controlled for. Here, because each day of recording, three arenas were recorded simultaneously, and sharing the same water, the variables describing fish from the same tank or arena might not be independent, so tanks and arenas were both included as random effects.

All statistical analyses were conducted in the R environment [576] and LMM were fitted using the package lme4 (linear modelling) [577]. The level of significance was set to $\boldsymbol{\alpha}<0.05$. The residuals represent the variability that is not due to the random effects included in the model; for the model to be valid, the residuals must be normally distributed. This was verified through graphical procedures (Q-Q plot and histograms of residuals; see Figure 57).


Figure 57: Graphical control of the validity of the model for one of the principal components: the Q-Q plot is close to being linear (a), and the residuals are normally distributed (b). Similar distributions were observed for the other principal components.

### 3.4. Results

The first 3 principal components accounted for $84.7 \%$ of the variance. Fish that scored high on PC1 tended to (i) be close to other fish, (ii) have a high number of local neighbours, and (iii) explore less of the arena. Fish that scored high on PC2 tended to spend most of their time at the edge of the arena. Fish that scored high on PC3 tended to spend little time still, and explored a large proportion of the arena. On this basis I labelled PC1 'sociability', PC2 'shyness', and PC3 'activity' (see Table 3).

Table 3: Factorial decomposition of the first three principal components (PC). These three PC represent $84.7 \%$ of the variance in the data. The most important variables for each component are in bold

|  | PC1 <br> 'sociability <br> , | PC2 <br> 'shyness <br> , | PC3 <br> 'activity |
| :--- | :---: | :---: | :---: |
| proportion of time near the arena edge TEd | 0.064 | $\mathbf{0 . 9 8 4}$ | -0.021 |
| median speed $m s$ | 0.025 | -0.031 | $\mathbf{0 . 5 2 0}$ |
| proportion of time spent still $T S$ | -0.026 | 0.001 | $\mathbf{- 0 . 7 0 9}$ |
| time taken to explore 10\% of the arena Exp10 | 0.245 | 0.098 | -0.114 |
| overall \% of the arena explored Exp $T$ | $\mathbf{- 0 . 6 0 4}$ | 0.134 | 0.361 |
| median distance to nearest neighbour $D N N$ | -0.271 | -0.049 | -0.137 |
| mean number of local neighbours $L G S$ | $\mathbf{0 . 7 0 4}$ | -0.026 | 0.254 |

All three PC were statistically similar across the three treatments ( $\mathbf{p}>0.05$; see Table 4 and Figure 58). R compares the different conditions to the intercept. For numerical variables, it is the estimated value of the response variable when the value of the fixed effects is 0 . For categorical variables, the intercept is considered by default to be the condition that comes first in the alphabet (here: ambient). Thus the estimates for each line correspond to the difference with the estimate for the 'ambient' condition. The $t$ value is defined as the estimate divided by the standard error.

Table 4: Results of model testing for differences in the behaviour of fish across experimental conditions. Presented are the estimates, associated standard error (s.e.), t-value and p-value with respect to PC1 'sociability', PC2 'shyness' and PC3 'activity.

| variable | estimate | s.e. | $\boldsymbol{t}$-value | $\boldsymbol{p}$-value |  |
| :--- | :--- | :--- | :--- | :--- | :---: |
| PC1 | 0.303 | 0.438 |  |  |  |
|  |  |  |  |  |  |
| intercept | -0.321 | 0.620 | -0.519 | 0.604 |  |
| ambient versus OA | -0.625 | 1.323 | -0.472 | 0.637 |  |
| ambient versus ambient $\rightarrow$ OA |  |  |  |  |  |
| PC2 | 0.176 | 0.274 |  |  |  |
| intercept | -0.407 | 0.388 | -1.050 | 0.293 |  |
| ambient versus OA | -0.115 | 0.696 | -0.165 | 0.869 |  |
| ambient versus ambient $\rightarrow$ OA |  |  |  |  |  |
| PC3 |  |  |  |  |  |
| intercept | 0.178 | 0.214 |  | 0.245 |  |
| ambient versus OA | -0.351 | 0.302 | -1.162 | 0.769 |  |
| ambient versus ambient $\rightarrow$ OA | -0.184 | 0.629 | -0.293 |  |  |



Figure 58: Comparison across the three principal components of the three conditions. Linear mixed model analysis revealed that treatment type does not predict variation in (a) PC1 'sociability', (b) PC2 'shyness' or (c) PC3 'activity'. See Table 3 for details of how behavioural variables load onto PC axes and Table 4 for statistical comparisons.

### 3.5. Computer model

Although observing fish in a simple environment makes their automated detection simpler, there is a risk that the environment might have impacted their behaviour: for example, the size of the arena could have constrained the average inter-individual distance or the size of the local group. If that were the case, then fish would be found to behave in the same way regardless of whether their behaviour would be distinct in the wild. In order to ensure that the results did not depend on the experimental setting and could be generalised to a natural environment, I built an artificial model of the experiment.

The goal of the model is to simulate the individual behaviour of the fish, without implementing any interaction rule. If the group variables computed for the simulated fish were found to differ from those of the real fish, it would confirm that the behaviour of the real fish is indeed impacted by that of the other fish in the arena, without making any hypothesis regarding the interaction rules actually followed by the real fish. It would show that fish are able to behave in ways that differ from the behaviours observed in all three experimental conditions, and that the reason fish behave in similar ways regardless of the condition is that these differences of environments do not impact the fish behaviour. In other words, the goal of this model is to show that there is a difference in behaviour between a simulated fish that does not have any social behaviour and the fish observed experimentally. Because fish were found to behave in the same way regardless of conditions, the individual behaviour of the simulated fish is based on data gathered from the entire recordings for all three conditions. The behaviour of the group in the computer model is to be compared with the experimental group data, regardless of conditions.

In a first version of the model (model 1), I considered that in the absence of other fish, the instantaneous movement of a fish could be described by its position (since the environment is symmetrical, the position can be reduced to its distance to the centre), its angle relative to the radius, and its previous speed and angular speed. In all of the recordings, I extracted in a table these four variables for all fish for all time points, as well as two related to its next movement: the speed and angular speed at the next time point.


Figure 59: The parameters that describe the state of a fish. $r$ is the distance of the fish to the centre of the arena, $\alpha$ is its angle to the radius, $v$ is its instantaneous speed, $\omega$ is the angle of the rotation toward the next position of the fish. Any state a fish can be in can be described by a quadruplet $\{r ; \alpha ; v ; \omega\}$, with $\omega$ being the instantaneous rotation. By convention, if the fish is not moving, $\omega$ is considered to be $0 . \alpha$ is also considered null if the fish is close to the centre of the arena. $\theta$ represents the angle of the fish compared to the angle of the arena.

Let F be the function that to a four dimensional vector $(r, \alpha, v, \omega)$ describing the state of a fish (see Figure 59) associates a velocity vector $\left(v_{+}, \omega_{+}\right)$. The data table can be considered as a list of 31 millions couples of quartets ( $r, \alpha, v, \omega$, first four columns) and the speed vectors ( $v_{+}, \omega_{+}$, last two columns) that are their image through F. Some representations of the data can be found in Figure 60. The data show that fish are most often near the edge of the arena (the proportion of data recorded near the edge is higher than the proportion of the total surface that is near the edge), and although the distribution of orientations is close to being uniform, there are two small spikes around $\pi / 2$ and $-\pi / 2$, corresponding to fish swimming parallel to the edge. Rotation angles are mostly null; the comparison with speeds shows that speeds significantly higher than 0 are almost only associated with small rotation angles, meaning that large rotation angles corresponds to very low speeds and thus mostly corresponds to noise in the recording.


Figure 60: Distribution of a sample of 25000 data points of the experimental data, represented along the four dimensions distance, orientation, speed and rotation. R is in $\mathrm{cm}, \alpha$ in $\mathrm{rad}, \mathrm{v}$ in $\mathrm{cm} / \mathrm{s}$ and $\omega$ in $\mathrm{rad} /$ frame (since $\omega$ is the instantaneous rotation, the time scale should be the frame rather than the second). Each row and column represents one of the dimensions; the scatter plot at the interaction of a row and column represents the distribution of the data sample along the two dimensions. On the diagonal are the histograms of the individual dimensions; the vertical scales go from 0 to respectively 6000, 2500, 13000 and 10000.

At $t=0$, each fish is initialised with a random state $s_{i}(0)$ drawn from the table, as well as a random angle $\theta_{i}(0)$ that describes its position in the arena. Each state corresponds to coordinates of a point $p$ in the 4-D space that describes all possible fish states (the combinations of possible distance to the centre of the arena, angle of the fish, instantaneous speed and instantaneous rotation). The 10 points $p_{i}\left(v_{+}, \omega_{+}\right)$of the data table that are closest to the current state of the fish in the 4-D state space are found, and the velocity vectors associated with these points through function F (last two columns of the table) are utilised to deduce the next velocity $v$ and instantaneous rotation $\omega$ of the fish:

$$
\begin{align*}
& v(t+1)=\frac{\sum_{i}^{10} v_{+}\left(n_{i}\right)}{10}+\varepsilon_{1}  \tag{Equation 18}\\
& \omega(t+1)=\frac{\sum_{i}^{10} \omega_{+}\left(n_{i}\right)}{10}+\varepsilon_{2}
\end{align*}
$$

Equation 19

In Equation 18 and Equation 19, $n_{i}$ represents one of the 10 entries from the data table that are the closest to the current state of the fish; $v_{+}$and $\omega_{+}$represent the first component (speed) and second component (instantaneous rotation) of the output of the function $F$ for those entries; $\varepsilon_{1}$ and $\varepsilon_{2}$ represent noise. The coordinates of the fish at $t+1$ are then determined by the following equations:

$$
\begin{array}{cc}
r(t+1)=\sqrt{r(t)^{2}+2 v(t) \times \cos (\alpha(t)+\omega(t))+v(t)^{2}} & \text { Equation } 20  \tag{Equation 20}\\
\theta(t+1)=\operatorname{atan}\left(\frac{r(t) \times \sin (\theta(t))+v(t) \times \sin (\alpha(t)+\theta(t)+\omega(t))}{r(t) \times \cos (\theta(t))+v(t) \times \cos (\alpha(t)+\theta(t)+\omega(t))}\right) & \text { Equation 21 }
\end{array}
$$

The new state of the fish is the following vector:

$$
s(t+1)=\left(\begin{array}{c}
r \\
\alpha \\
v \\
\omega
\end{array}\right)(t+1)=\left(\begin{array}{c}
r(t+1) \\
\alpha(t)+\theta(t)+\omega(t)-\theta(t+1) \\
v_{+}(s(t)) \\
\omega_{+}(s(t))
\end{array}\right) \quad \text { Equation } 22
$$

If the radius $r(t+1)$ is larger than the radius of the arena, other random values are drawn for $\varepsilon_{1}$ and $\varepsilon_{2}$, until the new position of the fish is valid. The space from which these variables are drawn expands with every iteration so the program does not get stuck (since the new variable is computed by averaging the outputs of $F$, rather than directly drawing one of the outputs, this new coordinate is not necessarily close to a valid new state).

The successive coordinates of each fish are thus computed for every frame until the end of the simulation ( 90000 frames). The advantage of model 1 is that the behaviour of every fish closely matches that of a modal fish if it did not consider the behaviour of the other fish, since the model draws from all information available to the fish and from actual observed behaviour. Unfortunately, model 1 requires a lot of computation in
order to determine the 10 closest states, due to the size of the data base, so it cannot be used. One possible solution would be to only consider a random sample of those states, but in this case the search space would be too sparse, due to the curse of dimensionality [578]: it would be unlikely to find 10 points that would be sufficiently close to the desired state on all 4 dimensions. Thus a simpler model is required in order to generate enough data to compare with the experimental data and confirm that a model with no interaction between fish does not accurately describe the experimental data.

Model 2 is a simplified version of model 1 : at every time step, an output from the data table is drawn randomly and the new state of the fish is computed with Equation 20 and Equation 21, independently of its previous state. In case of invalid coordinates (the fish would leave the arena), the new coordinates are replaced with the symmetrical point along the edge of the arena (Figure 61). In contrast with model 1, model 2 is much faster, and 261 -hour simulations of shoals of 10 fish can be created in 2 minutes. Comparison of the models can be found in Figure 62.


Figure 61: How model 2 handles impossible trajectories. In this example, the velocity vector would place the fish outside of the arena (red star); instead, the fish goes to the symmetric of that position relative to the tangent to the edge.


Figure 62: Comparison of trajectories obtained with real fish (a), model 1 (b) and model 2 (c) for one fish observed over 20 minutes.

I computed for data obtained with the model the same 7 variables as with the experimental data. For each variable, I compared the values obtained with the real data (without distinction of conditions) and the artificial ones. All variables showed significant differences ( $\boldsymbol{p}<0.01$ see Figure 63 and Table 5) between real and artificial data. For some variables, like the median speed and the time spent still, this is a direct consequence of the way the artificial model was designed. Since the instantaneous velocities of the artificial fish are drawn from the distribution of all the instantaneous velocities of the real fish, the median speed of each artificial fish is the median of a set of 90000 numbers drawn randomly from the same distribution; as such, it is highly likely to be close to the median of the instantaneous speeds of the real fish, for all artificial fish. As a consequence, the simulated data show almost no variance between the $m s$ variables of the 260 artificial fish ( $s t d=61 \mu \mathrm{~m} / \mathrm{s}$ ). For the real fish on the contrary, the quicker fish will tend to have a median speed that is much higher than for the slower fish, and there is a much higher variance for median speed of the real fish $(s t d=8.5 \mathrm{~mm} / \mathrm{s})$. Similarly for the time spent still, the artificial fish tend to have the same proportion of instantaneous velocities that are below the threshold of immobility, so here again, we have an extremely low variance for the artificial fish and a significant difference with the real fish.

Compared with the simulation, real fish showed a greater tendency to follow the edge of the arena (Figure 63a; Table 5) and explored a smaller proportion of the arena (Figure 63d, e; Table 5). Although the DNN are comparable between the experimental data and the model ( 4.1 cm on average in both cases; the difference computed of 0.05 cm is below the resolution of the video and thus not meaningful. See Figure 54b), there is significant variation among the experimental data, compared to the simulation. Moreover, the experimental data show that fish tend to be part of larger groups than their simulated counterparts (see Figure 54c); this confirms that there are behavioural differences between the fish and the model. The real fish also showed greater variability in these measures than the model fish (Levene's test: $\boldsymbol{p}<0.001$ in all cases).

Thus, in all three experimental conditions, the fish differed from models that were programmed in a similar environment but without social behaviours. This shows that the experimental design did not preclude the expression of social behaviour, yet fish in all three conditions were found to behave in the same manner. We can thus conclude
that the similarity in behaviour is not a consequence of environmental constraints, but rather of the absence of effect of ocean acidification on the behaviour of juvenile sea bass.

Table 5: Results of model testing for differences in the behaviour of real and simulated fish for different behavioural variables. Presented are the estimates, associated standard error (s.e.), tvalue and p -value.

| Variable | Estimate | s.e. | t-value | p-value |
| :---: | :---: | :---: | :---: | :---: |
| Proportion of time near the arena edge |  |  |  |  |
| intercept | 0.513 | 0.013 |  |  |
| real versus model | -0.265 | 0.019 | -14.16 | $<0.001$ |
| Time taken to explore $10 \%$ of the arena |  |  |  |  |
| intercept | 0.113 | 0.004 |  |  |
| real versus model | -0.100 | 0.006 | -15.43 | <0.001 |
| Overall \% of the arena explored |  |  |  |  |
| intercept | 0.565 | 0.009 |  |  |
| real versus model | 0.340 | 0.013 | 26.86 | <0.001 |
| Proportion of local neighbours |  |  |  |  |
| intercept | 2.221 | 0.037 |  |  |
| real versus model | -0.447 | 0.053 | -8.450 | <0.001 |
| Distance to nearest neighbour |  |  |  |  |
| intercept | 4.132 | 0.053 |  |  |
| real versus model | -0.046 | 0.075 | -0.618 | <0.001 |



Figure 63: Differences in behaviour, reaction to the environment and interindividual interactions for sea bass (experiments) and simulated fish. Experimental data with real fish suggest that, compared to the model, they are more likely to be found close to the edge of the arena (a), explore less of the arena (e) and explore it more slowly (d), and have a higher number of neighbours nearby (c). Distances to the nearest neighbour are similar on average between real and simulated fish, although there is more variation in the experimental data. Statistical comparisons are presented in Table 5.

### 3.6. Discussion

In this study, I investigated the effect of near-future oceanic conditions upon the behaviour of juvenile European sea bass. The fish showed statistically similar swimming patterns, and presented the same kinds of interaction with their environment
and with each other, regardless of whether they were raised and tested in ambient conditions, OA conditions, or raised under ambient conditions and tested in OA water. Overall, this indicates behavioural resilience to OA.

Most of the research on the effect of high temperature and acidity on European sea bass so far has been focused on extreme conditions. Grøttum et al. [559] for instance found that a concentration of $137.2 \mathrm{mg} . \mathrm{L}^{-1} \mathrm{CO}_{2}$ would be lethal to sea bass, but this level of acidity is much higher than the most pessimistic models of OA, and does not reflect any reality to be expected on Earth. Similarly, Cecchini et al. [560] submitted European sea bass to progressively higher concentration of $\mathrm{CO}_{2}$ and observed diminished feed intake for extreme acidification ( $\mathrm{pH}=6.48$ ). But in conditions matching $2100 \mathrm{OA}(\mathrm{pH}=7.52)$, sea bass showed increased feed intake. In a study by Vinagre et al. [557], sea bass showed thermal stress above $24^{\circ} \mathrm{C}$. This temperature was chosen to reflect the temperature expected for estuarine nurseries along the Portuguese coast in the summer post climate change. This choice is relevant, as juvenile sea bass usually hatch in estuaries [547]. The Portuguese coast is however among the warmest places within sea bass habitat, so these results might not be applicable to estuaries further north along the coast of Europe. The temperature tested here, $19^{\circ} \mathrm{C}$, corresponds to an average of Atlantic Ocean surface temperature under a $+2^{\circ} \mathrm{C}$ warming. Experiments using similar definitions of high and low temperatures and acidity as we did here found no effect of OA on larval morphology development [563] and no change in gene expression [561]. Malavasi et al. [558] observed that shoal cohesiveness was actually highest around $22^{\circ} \mathrm{C}$.

Nevertheless, the results presented here describe the behaviour of sea bass in a very simplified environment, and thus should not be overgeneralised. The fish tested here were all juveniles, because it is around that age that shoaling behaviour emerges in European sea bass [553], so the result may be generalised to adults; however, the size of the shoals observed here ( 10 fish) is extremely small compared to the shoals that occur in the wild [546]. The fish were recorded in a 2-D setting, and it could be useful to test whether these results could be reproduced in a 3-D environment. Partridge et al. [73] describe an experimental setting in which shoal structure can be studied in 3 dimensions. The results described here highlight the edge following behaviour in juvenile sea bass. Gathering around physical barriers is thought to play a role in the
emergence of shoaling behaviour in sea bass [553], so observing juveniles in a more complex environment could shed some light on the learning process for shoaling.

Ocean acidification, especially coupled with global warming, is a complex phenomenon to understand, as it affects many different aspects of the environment. This study offers no insight regarding the impact that the consequences of OA on the ecosystem might have on sea bass. Global warming will have an impact on oceanic currents [485], which could certainly have an impact on shoal cohesion. At all ages, crustaceans represent an important part of sea bass' diet [546]; although crustaceans are thought to be more resilient to OA than most calcifying organisms, OA has been found to make their shell stronger [511]. The consequences of this on sea bass foraging behaviour is currently unknown. OA can also affect fish on a neurological level, resulting in maladaptive behaviour regarding predators [523], [528], [531], [579]. This possible impact of OA has yet to be studied in sea bass.

Despite the simplifying hypotheses regarding the environment, this study returns important results regarding sea bass shoaling behaviour. Not only does it prove that sea bass behave in a similar manner in this experimental setting regardless of the pH and temperature conditions, but through the use of a computer model, I have shown that sea bass exhibiting no social behaviour would behave in a very different way. In particular, fish in the model exhibited less shoaling behaviour. More specifically, by showing significant behavioural differences between the real and model fish, I proved that juvenile sea bass do exhibit inter-individual interactions within this simple setting, and that these interactions are not impacted by ocean acidification.

### 3.7. Conclusion

Collective behaviour results from the combination of individual characteristics of the members of the group, the way they interact, and the properties of their environment. In this chapter, I analysed the effect a modification of the environment properties would have on individual and group behaviour. Ocean acidification and global warming are slow processes that impact many aspects of the environment. There are complex interactions between their diverse consequences, which makes accurate predictions
difficult. The year-2100 ocean does not exist anywhere at the moment, yet we need to determine its consequences as early as possible. This requires the use of models.

Models can only be based on our understanding of the real environment. Experimenters might neglect to include in their model aspects of the environment they do not expect to play a role in the behaviour studied. Thus, designing a useful model of an environment that does not exist yet is difficult. One possible approach is to test the organisms in a wide variety of settings, even beyond the range described by the most advanced predictions. This approach could return contradictory results though. Moreover, many dimensions would need to be tested, resulting in an immense number of tests. Another approach for designing future model would be to model the entire ecosystem. But to be accurate, the evolution of the system through time should be accounted for: the most accurate map is the territory, and the only way to obtain a perfectly reliable model of year- 2100 oceans would be to wait until 2100 .

The time required to adapt to environmental change is an important variable however. For instance, because they have generation times of about 1 day, unicellular algae and bacteria will go through tens of thousands of generation as the $\mathrm{pCO}_{2}$ slowly increases; this is thought to be sufficient for adaptive processes to become relevant regarding environmental change occurring over decadal timescales or longer [580]. This is not true for vertebrates like sea bass. For these, it is the individual development that should be taken into account. For this reason, this study compared specimens observed in the same conditions they were reared in, as well as some that were observed in OA conditions after being reared in ambient water. An alternative method is to make the transition to the new condition progressive [514], [548].

An individual behaviour can be impacted by its environment and by the way it interacts with its conspecifics. Organisms can be studied in a different environment, but they cannot be observed outside of any environment. The experimental environment can be made as simple as possible so as to minimise the influence of the environment on the individuals' behaviour, but the possibility of an influence of the environment cannot be removed altogether. As a result, it is not possible to isolate the influence of interindividual interaction in the resulting behaviour. The influence of the environment can be isolated by observing the organisms in isolation. But in order to determine whether
there are interactions with conspecifics when the organism is observed in a group, it is not enough to compare the organism's behaviour as part of a group to its behaviour when on its own. Indeed, although all the variables that describe individual behaviour (speed, etc.) might be identical, it is not possible to compare group variables (interindividual distances, etc.) between the two conditions, as the group variables cannot be defined in the lone organism condition. So it is not possible to isolate the role of the environment in the resulting behaviour, as there is no control for what the absence of collective behaviour would be.

My approach in this chapter was to develop a behavioural model based on the data collected on the organisms in groups, hypothesizing that there were no interactions with conspecifics: the agent's behaviour in the model is entirely defined by its interaction with the environment (typically, it is impossible to leave the arena). Thus, the variables that describe the individual behaviour of the agents are identical to those describing the individual behaviour of real fish, but the variables that describe the behaviour of the shoal are completely random (within the range authorised by the geometry of the experimental setting), as they would be if the fish were moving independently, with no collective behaviour. Differences between the global behaviour of the fish and that of the model's agents prove the existence of a collective behaviour in juvenile sea bass even in such a simple and artificial environment.

Resorting to a simple environment made it possible to predict the possible ways the organisms may interact with the environment and simulate this behaviour in a computer model. Even in the most advanced model (model 1), the rules describing the interactions of the agent with its environment were entirely defined by four parameters. This would not have been possible in a more complex environment. For this reason, this study highlights what can be gained by associating observation in the wild (formulation of the hypothesis and defining the parameters of the experiment, such as the age of the fish), lab experiment (data collection) and computer simulation (confirmation/refutation of the hypothesis).

# CHAPTER 4 - NO I IN TEAM? IMPACT OF INDIVIDUAL CHARACTERISTICS ON HUMAN GROUP PERFORMANCE IN A FORAGING TASK 

### 4.1. Introduction

Individuals from many different animal species that live in groups coordinate their behaviours in time and space [177], [200], [581], [582]. The collective patterns observed at the scale of the group can be explained as an emergent property of local interactions between individuals [200]. Such phenomena can be represented by computer models in which all individuals in the group (often called "agents") are created with similar properties, interacting with each other and their environment according to precise sets of rules [198], [200], [583]. These models of collective motion have led to new insights into the mechanisms for the behaviour of various species, over a variety of sizes, e.g. from cells [584]-[587], through to fish [140], [588], birds [589], [590], ungulates [591]-[594], and recently to large marine mammals [595], [596].

Humans are a particular example of species that forms self-organising groups, and application of such methods to humans has proven fruitful in understanding the psychology of crowd behaviours [597], city planning [598], transport engineering [599] and virtual environments [600]. Data collected from direct observation [601], video recordings [602] or GPS [603], in direct observation of experiments, have informed the development of a suite of different models. Most of those models either consider the crowd as a gas [604] or as a group of particles subjected to so-called social forces [605], or consider its individual members as agents with specific behaviours [606]. These approaches have yielded important insights regarding how building evacuations should be conducted for example [607], and, in particular, in the management of 'bottlenecks' in crowds, to avoiding potentially catastrophic panic movements [208]. Another example of application is in human-robot teamwork scenarios: an agent can use a model
of human behaviour to predict the next actions of the people its interacting with, and plan its own next action accordingly [608].

In general, theoretical (agent-based) models of animal (including human) collective behaviours have all agents following the same local interaction rules [193], [201], [609], [610], but more recent models do allow for variations in the agents and their general patterns of behaviour [195], [611], [612]. One of the ways agents can be differentiated is through their personality [613], and models have found that such differences (e.g. a greater or lesser tendency to be social) can determine the outcome of collective behaviours [614]. This phenomenon is supported by much empirical research. In sheep (ovis aries) herds for instance, the proportion of bold individuals increases the likelihood of the herd splitting into subgroups to graze on different patches [615]. In three-spined sticklebacks (gasterosteus aculeatus), bold individuals resume foraging after an attack quicker than shyer individuals, and display lower shoaling tendencies [616]. A variety of individual-specific traits such as age, size, sex and dominance can also play a role in determining the position of an individual within a group in diverse species [617]. Furthermore, the geometric structure of the group may be impacted by the presence of individuals with specific leadership traits independent or correlated to those other traits mentioned above [618]. Those leaders may emerge from a difference of interest from the others [619] or from access to different information [620], [621], or from different traits, such as personality [616], [622], or traits associated with dominance [623], such as status within the group [155], [624], size [625] or sex [626].

Different models have been developed to represent personality differences in human, such as Eysenck's 3-factor model [614], the Myers-Briggs Type Indicator (MBTI) [627] or the Five-factor model (also known as the OCEAN model, for Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism, which are said five factors) [608], [628]. They can mediate patterns of interaction and collective behaviour. In a collective resources dilemma [629], for example, extraversion is negatively correlated with cooperation, whereas agreeableness is positively correlated with cooperation. Gender can also impact on individuals' strategies in a public goods game [630] and perceptions of team norms change more among demographically diverse groups (in terms of sex, race and nationality) [631]. Height is known to have an influence on leadership, with taller people being more likely to hold managerial
positions [632], and Lindeman et al. [633]showed that, based on their pictures alone, taller women were regarded as better managers than shorter ones. Blaker et al. [634] found the same results regarding the links between height and perceived leadership for both genders, and postulated that this was a consequence of the advantage that height provided in ancestral human environments.

Obtaining precise data on the individual characteristics of members of a crowd in the contexts in which they tend to be studied (e.g. building evacuations) is difficult. Another situation associated with human collective motion has, however, gathered more precise data regarding the individual characteristics and behaviour of the agents within the crowd: foraging. Indeed, an important idea in the study of foraging in general and human foraging in particular is the Optimal Foraging Theory (OFT) [635]. The idea of OFT is that through Darwinian evolution, the foraging strategy that minimises the cost (time, effort and risk) and optimises the gain in calories will be the strategy that will tend to be followed by the population. Hence the strategy of the group should converge toward the strategy that best suits the individual strengths of its members [636]. The behavioural data can be obtained from direct observation of the path taken, time measurement of the foraging [636], observing subjects' foraging in a virtual environment [637], or equipping the subjects with GPS units [638]. Several of such studies [637]-[639] have found that the exploration patterns of the foragers could be represented by standard mathematical models which are compatible with OFT [639]. Studies of human foraging have also found that foraging strategies varied with age [636], gender [640] and group size [641] and teams of people in social foraging games, teams of people quickly achieve coordination to maximise their foraging returns, relying on communication to achieve consensus on the best foraging patches [642].

In this chapter, the behaviour of teams of different numbers of subjects is observed during a social foraging task. I specially investigate if and how inter-individual differences in age, sex, size and personality types impact individual foraging strategies and group-level outcomes.

In the Wisconsin Card Sorting Test (WCST), a test of the ability to adapt one's strategy in the face of changing feedback [643], [644], agreeableness has been found to be negatively related with maintaining a losing strategy, and both a high agreeableness and
a high conscientiousness were found to be related with good scores at this test [645]. Since the WCST present similarities with the foraging experiment (subjects are asked to find the best strategy in an initially unknown environment that is subject to change), I expected people with a high level of agreeableness to make correct decisions regarding their choices of patch, and similarly, groups composed of a higher proportion of team members scoring high on these traits to do better.

Since height [632], [633], agreeableness [646], and extraversion [647] have been associated with leadership: I expected subjects to follow those who match this profile.

### 4.2. Methods

### 4.2.1. Study subjects and foraging task

This study is a follow-up to the work presented in [642], where groups of 4, 8, 12 and 24 human subjects were asked to collect tokens from any of six "foraging patches". Subjects were recruited among visitors to the Royal Veterinary College Open Day on 7th May 2011, 111 females, 33 males, age 16 to 54.

Subjects were also submitted to a short questionnaire to determine their personality according to the five-factor model (FFM) [648]. The FFM describes personality according to five orthogonal dimensions: openness, conscientiousness, extraversion, agreeability and neuroticism (OCEAN) [649]. The FFM is one of the most prevalent models of personality [650]-[653], along with the Myers-Brigg Type Indicator (MBTI) [654]-[656]; correspondences have been found between the two frameworks [657], [658], showing them to be both accurate assessments of personality. Ahrndt et al. [608] provide a comparison between the features of both models, and conclude that the FFM is better suited for models of crowds. Subjects' personalities along the FFM are often evaluated through the use of the Revised Neuroticism, Extraversion, Openness Personality Inventory (NEO-PI-R) [659]-[665], a 240 questions long questionnaire. Due to its length, the NEO-PI-R was considered inappropriate for this experiment. Therefore, the questionnaire chosen for the experiment is the Big-Five Inventory 10 (BFI-10) [666], a questionnaire based on the NEO-PI-R that includes only 10 questions.

Each trait is measured by two questions (on a scale from 1 to 5 ), with one of the questions measuring the presence of the trait directly, and the other measuring it in reverse (a high score on those questions means that the subject is low on that trait; see Table 6). This test thus measures subjects' personality on a scale between 2 and 10 for each of the five dimensions: openness, conscientiousness, extraversion, agreeableness and neuroticism.

Table 6: The BFI-10 questionnaire. Subjects are asked how well the following statements describe their personality. The scores were not included in the questionnaire. The letter R in the score column indicates that the item should be reverse-scored.

| I see myself as <br> someone who ... | Disagree <br> strongly | Disagree <br> a little | Neither <br> agree <br> nor <br> disagree | Agree a <br> little | Agree <br> strongly | Score |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| $\ldots$ is reserved | 1 | 2 | 3 | 4 | 5 | Extraversion, R |
| ...is generally <br> trusting | 1 | 2 | 3 | 4 | 5 | Agreeableness |
| $\ldots$..tends to be lazy | 1 | 2 | 3 | 4 | 5 | Conscientiousness, R |
| ...is relaxed, <br> handles stress well | 1 | 2 | 3 | 4 | 5 | Neuroticism, R |
| ...has few artistic <br> interests | 1 | 2 | 3 | 4 | 5 | Openness, R |
| ...is outgoing, <br> sociable | 1 | 2 | 3 | 4 | 5 | Extraversion |

The experiment consisted of a foraging game in which subjects collected tokens from different patches, and gathered them at a home base. The environment consisted of four circular arenas, with a 10 metre diameter. At the centre of the arena was a bin, the "home base", and six other bins, or "patches", were regularly located on the circle (each being 5 m away from the nearest two, see Figure 64a). There were three trials; every
subject took part in one trial, in one arena. In each trial, the first arena was occupied by 4 people, the second by 8 , the third by 12 and the last one by 24 subjects.

In each of the ten minutes trials, each patch initially contained 300 tokens, both good (green) and bad (purple) in various proportions: in all conditions, one patch had $95 \%$ of good tokens, two had $65 \%$, one had $50 \%$, one had $35 \%$ and the last one $5 \%$ of good tokens; but the order of the bins differed between arenas. Subjects were tasked to collect one token per trip only and get it back to the home base; they were instructed to collect as many green tokens and as few purple tokens as possible (subjects could not see the tokens before collecting them, but once they had picked one up and seen its colour, they had to take it back to the home base regardless of the colour). No communication was allowed among subjects, who were initially naïve to the quality of the patches.

The experiment was run simultaneously in all four environments with three repeats (the distribution of tokens in all patches being reinitialised before every run). All participants were provided with baseball caps that were outfitted with data loggers which collected geolocation data at 1 Hz [111] (see Figure 64b). These data were used to explore the decisions and performance of subjects as described below.


Figure 64: a: Representation of the foraging arena. B represents the home base, p1-6 represent the six patches; the patches were located at a distance $\mathrm{d}_{\mathrm{B} \rightarrow \mathrm{p}}=5 \mathrm{~m}$ from the home base. The shaded portion of a patch represents the proportion of good tokens it contains: $5 \%, 35 \%, 50 \%, 65 \%$ or $95 \%$. Although the same six ratios of good and bad forage were found in every arena, they were not necessarily in the same position. Image taken from [642]. b: Photograph from one of the trials: the 24 subject arena can be seen in the foreground, and another arena can be seen in the background. The data loggers are visible on top of the subjects' caps. One of the patches can be seen clearly in the bottom right corner: subjects were not able to see the tokens before drawing them.

### 4.2.2. Subject movement data

The data loggers recorded the coordinates of the caps throughout all three experiments. So for each logger, the data corresponding to each experiment was sampled, with the starting time of the sample being the same for all subjects within an arena and a trial. Out of 144 subjects, 13 had to be removed due to faulty loggers (the signal was lost during the experiment). Among the remaining loggers, the recording errors were then removed by discarding any data points that were farther from 7 metres from the home base (this includes any time when the logger correctly detected that subjects were outside of the arena). Indeed, in the conditions with the most participants, the number of people already present at a patch at any given time could prevent the subject from directly going there; they would instead have to walk around the other subjects, to reach the patch from behind: they would then be between 6 m and 7 m from the home base. But this strategy never required going further than 2 metres behind the patch.

Data were also cleaned by eliminating speeds that were deemed unlikely. Speeds were calculated with the following formula:

$$
\begin{equation*}
s(t)=\sqrt{\sum_{k=1}^{2}\left(X_{k}(t+\Delta t)-X_{k}(t)\right)^{2}} / \Delta t \tag{Equation 23}
\end{equation*}
$$

$X_{1}(t)$ and $X_{2}(t)$ being the abscissae and ordinates and time $t$ respectively. Subjects were instructed to keep a walking speed. Literature indicates that walking speed in a crowd is usually between $1.2 \mathrm{~m} / \mathrm{s}$ and $1.5 \mathrm{~m} / \mathrm{s}$ [667]-[669], with maximum walking speed being $2.53 \mathrm{~m} / \mathrm{s}$ on average for men in their 20 s (and lower for other populations) [670]. Hence, all data associated with speeds above $3.00 \mathrm{~m} / \mathrm{s}$ were also removed. For more than $80 \%$ of subjects, the data eliminated represented less than $10 \%$ of the duration of the experiment. The distribution of speeds is presented in Figure 65.


Figure 65: Distribution of speeds in the corrected data set. Walking speeds are distributed normally around $1 \mathrm{~m} / \mathrm{s}$, as represented by the normal distribution in red (mean: $1.05 \mathrm{~m} / \mathrm{s}$, standard deviation: $0.28 \mathrm{~m} / \mathrm{s}$. The mean squared distance is $2.61 \times 10^{-6}$ ). The other behaviours, such as picking up a token, form a second peak near $0 \mathrm{~m} / \mathrm{s}$.

A foraging trip by a subject was defined to be a closed-loop trajectory that starts and finishes at the home base and visits one foraging patch. Under ideal conditions, the trajectory would look like that shown in Figure 66a: the extremities of all foraging trips are close to the home base on one side, and close to a patch on the other. However, the coordinates recorded by the loggers sometimes happened to exhibit spatial drift, with $12 \%$ of trips showing drift, as demonstrated in Figure 66b: the parallel tracks show that most of the trips are to patch $\mathrm{n}^{\circ} 3$, yet many among them appear to be closer of patch 4 . As a consequence, it was not possible to automatically determine to which patch subjects were walking.


Figure 66: Two examples of trajectory; every colour a patch, a darker shade represents a more important drift. a: an example of a subject who visited all the patches, with almost no drift. $\mathbf{b}$ : a subject with important drift; even though some of the foraging trips are detected very close to patch 4, they were associated to patch 3 because their direction is mostly parallel to a line between the home base and patch 3, indicating that the dark red tracks should be closer to the red ones.

I observed every trajectory and annotated them with the patches visited by the subjects at all times. The beginning and end of every trip was then detected by considering the projection of the trajectory onto the arm the subject is walking on. A trip starts at a local minimum (the home base). It then goes through a local maximum that is located at least 3.5 m away from that minimum (as subjects don't necessarily have to be on the patch or the home base to pick up / drop a token, I found that the local extrema were not necessarily 5 m apart). Finally, the projection of the subjects' trajectory goes back to a local minimum. If this loop is completed in less than 17 s , it is considered to be a trip. Indeed, $0.6 \mathrm{~m} / \mathrm{s}$ is a slow walking speed [670], so walking the 10 m required to collect a token and drop it at the home base at normal speed shouldn't require more than $\frac{10 \mathrm{~m}}{0.6 \mathrm{~m} / \mathrm{s}} \approx 17 \mathrm{~s}$. The effort is defined as the number of trips divided by the proportion of recorded data. For subjects whose loggers recorded noisy data, few trips were detected, the recorded data being very sparse. For this reason, those subjects' effort is very low: 4 subjects have an effort value below 10. In the majority of cases though, the effort is a reasonable estimation of the total number of trips made by the subject throughout the experiment, with a median value of 41.88 trips made in 10 minutes.

After each trial, every team was given a score, equal to the number of good tokens collected, but it was not possible to directly measure individual scores. I considered
instead that on average, each subject collected on each trip a number of good tokens equal to the proportion of good tokens in the patch. For instance, when a subject visited a patch of medium quality (i.e., containing $50 \%$ green and $50 \%$ purple tokens), their score was incremented by 0.5 , the average value of a token from that patch, regardless of whether the subject collected a purple or a green token. In order to take into account the missing data, I defined the accuracy as the ratio of the score by the number of trips. It represents the average quality of the choice of patches.

Since I was interested in the decision process of the subjects, an important variable (switch $N b$ ) was the number of times a subject would choose to go to a different patch from the one previously visited. Finally, in order to understand whether subjects drew information from their environment, I collected data on the location of the other subjects. medDNN is the median over time of the distance to the nearest neighbour $D N N_{i}$, which is defined as

$$
\begin{equation*}
\operatorname{DNN}_{i}(t)=\min _{j} \sqrt{\sum_{k=1}^{2}\left(X_{k, i}(t)-X_{k, j}(t)\right)^{2}} \quad \forall i, j i \neq j \tag{Equation 24}
\end{equation*}
$$

with subject $i$ and $j$ being respectively at the coordinates ( $X_{1, i}, X_{2, i}$ ) and ( $X_{1, j}, X_{2, j}$ ) at time $t$. This is the only variable that relies on accurate measurements by the GPS loggers. Figure 67 b presents the distribution of the accurate data among participants, i.e. the number of time points for which data were recorded at the actual coordinates of the subject. I consider that data are accurate for a trip if the location of the subject when they are at the patch matches the actual coordinates of the patch (i.e. no GPS drift). For comparison, Figure 67a represents the proportion of usable data: the data that were not necessarily recorded at the proper coordinates, but do not impact the detection of trips for instance.


Figure 67: a: Cumulative distribution of usable data; $\mathbf{b}$ : Cumulative distribution of accurate data. Data is considered accurate for the times that do not include drift.

Another way of measuring group dynamics was to count the number of people going to the same patch as the subject. All subjects who were walking toward the patch or coming back from it were included in the count (see Figure 68).


Figure 68: Position of all subjects at one instant. The rectangles represent the patches, with the grey area representing the proportion of 'good' tokens. B represents the home base, at coordinates $(0 ; 0)$. The * represent the subjects, their colour represent the patch they are going to or coming back from, so all subjects in the same colour are foraging in the same patch.

A summary of the variables is presented in Figure 69.


Figure 69: Distribution of the variables that describe subjects' behaviour: the number of trips recorded for the subjects (a) and the total number of trips as estimated from the usable data (or 'effort', b), the score as estimated from the recorded trips (c), the estimated proportion of successful trips ('accuracy', d), the median distance to the nearest neighbour (DNN, e), the median proportion of the group visiting the same patch (f), the number of times a subject switches to a different patch (g). The proportion of usable data is shown in $\mathbf{h}$.

### 4.2.3. Data analysis

The performance of the subjects was measured by their accuracy. I ran a linear mixed model (LMM) to explain this variable as a function of the individual properties of the subjects: sex, age, height, and the 5 factors that describe their personalities (model A). To account for the non-independence of data coming from the same arena, or from the difference in group sizes, group ID as well as the number of subjects in the arena were included as random variables in all models.

I then observed the influence of individual behaviour on the accuracy. The behaviours selected as fixed effects for this model were the effort, the number of time the subject switched to a different patch (switchNb), the median number of subjects visiting the same patch (medNbr), and the median inter-individual distance (medDNN) (model B).

I tested the influence of the individual properties of the subjects (age, height, sex, personality dimensions as fixed effects) on the factors from model B that were found to have an impact on the performance, in order to check for any indirect influence of individual properties on performance (model C). I tested in particular whether agreeableness had an impact on the individual strategy (represented by switchNb). I also tested whether the number of subjects going to a patch was correlated with the instantaneous walking speed of the subjects, to control whether crowd effects were impacting subjects' locomotion, and whether conscientiousness would impact the choice of strategy. Finally, I planned to analyse whether the composition of the group along the variables highlighted by model A had an impact on the performance of the group (in other word, whether the influence of factors like personality could be a group effect, or only impacted subject behaviour and performance at the individual level. Model D). A summary of the different models can be found in Table 7.

Table 7: The different analyses planned. The response variable in model C depends on the results of model B , although the effect of agreeableness and conscientiousness on the strategy will need to be tested, even if SwitchNb is found to have no effect on the accuracy in model B. The details of model D will depend on the results of model A .

| Model | Response <br> variable | Fixed effect | Random effect |
| :--- | :--- | :--- | :--- |
| Model A | Accuracy | sex, age, height, Big 5 | Group ID, group size |
| Model B | Accuracy | Effort, switchNb, medNbr, medDNN | Group ID, group size |
| Model C | [model B] | sex, age, height, Big 5 | Group ID, group size |
| Model C - <br> strategy | SwitchNb | Agreeableness, conscientiousness | Group ID, group size |
| Model D | Accuracy | Group personality | Group ID, group size |

All statistical analyses were conducted in the R environment [576] and LMMs fitted using the package lmertest (Tests in Linear Mixed Effects Models) [671]. The level of significance was set at $\boldsymbol{\alpha}<0.05$ and model diagnostics were performed using graphical procedures (Q-Q plot and standardized residuals versus fitted values; see Figure 70).


Figure 70: Graphical control of the validity of the model for one of the principal components: the Q-Q plot is close to being linear (a), and the residuals are normally distributed (b). Similar distributions were observed for the other models.

### 4.3. Results

Since the goal of model A is to determine the influence of each of the individual characteristics on the score, they should not be aggregated into principal components. Table 8 shows that among the individual characteristics, only height and sex are correlated (Spearman coefficient: $0.51, \boldsymbol{p}<0.001$ )

Table 8: Spearman correlation coefficients $\rho$ between the individual characteristics (the letters are the 5 personality dimensions). Only height and sex present a moderate correlation ( $\rho>0.4$ ).

|  | N | A | E | C | O | Height | Age |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Sex | -0.21 | -0.01 | -0.10 | -0.01 | 0.01 | $\mathbf{0 . 5 1}$ | 0.06 |
| Age | -0.09 | -0.02 | -0.13 | 0.13 | -0.12 | 0.08 |  |
| Height | -0.28 | -0.02 | 0.07 | 0.01 | 0.03 |  |  |
| O | -0.13 | 0.19 | 0.06 | 0.21 |  |  |  |
| C | -0.22 | 0.19 | 0.26 |  |  |  |  |
| E | -0.28 | 0.10 |  |  |  |  |  |
| A | -0.22 |  |  |  |  |  |  |

In model A, the intercept corresponds to the estimated accuracy of a 0 year old woman 0 m tall with 0 in all five personality dimensions, regardless of the trial or the number of people in the group (of course, the model has no validity for such numbers). The estimate for each dimension corresponds to the slope associated with the dimension. None of the variables tested in model A were found to predict the score $(\boldsymbol{p}>0.05$, see Table 9).

Table 9: Results of model testing for the influence of individual characteristics on the subject's score. Presented are the estimates and associated standard error (s.e.), t -value and p -value.

| Predictor variable | Estimate $\pm$ s.e. | t -value | p -value |
| :--- | :--- | :--- | :--- |
| Intercept | $0.80 \pm 0.33$ |  |  |
| Sex (M) | $-0.034 \pm 0.039$ | -0.86 | 0.392 |
| Age (year) | $-0.0016 \pm 0.0012$ | -1.35 | 0.178 |
| Height (metres) | $-.00068 \pm 0.018$ | -0.38 | 0.702 |
| Openness | $0.00028 \pm 0.0074$ | 0.038 | 0.969 |
| Conscientiousness | $0.0070 \pm 0.010$ | 0.67 | 0.502 |
| Extraversion | $-0.014 \pm 0.0085$ | -1.7 | 0.098 |
| Agreeability | $0.00031 \pm 0.0077$ | 0.040 | 0.967 |
| Neuroticism | $-0.014 \pm 0.0071$ | -1.9 | 0.055 |

There is a correlation between the number of neighbours and the distance to the nearest neighbour (Spearman's rank correlation coefficient: -0.60, $\boldsymbol{p}<0.001$ ): if more subjects
are visiting the same patch, the density along the path between the patch and the home base will be higher, hence subjects are more likely to be close to one another. This correlation means that the number of neighbours and the distance to the nearest neighbour cannot be used in the same linear model. For this reason, model B was split into two models, model B1 (medNbr) and model B2 (medDNN). The effort was positively correlated to a better accuracy ( $\boldsymbol{p}<0.005$, in both models): the more patches subjects visited, the better their score. There was no relation between the number of switches between patches and the accuracy in either model ( $\boldsymbol{p}>0.05$ ). The number of neighbours was positively correlated to the accuracy: subjects who visited the same patches as a large part of the group collected more tokens ( $\boldsymbol{p}<0.001$ ). Similarly, the distance to the nearest neighbour was negatively correlated to the accuracy ( $\boldsymbol{p}<0.001$ ). The detailed results of these tests can be found in Table 10.

Table 10: Results of model testing for the influence of the behaviour of the subjects on their score. Presented are the estimates, associated s.e., $t$-value and p-value.

| Model | Predictor <br> variable | Estimate $\pm$ s.e. | t -value | p -value |
| :---: | :--- | :--- | :--- | :--- |
| Model B1 | Intercept | $0.28 \pm 0.07$ |  |  |
|  | effort | $0.0039 \pm 0.0011$ | 3.478 | $<0.001$ |
|  | medNbr | $0.053 \pm 0.010$ | 5.231 | $<0.001$ |
|  | switchNb | $-0.00024 \pm 0.0021$ | -0.116 | 0.91 |
|  | Intercept | $0.57 \pm 0.08$ |  |  |
|  | effort | $0.0035 \pm 0.0011$ | 3.071 | 0.003 |
|  | medDNN | $-0.094 \pm 0.022$ | -4.252 | $<0.001$ |
|  | switchNb | $-0.0020 \pm 0.0021$ | -0.939 | 0.35 |

As explained previously, the goal of the different versions of model C was to test the influence of individual characteristics on the variables that had been found to affect the score in model B (effort, medNbr and medDNN), as well as the influence of conscientiousness and agreeableness in particular on the strategy (switchNb). None of the individual properties were found to affect either the effort or the number of
neighbours $(\boldsymbol{p}>0.05)$. Moreover, neither conscientiousness nor agreeableness affected the number of switches ( $\boldsymbol{p}>0.05$ for both). However, the level of extraversion of the subject was found to be associated with greater distances to the nearest neighbour ( $\boldsymbol{p}<0.05$, see Table 11).

The goal of model D was to determine whether the effect of the individual characteristics on accuracy was due to the emergence of particular collective behaviours in the groups in which such characteristics were more present. As no individual property was found to affect the performance in model A, I did not test their impact at the level of the group.

Finally, I found a correlation between the number of subjects going to a patch and the instantaneous walking speed of the subjects. (Spearman's rank correlation coefficient: $0.34)$, which indicates that subjects tended to walk more slowly when walking along arms that contained more people.

Table 11: Results of model testing for the influence of individual characteristics of the subject on their effort, and on the median number of neighbours. Presented are the estimates, associated s.e., t -value and p -value.

| Response variable | Predictor variable | Estimate $\pm$ s.e. | t-value | p-value |
| :---: | :---: | :---: | :---: | :---: |
| effort | Intercept | $34.4 \pm 14.2$ |  |  |
|  | Sex (M) | $1.17 \pm 2.45$ | 0.477 | 0.63 |
|  | Age | $-0.060 \pm 0.089$ | -0.672 | 0.50 |
|  | Height (metres) | $2.88 \pm 5.60$ | 0.514 | 0.61 |
|  | Openness | $0.61 \pm 0.54$ | -1.122 | 0.26 |
|  | Conscientiousness | $0.75 \pm 0.76$ | 0.978 | 0.33 |
|  | Extraversion | $0.29 \pm 0.62$ | 0.469 | 0.64 |
|  | Agreeability | $-0.031 \pm 0.58$ | -0.054 | 0.96 |
|  | Neuroticism | $0.25 \pm 0.52$ | 0.472 | 0.64 |
| medNbr | Intercept | $1.60 \pm 1.57$ |  |  |
|  | Sex (M) | $-0.013 \pm 0.25$ | -0.051 | 0.96 |
|  | Age | $-0.010 \pm 0.009$ | -1.072 | 0.28 |
|  | Height (metres) | $0.205 \pm 0.58$ | 0.353 | 0.73 |
|  | Openness | $-0.0068 \pm 0.057$ | -0.118 | 0.91 |
|  | Conscientiousness | $0.063 \pm 0.081$ | 0.780 | 0.44 |
|  | Extraversion | $-0.053 \pm 0.065$ | -0.808 | 0.42 |
|  | Agreeability | $0.038 \pm 0.060$ | 0.632 | 0.53 |
|  | Neuroticism | $-0.068 \pm 0.054$ | -1.246 | 0.22 |
| medDNN | Intercept | $2.01 \pm 0.79$ |  |  |
|  | Sex (M) | $0.058 \pm 0.112$ | 0.521 | 0.60 |
|  | Age | $0.027 \pm 0.004$ | 0.669 | 0.50 |
|  | Height (metres) | $0.060 \pm 0.255$ | 0.237 | 0.81 |
|  | Openness | $-0.023 \pm 0.025$ | -0.901 | 0.37 |
|  | Conscientiousness | $-0.036 \pm 0.036$ | -1.000 | 0.32 |
|  | Extraversion | $0.064 \pm 0.028$ | 2.244 | 0.03 |
|  | Agreeability | $-0.022 \pm 0.026$ | -0.852 | 0.40 |
|  | Neuroticism | $-0.026 \pm 0.024$ | -1.095 | 0.27 |
| switchNb | Intercept | $0.502 \pm 0.099$ |  |  |
|  | Conscientiousness | $0.0023 \pm 0.0092$ | 0.251 | 0.80 |
|  | Agreeableness | $0.0023 \pm 0.0073$ | 0.320 | 0.75 |

### 4.4. Discussion

This research project looked into how individual characteristics would impact the strategy, the organisation and the efficiency of human teams in a foraging activity.

Previous analysis conducted with this data set found that all teams quickly located and focused their efforts on the best or two best patches [642]. My analysis confirmed this result, the number of subjects visiting one patch being a good predictor of their accuracy. This result is coherent with Offerman and Schotter's results [672], who found that when presented with a decision with uncertain payoffs, subjects tend to imitate successful subjects. The influence of the group on individual behaviour can be observed in different domains, such as the cultural market. Salganik et al. [673] presented an experiment in which subjects had to rate songs, and were given the ability to download them. In one condition ("independent"), the choice of songs was made fully independently from the other subjects, whereas in another condition ("social influence"), subjects could see the number of time any song had been downloaded. Although the best-rated songs tended to be downloaded more in every condition, the study found that the distribution of proportion of downloads was significantly less balanced (higher Gini coefficient) in the social influence condition than in the independent condition. In other words, when subjects could observe each other's behaviour, they tended to converge on the best-rated option rather than explore.

Analysing subjects' strategy proved difficult. One cause of the difficulty was that the trajectories were not always simple back and forth trips in straight line between the home base and one of the patches as I expected. In large groups in particular, the density of the crowd should be considered, as subjects can rarely go in a straight line: since subjects tend to find the best patches quickly and to focus on those, the line between the home base and the best patch tends to be crowded. Even more so at the home base itself: since many subjects tend to be around the home base at all times, subjects often had to move around it in order to deposit their tokens. Subjects tend to move more slowly amidst crowds [605], and in this experiment, we can observe that subjects walked more slowly when travelling along the same arm as many other subjects. This is probably the reason why in the 24 -subjects trials, subjects tended to split themselves between the best two patches, rather than all going to the patch with the highest proportion of good tokens: going through the crowd represents an added cost.

Furthermore, it was sometimes difficult to infer decisions on the part of the subjects from their trajectories alone, and thus the beginning of a trip could rarely be determined with a precision inferior to a few seconds. For these reasons, it was not possible to
determine who the leaders in the group were, and I was not able to assess the influence of agreeableness, extraversion and height on leadership in this task. If two subjects decide to start visiting the same patch within a few seconds of one another, this does not necessarily mean that the second subject is following the first: the opposite is actually possible (if the first subject was able to guess the destination of the second one from visual cues earlier than I could determine it from GPS data). Another possibility would be a phenomenon of social amplification. Amé et al. [674] have shown that cockroaches gather according to a non-linear process, and that once the group reaches a certain size, the social signal represented by this aggregation becomes amplified, and draws larger numbers of cockroaches. Similar dynamics have been observed in humans [675]. Hence, when two or more subjects decide to move at the same time to a patch that is already being foraged, it is possible that neither is following the other, but that they are actually both responding to the presence of the subjects already present at that patch. Testing this hypothesis would require conducting this experiment with a larger number of subjects, or with better means of determining their strategy, for instance by being able to observe where the subjects are looking just before they start moving to a new patch [676].

In a study on the impact of personality on cooperation, Koole et al. [629] found that agreeableness was positively related with cooperation, but extraversion was not. Here, extraversion was found to have an effect on the distance to the nearest neighbour, but not on the number of neighbours. This difference is difficult to explain, and would necessitate further research. One possible explanation is that extraverted people keep larger distances than other people when in smaller groups, in which the distance to the nearest neighbour is not necessarily equivalent to the density of the group.

My analysis did not show any influence of agreeableness on the task or the performance. One difference between the two protocols was that in Koole et al.'s experiment, subjects were made aware of the limited state of the resources [629]. In King et al.'s design on the other hand [642], subjects were not made aware that the amount of resources per patch was limited, and in fact, the smaller groups did not deplete any patch. Moreover, the subjects were tasked to gather tokens as a group: at the end of every trip, the token was placed in a common pool at the home base, highlighting the collaborative aspect of the task. Koole et al. on the other hand gave feedback to the
subjects after every turn, informing them on whether the common pool was being overused, thus highlighting the tragedy of the commons in this scenario [677]. In other words, in the experiment described in this chapter, a greedy approach is encouraged: not only the subjects were invited to collect as many tokens as possible throughout the trial (although they were not allowed to take more than one per trip), they were also invited to gather the "good" tokens in priority. Greedy strategies have disastrous effects in tragedy of the commons scenarios [678], which instead require cooperation, which is emphasized by the agreeableness dimension [679]. Hence, it is possible that the effect of agreeableness is task-dependent.

An important aspect of the task was the necessity to adapt one's strategy as the tokens successively collected confirm or infirm one's assessment of the quality of the patch. According to Jensen-Campbell et al. [645], this aptitude too is associated with a high agreeableness. In this case too, context might be the reason for the difference with the results observed here, since Jensen-Campbell et al.'s results related to an individual task. Since agreeableness is associated with cooperation, results in an individual task might not be applicable to a collective task.

### 4.5. A human Multi-Armed Bandit problem

In probability theory, there is a class of problem called "multi-armed bandit problems" (MAB) [680]. If a slot machine $i$ (or one-armed bandit) yields an unknown expected value $\mu_{i}$ with every use, what is the optimum strategy for maximising returns from a system of $N$ slot machines, each with a different expected value? An optimal strategy for this problem would have to solve the trade-off between "exploitation" (the use of the machine whose return is known to be decent) and "exploration" (trying different machines enough times to evaluate their returns compared to the machines that have already been tested).

Besides gambling strategies, this problem finds applications in various domains. In clinical trials, different drugs are initially thought to have some likelihood of effectively treating a disease. Exploitation means continuing with the already vetted drugs, with its known side-effects and chances of failure, and exploration means risking patients' lives
with a new drug. Determining the best drug requires tests, but those tests can be costly in terms of human lives and patients' quality of life. Other fields of application include web site optimisation [681], portfolio design [682], or personalised recommendation [683]. MAB have also been applied to the study of animal behaviour, as individuals [684] or as a group [685]. The study presented in this chapter can be considered as a sixarmed bandit problem.

A simple algorithm for a MAB problem is the epsilon-greedy approach. A simple greedy approach would go for the arm $\alpha$ with highest estimated probability and ignore the other arms, even if the first few tries were not representative of the actual quality of the arm, and those arms would actually return better results than the arm $\alpha$. An epsilongreedy algorithm avoids this difficulty by choosing an arm at random a fraction $\varepsilon$ of the time.

Another popular algorithm is the Upper Confidence Bound algorithm (UCB) [686]. It relies on the principle of optimism under uncertainty: the idea is to go for the arm that could be the best one, based on the current knowledge of the system. Either it is actually better than the others, or its high upper confidence bound means that its confidence interval is too large compared to the others, in which case ignoring it would be a risky strategy.

One of the first algorithms presented to solve the exploration-exploitation dilemma was Thompson sampling [687]. The probability distribution associated with arm $i$ is supposed to be a Beta distribution, which is defined as $f_{\alpha_{i}, \beta_{i}}$ :

$$
\begin{gather*}
f_{\alpha_{i}, \beta_{i}}(\theta)=\frac{\theta^{\alpha_{i}-1}(1-\theta)^{\beta_{i}-1}}{B\left(\alpha_{i}, \beta_{i}\right)}  \tag{Equation 25}\\
B\left(\alpha_{i}, \beta_{i}\right)=\int_{0}^{1} t^{\alpha_{i}-1}(1-t)^{\beta_{i}-1} d t \tag{Equation 26}
\end{gather*}
$$

( $B$ is the beta function, and $\alpha_{i}$ and $\beta_{i}$ are the shape parameters of the distribution associated with arm $i$ ). If no information is known about the system before trying any arm, $\alpha_{i}$ and $\beta_{i}$ can be considered to be 1 for all arms, in which case the probability distribution associated with each arm is uniform. Otherwise, different initial probability distributions can be used to fit the priors. Each turn, we choose the arm $i$ that maximises the expected reward (considering the current estimation of the distributions), and we
receive a reward $r \in\{0 ; 1\}$. The probability distribution for that arm is then modified by increasing $\alpha_{i}$ if $r$ is 1 , or $\beta_{i}$ if $r$ is 0 , by 1 . Thompson sampling has been shown to be close to optimal [688]. Chapelle and Li [689] have found Thompson sampling to return a lower regret than UCB, which means it tends to be closer to what could be considered as the optimal strategy a posteriori.

Variants of the MAB problem include contextual bandits, which consider contextual elements for computing the priors [683], non-stationary or restless bandits [690], in which the rewards can change randomly, adversarial bandits [691], in which the reward for the different arms are decided by an adversary rather than constant, etc.

Modelling the foraging simulation as a MAB problem would allow for comparison of different strategies, and in particular to compare with the experimental results a model with or without information gathering from the other agents. Several models have attempted to combine multi-agent systems with MAB. Liu and Zhao [692] devised a system in which multiple agents are collectively increasing their score by playing with a MAB system that has fewer arms than agents. Agents get penalised if they choose the same arm at the same time, either by not getting a point, or by getting one point only for their collective effort, to represent the collision between players. This can be compared to how in the foraging experiment; subjects were walking more slowly when sharing an arm with more people. Liu and Zhao's model could be a way to represent the cost of sharing an arm without having to model the actual collisions and path planning.

Schlag [693] designed a multi-player MAB in which the different agents can observe the behaviour of the other agents to guide their decision. Contrary to the foraging experiment though, they can also observe the result of said behaviour (i.e. whether their score is increased or not). The agents in Schlag's model can only observe a fraction of the other agents. In the foraging task, it is unlikely that a subject would be able to pay attention to all 23 other subjects, so they might have proceeded in a similar fashion. In the condition of symmetric sampling, Schlag's agents always observed some of the same subset of agents, whereas in the random sampling condition, the choice of agents to be observed is completely random. As I was not able to determine the copying strategy of the subjects in the foraging experiment, it is not possible to say whether human subjects tend to follow randomly, or if they tend to follow the same subset of
subjects (the leaders). Schlag found that the best strategy was that of proportional imitation. This strategy consists in: only switching as a means to imitate other agents (no exploration of arms that are not being visited by others); never imitating agents that perform worse than oneself; decide to imitate any given agent with a probability that is proportional to how much better that agent performed. This strategy is clearly dependent on being able to observe the result of the other agents.

In contrast, Rendell et al. [694] implemented a 100-armed bandit, in which agents could either discover a new behaviour ("innovate"), perform a known behaviour ("exploit"), or observe another agent performing a behaviour known to them ("observe"). Only "exploit" could yield a payoff, but exploitation was only possible for behaviours that had already been discovered through observation or innovation. The most effective strategies were found to be those that relied heavily on observation. The reason for this is that only the agents that are "exploiting" can be observed, and the exploitation phase usually came once an agent had become confident they knew a behaviour with a high payoff. Hence, copying would generally grant an agent some information about a small sample of behaviour with a high payoff, whereas innovating would return information about a randomly chosen unknown behaviour. One difference between innovating and observing was that innovating would return accurate information regarding the payoff of the new behaviour, whereas observing had a small probability of associating an incorrect payoff to the newly observed behaviour. Yet the authors found that this made little difference: even when the probability of error was high, the agents who started to exploit the newly acquired behaviour tended to fare better than those who did not. This is coherent with what we observed in the foraging experiment, where the payoff of the patch visited by the other agents was unknown, yet the fact that a large proportion of agents were going to a patch made it likely to return a high payoff. Although the most successful strategies relied on copying, the authors found that the best among them were selective about when they copied. Every turn, the payoff associated with a behaviour had a low probability of changing; the best strategies started copying again shortly after the payoff of the behaviour they were exploiting changed. In the foraging experiment, the payoff from a patch could change under one circumstance: when it is depleted, the payoff would drop to 0 . In those circumstances, most of the subjects who were visiting that patch would switch to the next one with the most subjects.

An implementation of the foraging experiment as a MAB would not need to model the geometry of the problem. Every turn, agents would decide on a patch to forage in. The duration of a turn would be dependent on the number of agents visiting the patch at the same time. The data logs from the experiment did not show the subjects stopping to observe their surroundings, so observations should not require an entire turn as in [694]. Since the experimental data were not enough to determine the subjects' strategies in terms of observations, different strategies should be compared: observing the same agents throughout the experiment [693], observing patches and the number of agents currently visiting them, observing the other agents only when the payoff of the patch changes [694], etc. Outside of exterior information, the process through which agents update their priors following a visit to a patch should be coherent with humans' intuitive understanding of probabilities, rather than a precise computation of a beta distribution. Reverdy et al. [695] have reported that human subjects performed at best as well as the UCB algorithm. An important difference between that study and the foraging experiment is that the distribution of the arms was not random (see Figure 71). This was meant to represent familiarity with the environment and prior knowledge; it did not bias the results in favour of the humans however, as the version of UCB employed by the computer relied on correlated priors. Since the distribution of the patches in the foraging experiment was completely random, it is unclear whether human performance would be comparable to that of the UCB algorithm in these circumstances.


Figure 71: Two of the "landscapes" employed by [695]. The different arms were presented to the human subjects as a $10 \times 10$ empty grid. The $z$ axis represents the payoff of the square at the corresponding $x, y$ coordinates in the grid. Both landscapes present clear regularities that were exploited both by the human subjects and the computer algorithm.

In contrast, when Racey et al. [696] subjected humans and pigeons to a MAB problem with randomly distributed payoffs, they found that the humans' behaviour was consistent with a softmax decision rule [697] (whereas the pigeons performed poorly in this task, selecting arms independently of the rewards). Thus, the softmax decision process might be the most adapted algorithm for modelling human behaviour in such a task.

Comparing such simulated models with the behaviour of the subjects in the foraging experiment could yield new insight into the strategies they followed.

### 4.6. Conclusion

All animals need to feed, so foraging is almost universal. Yet foraging needs not be a social activity: in many species, one individual has to locate food, acquire it and consume it on its own [698]-[700]. In contrast, humans cooperate on a larger scale than most animals, their cooperation not being limited to relatives [701]. In modern Western civilisation, food production and distribution is part of a highly complex network of interactions which would not be possible without cooperation. But even in huntergatherer societies, foraging is a social activity [641].

In this chapter, I presented a foraging experiment, in which I studied how human subjects performed as a team. The experiment was conducted in a circular arena of 10 metres in diameter. In actual foraging circumstances, humans usually need to travel for much longer distances [639]. However, although nothing was done to encourage collaboration (in particular, subjects were asked not to talk to each other), this setting was enough to observe the emergence of collective behaviour (see [642]). One of the most complex examples of coordination between subjects was the formation of a human chain between the best patch and the home base. This particular event happened twice, in two trials that were conducted simultaneously, which suggest that this is a case of inter-group imitation (these events were not included in the analysis, as no movement can be observed on the recording, and this part of the experiment thus had to be discarded).

Collective behaviour emerges as the consequence of many individual interactions. One important aspect of the study was the analysis of individual behaviour in terms of personal characteristics, such as age, gender or personality. In particular, like other species [618], [622], [702], humans tend to follow leaders [703], [704], and one goal of the study was to determine whether being a leader was related to personal characteristics. Although extraversion was found to have an influence on individual behaviour, I was not able to determine the identities of leaders, let alone their characteristics. This highlights the importance of the choice of observation method for the study of collective behaviour: although some behaviours can be inferred from the trajectory of a subject, one GPS logger was found not to be enough to determine where (and on whom) subjects focused their attention or to understand their decision process. From a change in trajectory, it is possible to determine that the subject has decided to switch to a different patch (which allowed me to determine the beginnings and ends of foraging trips). But the data did not always make it possible to pinpoint the moment of the decision down to the second, and thus to determine what caused the change in strategy and infer the decision process.

Finally, I explored how a simulated model of this experiment could be implemented. Interestingly, this experiment matches the structure of both foraging experiments [637] and Multi-Armed Bandit (MAB) tasks [696]. A simulated model would only focus on simpler behaviours: decision processes would be instantaneous, agents would walk in straight lines, and of course, a simulated agent would not decide to form a chain or to wander into other patch fields, as it has happened in one of the trials. Nevertheless, it could provide a useful tool to understand the mechanisms of cooperation and the emergence of collective behaviour.

## CHAPTER 5 - SYNOPSIS; THE FUTURE

Collective behaviour is fundamental to the survival of many species. It predates life itself, as even complex chemical structures can occur through self-organisation by macromolecules [705]. It is a powerful process through which individual organisms can structure themselves as a complex system and maximise their fitness. Although the local rules followed by the individuals within the group are enough to explain the complex structures that can emerge from them, isolating those rules is not easy. One reason is that different local rules can produce similar group outcomes [212], [706]-[708]. This means that in order to understand collective motion, it is necessary to understand the behaviour of both the individual and the group.

In this thesis, I have demonstrated that aspects of the collective behaviour could be deduced from the study of a smaller group. Local and global properties can be measured with the help of diverse tools, depending on the nature of the group, the time and space scale, and the aspect of the behaviour to be studied. I also showed how computer models can complete our understanding of the links between the local and the global scale. I followed this approach for groups from 4 to tens of thousands, observed at scales from under 1 mm to 10 m , and time scales ranging from 10 minutes to several days (for frequencies ranging from 0.5 mHz to 25 Hz ).

In Chapter 2, this method was applied to mesenchymal stem cells: cells from the same lineage were observed with two different instruments. Fluorescence and bright field microscopy gave me information on cells spatial organisation, their dynamics, and their division process. Flow cytometry on the other hand returned information over large quantities of cells. If those methods could be combined, it would be possible to generalise the properties observed from microscopy to the larger populations that were observed through flow cytometry. However, I was able to prove through the use of models that the cells developed differently between the two conditions. This project illustrated the danger of observations perturbing the system, and that even though different models can be drawn from the same data, useful information can be deduced from the features they share.

In Chapter 3, I studied European sea bass under ambient conditions as well as the conditions that can be expected by the end of the century because of ocean acidification. Because this environment does not exist in the wild yet, this experiment could only be done in laboratory conditions, resorting to a smaller group. Using a computer model, I demonstrated that the individual interactions observed in those simplified conditions could be generalised to observations in the wild. This study illustrated the advantages and limitations of extremely simplified experimental environments.

Finally, in Chapter 4, humans were the species under study. I analysed the behaviour of groups of different sizes as they collaborated in a foraging task, and confirmed that the size of the group had an impact on the strategy. This experiment revealed that individual personality had no influence on the outcome of the task. It also highlighted the importance of the precise measurements in order to understand the rules followed by the subject.

Each of these studies generated large amounts of data. Historically, the difficulty of collecting data over large scales from considerable numbers of individuals with high precision has been a limit to the study of collective behaviour [64], [65], sometimes leading researchers to erroneous conclusions [66]. Nowadays, not only is it possible to collect large amounts of data at a high frequency [134], but computers can also be used to eliminate noise at high speed and to find patterns in the data [709], [710]. As measurement tools become more accurate, it will become possible to set up experiments that are closer to the natural conditions, and to apply to the wild observations methods that had to be confined to the lab in the past [178].

Yet detailed and accurate data sets will not be enough to understand collective behaviour, as the local rules followed by the organisms are not directly accessible to the researcher. It is possible by studying the relative movement of individuals within the group to disprove a possible rule of interaction [590], but not to prove that the rule is sufficient for explaining the emerging structure observed. Simulations are required for that. Simulations make it possible to quickly test various hypotheses regarding the link between the local and the global scale. Moreover, they make it possible to set up the system in such a way that all parameters are known, whereas even in laboratory conditions, the organism must be considered as a black box. Models are also a method to confirm that the global behaviour can indeed occur from the hypothesised individual behaviours alone [201].

The future will bring more opportunities to associate observations in the wild, lab experiments and simulated models. Bio-logging can already provide us with precise data on which to base computer models [106]. But developing the right model is complex when the influence of the environment is unknown. Laboratory experiments are still required in order to control for the multiple parameters that could impact the organism's behaviour, such as the environment [90], the composition of the group [148], or the behaviour of the conspecifics [156]. Thus observations from the wild (possibly completed by numerical data from biologging) would lead to the formation of hypotheses, which could then be tested in the lab on a reduced number of organisms under controlled conditions and in a simplified environment. The data produced in those experiments could be simply expanded to full-size artificial group, confirming or infirming the hypothesis.

It would even be possible to close the loop by introducing those rules inside the group: a robot could be programmed with those rules and placed within the group, so as to control how close its behaviour is from that of the rest of the group [157], [711], [712]. This approach makes it possible to test rules modelled after one species in a completely different group, as suggested by Strömbom and King [713]. This new paradigm would make it possible to study behaviour outside of its evolutionary context, thus isolating the purpose of the rules.

One particularly interesting species to study with such approaches is human beings, the only species that does not need the intervention of biologists to frequently walk around with a GPS sensor. By their use of smartphones, humans constantly generate large amounts of data [714] which can be harvested and analysed in order to better understand their behaviour. The aggregation of the data generated by the entire group of users makes it possible to understand the individual behaviour [715], and reciprocally, the individual actions can be used to predict emerging patterns at the group level [716]. But collective behaviour need not be collective motion only: humans generate far more than GPS data, and those other types of data also can be harvested and processed to better understand and influence individual and group behaviour [717]. For instance, virtual environments can be modified in a way that is barely noticeable by the user, so experimenters can study the influence of those environmental changes on inter-subjects interactions, as well as the group patterns that can emerge from them [718] (an example of a method called $\mathrm{A} / \mathrm{B}$ testing [719]). And the data generated can be used to program bots to infiltrate human virtual societies in order to influence their behaviour [720].

The use of large amounts of data to better understand, predict, and influence individual and collective behaviour will soon become more prominent in fields like user experience [721], marketing [722] and politics [723], as it will be in biology. Understanding the conclusions from such research might be the best way to become more aware of such influence and to protect ourselves from it.

APPENDIX

## ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research
CrossMark
click for updates
Cite this article: Duteil M, Pope EC, Pérez-Escudero A, de Polavieja GG, Fürtbauer I, Brown MR, King AJ. 2016 European sea bass show behavioural resilience to near-future ocean acidification. R. Soc. open sci. 3: 160656. http://dx.doi.org/10.1098/rsos. 160656

Received: 1 September 2016
Accepted: 6 October 2016

Subject Category:
Biology (whole organism)

## Subject Areas:

behaviour/environmental science

## Keywords:

Dicentrarchus labrax, environmental change, fisheries, motion tracking, social behaviour

## Author for correspondence:

A. J. King
e-mail: a.j.king@swansea.ac.uk

# European sea bass show behavioural resilience to near-future ocean acidification 

M. Duteil ${ }^{1,2}$, E. C. Pope ${ }^{1}$, A. Pérez-Escudero ${ }^{3,4}$, G. G. de<br>Polavieja ${ }^{3,5}$, I. Fürtbauer', M. R. Brown² and A. J. King ${ }^{1}$<br>${ }^{1}$ Department of Biosciences, College of Science, Swansea University, Singleton Park SA2 8PP, UK<br>${ }^{2}$ College of Engineering, Swansea University, Bay Campus, Swansea SA1 8EN, UK<br>${ }^{3}$ Instituto Cajal, Consejo Superior de Investigaciones Científicas, Madrid, Spain<br>${ }^{4}$ Department of Physics, Physics of Living Systems Group, Massachusetts Institute of Technology, Cambridge, MA, USA<br>${ }^{5}$ Champalimaud Neuroscience Programme, Champalimaud Center for the Unknown, Lisbon, Portugal

(DD) $\mathrm{AJK}, 0000-0002-6870-9767$

Ocean acidification (OA)—caused by rising concentrations of carbon dioxide $\left(\mathrm{CO}_{2}\right)$-is thought to be a major threat to marine ecosystems and has been shown to induce behavioural alterations in fish. Here we show behavioural resilience to nearfuture OA in a commercially important and migratory marine finfish, the Sea bass (Dicentrarchus labrax). Sea bass were raised from eggs at $19^{\circ} \mathrm{C}$ in ambient or near-future OA ( $1000 \mu \mathrm{~atm}$ $p \mathrm{CO}_{2}$ ) conditions and $n=270$ fish were observed 59-68 days post-hatch using automated tracking from video. Fish reared under ambient conditions, OA conditions, and fish reared in ambient conditions but tested in OA water showed statistically similar movement patterns, and reacted to their environment and interacted with each other in comparable ways. Thus our findings indicate behavioural resilience to near-future OA in juvenile sea bass. Moreover, simulated agent-based models indicate that our analysis methods are sensitive to subtle changes in fish behaviour. It is now important to determine whether the absences of any differences persist under more ecologically relevant circumstances and in contexts which have a more direct bearing on individual fitness.

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9. figshare.c. 3519006.

## 1. Background

Rising concentrations of carbon dioxide $\left(\mathrm{CO}_{2}\right)$ in the atmosphere over the past 200 years have led to ocean acidification (OA) that threatens marine ecosystems [1-3]. It is predicted that by the year 2100 up to half of the seawater at the ocean's surface will show partial pressures of $\mathrm{CO}_{2}$ exceeding 1000 microatmospheres ( $\mu \mathrm{atm}$ ), with detrimental implications for fisheries [4]. Consequently, there is a collective drive in the scientific community to better understand how future OA may impact upon marine fisheries and ecosystems [5].

Generally, the first response of organisms to any environmental change is to modify their behaviour which, in turn, shapes ecological effects [6]. However, it is challenging to incorporate behaviour into climate change studies, and this is likely limiting our ability to forecast the impacts of change [7-9]. Given the ecological and economic ramifications of positive or negative effects on commercially important marine fisheries, it is critical to investigate the behavioural effects of near-future oceanic conditions, using robust and relevant experiments to inform management strategies [6,7]. Recent work studying a variety of marine fish species has found both negative or neutral effects of nearfuture OA conditions on fish behaviour [10-15]. Such contradictory results may, at least in part, be explained by the different ways experiments are designed and conducted. For instance, some researchers expose fish to OA conditions and observe fish behaviour in ambient conditions (e.g. [10,16]) while others expose fish to OA conditions and also observe fish behaviour in OA water (e.g. [11]). Overall, this makes it difficult to disentangle OA effects from any behavioural alterations that are a response to any changed environment.

Here, we investigate the behaviour of juvenile European sea bass, Dicentrarchus labrax, a commercially important fish species with global landings and total aquaculture production worth over \$US 1 billion in 2012 (8990 and 153182 tonnes from capture fisheries and aquaculture, respectively [17]). We assessed fish behaviour under three different conditions: (i) raised and tested under ambient conditions, (ii) raised and tested under near-future OA conditions, and (iii) raised under ambient conditions but tested under near-future OA conditions. We used automated video tracking [18] and calculated behavioural measures that are commonly used as measures of inter-individual differences in behaviour within populations $[19,20]$ and are likely to be of biological and ecological relevance for obligate schooling migratory fish species like sea bass [21,22]. Significant differences in fish behaviour across our experimental conditions would provide evidence for behavioural alterations in the sea bass in response to OA [10-13,16,23]. By contrast, if fish showed statistically similar movements and interacted with their environment and each other in comparable ways, this would indicate behavioural resilience to OA. Given that the laboratory environment could constrain fish behaviour (e.g. the restriction of movement to the test arena), we also built an artificial agent-based model of the experiment, which we compared with our experimental data, allowing us to gauge whether our analysis method is sensitive to subtle changes in fish behaviour.

## 2. Material and methods

### 2.1. Subjects

Fertilized D. labrax eggs from a mixed spawn (multiple males and females) were incubated in water at $19^{\circ} \mathrm{C}$ and $585 \mu \mathrm{~atm} p \mathrm{CO}_{2}$ (ambient) or $1000 \mu \mathrm{~atm} p \mathrm{CO}_{2}$ (for full details of rearing conditions, see the electronic supplementary material and Pope et al. [24]); $n=180$ fish from $585 \mu \mathrm{~atm} p \mathrm{CO}_{2}$ conditions (mean length $\pm$ s.d. $=32.34 \pm 3.12 \mathrm{~mm}$ ), and $n=90$ fish (mean length $\pm$ s.d. $=35.35 \pm 4.61 \mathrm{~mm}$ ) from 1000 uatm $p \mathrm{CO}_{2}$ conditions were collected from their home tanks and put in groups of $n=10$ using a sweep net (one movement) and studied over a 10-day period from day 59 to day 68 of development, which coincides with metamorphosis into the early juvenile stage when schooling behaviour begins to develop. Following completion of experiments, fish were sacrificed by an overdose of anaesthetic (MS222; Acros Organics) and total length and weight were measured.

### 2.2. Experimental set-up

Groups of $n=10$ fish were placed in one of three circular Plexiglas 'test arenas' ( 30 cm diameter, 3 cm height) with permeable mesh bottoms (figure 1a). The test arenas were placed in a larger tank with a flowthrough water system to ensure conditions were maintained. Test arenas were separated with opaque plastic partitions and surrounded by a custom-built aluminium frame and white screen (PhotoSEL BK13CW White Screen). A Panasonic HDC-SD60 HD video camera (Panasonic Corporation of North


Figure 1. Data collection and processing. (a) Fish were filmed in circular test arenas placed in a larger tank with flow-through water system (see 'Experimental set-up' in Material and methods for details). (b) Frequency histogram of fish speeds ( $\mathrm{cm} \mathrm{s}^{-1}$ ) across all experiments when detected as moving; inset shows the proportion of total observations in which fish are detected as still or moving. (c) Example of a single fish's movement trajectory over a 60 min trial, with intensity of pixels indicating the amount of time the fish centroid spent on the pixel. (d) Data showing mean (red line) plus standard deviation (blue area) in exploration of the test arena (pixels explored) as a function of time for $n=260$ fish over 26 one-hour trials. (e) A single frame screenshot showing the camera view of $n=10$ fish from video collected at 25 Hz , with a focal fish highlighted by a red cross. Conspecifics within the median inter-fish distance are within the red line ( $n=2$ ), and the focal fish's nearest neighbour indicated by a blue cross. ( $f$ ) Frequency histogram showing inter-fish distances for $n=260$ fish over 26 one-hour trials; the modal inter-fish distance is indicated by the red vertical line.

America, Secaucus, NJ, USA) was positioned above each test arena (figure 1a). Four photographer's lights (each with $4 \times 25$ W 240 V 6400 K True Day light bulbs) lit the arenas from the outside, hence dispersing light evenly over the arenas and enabling optimum conditions for video recording. Video recordings were started approximately 60 s after the fish were placed into the arena and arenas were recorded for 60 min of footage ( 90000 frames at 25 Hz ; representing 90000 data points per fish) which was subsequently tracked to provide $x, y$ coordinates (below). We tested groups of fish reared at near-future OA conditions ('OA fish', $n=9$ groups, $n=90$ fish), ambient ocean conditions ('ambient fish', $n=9$ groups, $n=90$ fish) and fish reared under ambient conditions but tested in OA water ('ambient fish tested in OA', $n=8$ groups, $n=80$ fish, because of a processing error of one video file).

### 2.3. Generating $x, y$ fish coordinates

Experiments were recorded as AVCHD videos, and converted into mp4 using Freemake (v. 4.1.5.4) to allow tracking of fish position using idTracker [18]. idTracker software extracts a characteristic fingerprint from each individual being filmed, and then uses these fingerprints to identify every individual throughout the video. For each video, we restricted the analysis to the arena, and selected the minimum size and intensity threshold in a way that we detected the 10 fish and no artefacts in ten frames randomly chosen in the video. After tracking, idTracker returns the $x, y$ coordinates for each individual fish at temporal resolution of 25 Hz (i.e. the frame rate). The limited water depth ( 3 cm ; figure $1 a$ ) ensures accuracy in our tracking in two dimensions and calculation of subsequent behavioural variables (see below), but also restricted fish movement in the vertical plane.

### 2.4. Processing and interpolation of $x, y$ fish coordinates

Some errors were detected in the $x, y$ coordinates generated by idTracker and so we undertook the following data-processing stages in Matlab (v. R2014b): (i) scaling: we normalized the images so that all the data we processed were at a scale of 0.03 cm per pixel, with the centre at $(0 ; 0)$ and diameter of 1000
pixels ( 29 cm ). This enabled us to delete any points that lay outside the arena which typically occurred when the fish swam near the edge of the arena and their reflection was detected by idTracker as another fish. (ii) Fish label swapping: occasionally a fish was confused with another fish or with an artefact by idTracker.

For each frame and for each tracked object (fish), we calculated the instantaneous speed $s_{t}$ ( $s_{t}=$ $\sqrt{\left(x_{t}-x_{t-1}\right)^{2}+\left(y_{t}-y_{t-1}\right)^{2}}$, with $x$ and $y$ being the coordinates of the fish and $t$ designating the current time frame). We compared this speed with a maximum possible speed, considered to be the fastest speed detected during manual tracking of videos, at $0.91 \mathrm{~m} \mathrm{~s}^{-1}$. This speed is coherent with other observations of maximum speed for juvenile sea bass at this temperature [25]. These two stages of post-processing eliminated $0.1 \%$ of data points that were considered as erroneous. Third, we interpolated the sections of missing points in our $x, y$ coordinates that were sufficiently short (see electronic supplementary material, figure S1, for an example, for processing for one fish in one experiment). For every sequence of missing data shorter than one second and for which we knew the position of the fish for the five frames just before and just after that sequence, we computed a piecewise cubic interpolation to determine the polynomial which best approximated the missing values. This last step allowed us to determine 104143 more data points, giving us a total of 2563633 frames with all fish, which represented $76 \%$ of the entire dataset ( 31144046 data points; 8937683 for the ambient condition, 8689475 for the OA condition, and 13516888 for the ambient tested in OA condition).

### 2.5. Behavioural variables

We extracted data on fish activity (median speed and proportion of time spent still), how fish reacted to the test arena (proportion of time near the edge of the arena, time taken to explore $10 \%$ of the arena, and overall percentage of the arena explored), and associations of fish with each other (median distance to nearest neighbour, and mean number of local neighbours) (see figure $1 b-e$ for an overview of these variables).

Fish speed at each frame $\left(s_{t}\right)$ was calculated as described above. For our analyses, we calculated the median $s_{t}$ across all frames and fish. We considered that a fish stayed still between two frames if its movement was less than a pixel, which amounts to a speed of $0.7 \mathrm{~cm} \mathrm{~s}^{-1}$, and used this to calculate the proportion of time spent still during experiments. The proportion of time fish spent near the arena edge was based on the proportion of time fish were within 3 cm (length of fish to nearest cm ) of the arena edge ( $20 \%$ of the radius of the arena). To determine how much of the arena fish explored within 10 min , and over the whole experiment, we partitioned each test arena into a grid of $1 \times 1 \mathrm{~cm}$ squares (total number of squares per arena $=788$; electronic supplementary material, figure S1); $100 \%$ exploration meant that the fish had entered each of the 788 squares at least once. To describe how fish interacted with each other, we calculated the distance of each fish to all other fish in every frame, and also computed the distance to the closest neighbour. Using these data, we determined the median distance between all fish for each trial. Over the entire dataset median distance was 6.9 cm and, for each fish in each frame, we then calculated the number of fish within 6.9 cm of a focal fish. For analyses, we used the mean proportion of the group that was within a focal fish's radius (less than 6.9 cm ).

### 2.6. Simulation model

Because the simple environment in which the fish were tested could have constrained fish behaviour in some way (e.g. due to the geometry of the environment), we built an artificial model of our experiment. Like in the real experiment, we placed $n=10$ fish in an arena equivalent to that used in our experiments (figure 1) and determined the position of every fish in consecutive time intervals (frames) as: $x(t+1)=$ $x(t)+v(t) \times \cos (\theta(t)+\omega(t))$ and $y(t+1)=y(t)+v(t) \times \sin (\theta(t)+\omega(t))$, where $x$ and $y$ are the coordinates of the fish, $\theta$ is the current orientation of the fish, and $v$ and $\omega$ are, respectively, speeds and angular speeds uniformly drawn at random from the distributions of speeds and angular speeds observed for the real fish. This effectively builds the simplest model of fish movement. We ran this model for 60 min , for the same number of replicates $(n=26)$ as in our experiment to produce a dataset of equivalent size to our experimental data and computed the same variables we measured with the real fish, except fish median speed and time spent still, which were directly coded in our model and, therefore, on average, identical (electronic supplementary material, figure S2). Because any speed and angular speed can be selected (from those observed in real fish), it was possible that the computed trajectory of the simulated fish fell outside the test arena. In this case, these coordinates were replaced with the symmetrical point along the edge of the arena.

### 2.7. Data analysis

Because our seven fish variables (see above) were correlated, we normalized variables and combined them into principal component scores with orthogonal rotation (varimax) using principal components (PCs) analysis [26]. To test for differences in the behaviour of fish across our three experimental conditions, we fitted three linear mixed models (LMMs) with PC1 (model 1), PC2 (model 2) and PC3 (model 3) as the response variable, and experimental condition ('ambient', 'OA', 'ambient tested in OA') as a categorical fixed effect. We fitted experiment number as a random effect to control for the potential non-independence of data because we observed three experimental arenas at a time.

To test for differences in the behaviour of our real fish compared with our simulated fish, we ran five linear models, one for each of the variables that were not used to generate the simulation dataset: (i) proportion of time near the arena edge, (ii) time taken to explore $10 \%$ of the arena, (iii) overall \% of the arena explored, (iv) median distance to nearest neighbour, and (v) mean number of local neighbours, and fitted data type (real, simulation) as a fixed effect.

All statistical analyses were conducted in the R environment [27] and LMMs fitted using the package lme4 (linear modelling) [28]. The level of significance was set at $\alpha<0.05$ and model diagnostics were performed using graphical procedures ( $\mathrm{Q}-\mathrm{Q}$ plot and standardized residuals versus fitted values).

## 3. Results

Our seven correlated behavioural variables produced three uncorrelated PCs that represented $84.7 \%$ of the variance in our data (table 1). Fish that scored high on PC1 tended to (i) be close to other fish, (ii) have a high number of local neighbours, and (iii) explore less of the arena. Fish that scored high on PC2 tended to spend most of their time at the edge of the arena. Fish that scored high on PC3 tended to spend little time still, and explored a large proportion of the arena. On this basis we labelled PC1 'sociability', PC2 'boldness', and PC3 'activity'. All three PCs were statistically similar across the three treatments ( $p>0.05$; figure 2; table 2).

Comparing our experimental data with our simulated agent-based model, we found statistically significant differences for all variables tested ( $p<0.05$; table 3). Compared with our model fish, real fish showed a greater tendency to follow the edge of the arena (figure $3 a$; table 3), explored a smaller proportion of the arena (figure $3 b, c$; table 3 ) and showed a greater attraction to their neighbours, resulting in greater cohesion within groups (figure $3 d$; table 3 ) and shorter distances to nearby neighbours (figure $3 e$; table 3). The real fish also showed greater variability in these measures than the model fish (Levene's test: $p<0.001$ in all cases). An example movie of the real fish data and simulated fish data is provided as a movie in the electronic supplementary material.

## 4. Discussion

We investigated the effects of near-future oceanic conditions upon behaviour in a commercially important and migratory marine finfish, the European sea bass. Fish showed statistically similar movements, interaction with their environment, and interaction with each other, regardless of whether they were raised and tested in ambient conditions, OA conditions, or raised under ambient conditions and tested in OA water. Overall, this indicates behavioural resilience to OA, supporting work on individual subjects showing that increased $p \mathrm{CO}_{2}$ similarly has no significant effect on larval morphology in this species [24], and work on individual subjects (rather than groups) investigating swimming activity and/or kinematics in juvenile or larval fish; for example, studies of cobia (Rachycentron canadum) [29], mahi-mahi (Coryphaena hippurus) [30], Atlantic cod (Gadus morhua) [31] and herring (Clupea harengus L.) [32] do not find an effect of elevated $\mathrm{CO}_{2}$. For an obligate schooling migratory fish like sea bass, our finding is good news, and suggests that we can be optimistic about their ability to cope with anthropogenic changes to their environment in the near future.

Since the simple environment in which the fish were tested in the laboratory could somehow constrain behaviour, we also built an agent-based model of the experiment which we compared with our experimental data. Comparing the model and real data allowed us to gauge whether our analysis method is sensitive to subtle changes in fish behaviour. Our results strongly suggest that the juvenile sea bass are interacting with one another and the environment in biologically relevant ways, and that these behaviours result in very different behavioural patterns that emerge from a simple model of fish that move with the same speed and turning angles.


Figure 2. Behavioural resilience to near-future ocean acidification. Data variables extracted from tracked fish trajectories (figure 1c) were normalized and combined into principal components. Linear mixed models revealed that treatment type does not predict variation in (a) PC1 'sociability', (b) PC2 'boldness' or (c) PC3 'activity'. See table 1 for details of how behavioural variables load onto PC axes and table 2 for statistical comparisons.

Table 1. Factorial decomposition of the first three principal components (PCs). These three PCs represent $84.7 \%$ of the variance in the data. The most important variables for each component are in italics.

| variable | PC1 'sociability' | PC2 'boldness' | PC3 'activity' |
| :---: | :---: | :---: | :---: |
| median speed | 0.025 | -0.031 | 0.520 |
| proportion of time spent still | $-0.026$ | 0.001 | -0.709 |
| proportion of time near the arena edge | 0.064 | 0.984 | -0.021 |
| time taken to explore 10\% of the arena | 0.245 | 0.098 | -0.114 |
| overall \% of the arena explored | -0.604 | 0.134 | 0.361 |
| median distance to nearest neighbour | -0.271 | -0.049 | -0.137 |
| mean number of local neighbours | 0.704 | $-0.026$ | 0.254 |

Table 2. Results of model testing for differences in the behaviour of fish across our experimental conditions. Presented are the estimates, associated standard error (s.e.), $t$-value and $p$-value with respect to PC1 'sociability', PC2 'boldness' and PC3 'activity'.

| variable | estimate | s.e. | $t$-value | $p$-value |
| :---: | :---: | :---: | :---: | :---: |
| PC1 |  |  |  |  |
| intercept | -0.096 | 0.107 |  |  |
| ambient versus OA | 0.144 | 0.144 | 1.002 | 0.316 |
| ambient versus ambient $\rightarrow$ OA | 0.081 | 0.106 | 0.762 | 0.446 |
| PC2 |  |  |  |  |
| intercept | 0.583 | 0.047 |  |  |
| ambient versus OA | $-0.079$ | 0.063 | -1.254 | 0.210 |
| ambient versus ambient $\rightarrow$ OA | 0.040 | 0.045 | 0.882 | 0.377 |
| PC3 |  |  |  |  |
| intercept | -0.026 | 0.055 |  |  |
| ambient versus OA | 0.003 | 0.078 | 0.040 | 0.968 |
| ambient versus ambient $\rightarrow O A$ | -0.093 | 0.081 | -1.147 | 0.252 |

Rearing fish in a competitor- and predator-free industrial setting where fish are not resource limited may have restricted our ability to detect differences. In fact, the importance of risk-taking behaviours (PC2) and social tendencies (PC1) measured here will likely be exaggerated during migratory movements [33-35] or under predation pressure [20,36]. Therefore, it will be important to determine whether the absences of any differences that we find here persist in adult fish and under more ecologically relevant circumstances and in contexts which have a more direct bearing on individual fitness. Indeed, while


Figure 3. Differences in behaviour, reaction to the environment and interspecific interactions for sea bass (experiments) and simulated fish (agent-based model). Experimental data with real fish suggest that they (a) are more likely to be found close to the edge of the arena, (b) explore less of the arena, (c) explore more quickly, $(d)$ have higher number of neighbours nearby and ( $e$ ) are closer to other fish. Statistical comparisons are presented in table 3.

Table 3. Results of model testing for differences in the behaviour of real and simulated (model) fish for different behavioural variables. Presented are the estimates, associated standard error (s.e.), $t$-value and $p$-value.

| variable | estimate | s.e. | $t$-value | $p$-value |
| :---: | :---: | :---: | :---: | :---: |
| proportion of time near the arena edge |  |  |  |  |
| real versus simulated data | 0.123 | 0.018 | 6.894 | $<0.001$ |
| time taken to explore 10\% of the arena |  |  |  |  |
| real versus simulated data | 230.7 | 23.68 | 9.745 | $<0.001$ |
| overall \% of the arena explored |  |  |  |  |
| real versus simulated data | $-0.223$ | 0.013 | $-16.65$ | $<0.001$ |
| proportion of local neighbours |  |  |  |  |
| real versus simulated data | 0.090 | 0.007 | 13.56 | $<0.001$ |
| distance to nearest neighbour |  |  |  |  |
| real versus simulated data | $-0.258$ | 0.076 | $-3.380$ | $<0.001$ |

individual variation in behavioural responses to high $\mathrm{CO}_{2}$ clearly exists (e.g. [37]), it is unknown how quickly favourable genotypes enabling the expression of particular behaviours could be established in populations. Extending the sorts of experiments presented here to explore within- and betweenindividual variation in behaviour across time and contexts (and preferably across generations) would enable researchers to address such questions. Similarly, our experiments were specifically designed
to investigate the effects of OA conditions on sea bass behaviour as if they were born into and live under such conditions in the near future. Of course, exposing the same fish to an acute change in OA may well have resulted in a behavioural change for the variables we investigated, or in other behaviours. For example, temperate wrasses (Ctenolabrus rupestris) exposed to OA (but tested in control water) show reduced avoidance of predator odour but no change in behavioural lateralization or swimming activity compared to control fish [38], and juvenile Atlantic cod (Gadus morhua) avoid OA water irrespective of how long they experience elevated $\mathrm{CO}_{2}$ [31]. Exposing subjects to acute shifts in $p \mathrm{CO}_{2}$ provides interesting case studies which add to our understanding of behavioural responses to rapid environmental change, but their ecological relevance to real-world OA is not always clear.

Overall, while we urge caution in over-generalizing our findings, our experimental investigations of the effects of near-future oceanic conditions on juvenile sea bass behaviour represent the most comprehensive study of its kind to date. We have used sophisticated automated tracking from video to produce a large and robust dataset to compare fish under different experimental conditions. We hope that the methodology and approach we have adopted here will inspire the development of even more elaborated experiments and models to understand how future OA may impact upon individual and social behaviour and how behaviour may shape ecological effects.

Ethics. This work was approved by Department of Biosciences Ethics Committee, Swansea University.
Data accessibility. Data are available from the Dryad Digital Repository http:/ /dx.doi.org/10.5061/dryad.2dc8k [39].
Authors' contributions. A.J.K. and E.C.P. conceived and performed the experiments and collected the video data. A.P.-E., G.G.d.P. and M.D. tracked fish trajectories from video. M.D. processed the data and extracted parameters for analyses, and constructed the simulation model with input from M.R.B. I.F., A.J.K. and M.D. analysed the data. M.D. and A.J.K. led writing of the paper with input from all authors, who read and approved the final manuscript. Competing interests. We declare we have no competing interests.
Funding. Experimental work was funded through UKOARP grant NE/H01750X/1, co-funded by the Natural Environment Research Council (NERC), the Department for Environment, Food and Rural Affairs (Defra), and the Department of Energy and Climate Change (DECC) and a NERC Fellowship awarded to A.J.K. (NE/H016600/3). M.D. was supported by a Swansea University PhD scholarship, A.J.K. by a NERC Fellowship (NE/H016600/3), I.F. by a German Research Foundation Fellowship (DFG; FU-985/1-1) and M.R.B. by an EPSRC grant (EP $\backslash$ M00061 $\backslash 1$ ). Acknowledgements. Thanks to Maria Scolamacchia, Jake Scolding, Alex Keay, Robin Shields and Kevin Flynn for logistic and technical support, and to A.J.K.'s and I.F.'s research teams for useful discussion and comments. Thanks also to Josefin Sundin and one anonymous referee for their constructive feedback and to Alice Power, the Editorial Coordinator.

## References

1. Doney SC, Fabry VJ, Feely RA, Kleypas JA. 2009 Ocean acidification: the other $\mathrm{CO}_{2}$ problem. Annu. Rev. Mar. Sci. 1, 169-192. (doi:10.1146/annurev. marine.010908.163834)
2. Fabry VJ, Seibel BA, Feely RA, Orr JC. 2008 Impacts of ocean acidification on marine fauna and ecosystem processes. ICES J. Mar. Sci. 65, 414-432. (doi:10.1093/icesjms/fsn048)
3. Vitousek PM, Aber JD, Howarth RW, Likens GE, Matson PA, Schindler DW, Schlesinger WH, Tilman D. 1997 Human alteration of the global nitrogen cycle: sources and consequences. Ecol. Appl. 7, 737-750. (doi:10.2307/2269431)
4. McNeil BI, Sasse TP. 2016 Future ocean hypercapnia driven by anthropogenic amplification of the natural $\mathrm{CO}_{2}$ cycle. Nature 529, 383-386. (doi:10.1038/
nature16156)
5. Dupont S, Portner H. 2013 Get ready for ocean acidification. Nature 498, 429. (doi:10.1038/ 498429a)
6. Nagelkerken I, Munday PL. 2016 Animal behaviour shapes the ecological effects of ocean acidification and warming: moving from individual to community-level responses. Glob. Change Biol. 22, 974-989. (doi:10.1111/gcb.13167)
7. Russell BD, Harley CDG, Wernberg T, Mieszkowska N, Widdicombe S, Hall-Spencer JM, Connell SD. 2012 Predicting ecosystem shifts requires new approaches that integrate the effects of climate change across entire systems. Biol. Lett. 8, 164-166. (doi:10.1098/rsbl.2011.0779)
8. Nussey DH, Wilson AJ, Brommer JE. 2007 The evolutionary ecology of individual phenotypic plasticity in wild populations. J. Evol. Biol. 20, 831-844. (doi:10.1111/j.1420-9101.2007.01300.x)
9. Visser ME. 2008 Keeping up with a warming world; assessing the rate of adaptation to climate change. Proc. R. SoC. B 275, 649-659. (doi:10.1098/rspb. 2007.0997)
10. Ferrari MCO, McCormick MI, Munday PL, Meekan MG, Dixson DL, Lonnstedt 0, Chivers DP. 2011 Putting prey and predator into the $\mathrm{CO}_{2}$ equation-qualitative and quantitative effects of ocean acidification on predator-prey interactions. Ecol. Lett. 14, 1143-1148. (doi:10.1111/j.1461-0248. 2011.01683.x)
11. Jutfelt F, de Souza KB, Vuylsteke A, Sturve J. 2013 Behavioural disturbances in a temperate fish exposed to sustained high- $\mathrm{CO}_{2}$ levels. PLoS ONE 8, e65825. (doi:10.1371/journal.pone. 0065825)
12. Manriquez PH, Elisa Jara M, Loreto Mardones M, Torres R, Navarro JM, Lardies MA, Vargas CA, Duarte C, Lagos NA. 2014 Ocean acidification affects predator avoidance behaviour but not prey detection in the early ontogeny of a keystone species. Mar. Ecol. Prog. Ser. 502, 157-167. (doi:10.3354/meps10703)
13. Munday PL, Cheal AJ, Dixson DL, Rummer JL, Fabricius KE. 2014 Behavioural impairment in reef fishes caused by ocean acidification at $\mathrm{CO}_{2}$ seeps. Nat. Clim. Change 4, 487-492. (doi:10.1038/ nclimate2195)
14. Heuer RM, Grosell M. 2014 Physiological impacts of elevated carbon dioxide and ocean acidification on fish. Am. J. Physiol. Regul. Integrat. Comp. Physiol. 307, R1061-R1084. (doi:10.1152/ajpregu.00064. 2014)
15. Clements JC, Hunt HL. 2015 Marine animal behaviour in a high $\mathrm{CO}_{2}$ ocean. Mar. Ecol. Prog. Ser. 536, 259-279. (doi:10.3354/meps11426)
16. Nilsson GE, Dixson DL, Domenici P, McCormick MI, Sorensen C, Watson S-A, Munday PL. 2012 Near-future carbon dioxide levels alter fish behaviour by interfering with neurotransmitter function. Nat. Clim. Change 2, 201-204. (doi:10.1038/nclimate1352)
17. FA0.2011-2016 Fisheries and aquaculture software. FishStatJ-software for fishery statistical time series. Updated 21 July 2016. Rome, Italy: FAO Fisheries and Aquaculture Department. http://www.fao.org/ fishery/statistics/software/fishstatj/en.
18. Perez-Escudero A, Vicente-Page J, Hinz RC, Arganda S, de Polavieja GG. 2014 idTracker: tracking individuals in a group by automatic identification of unmarked animals. Nat. Methods 11,743-748. (doi:10.1038/nmeth.2994)
19. Budaev SV. 1997 'Personality' in the guppy (Poecilia reticulata): a correlational study of exploratory behavior and social tendency. J. Comp. Psychol. 111,399-411. (doi:10.1037/0735-7036.111. 4.399)
20. Fürtbauer I, Pond A, Heistermann M, King AJ. 2015 Personality, plasticity and predation: linking endocrine and behavioural reaction norms in stickleback fish. Funct. Ecol. 29, 931-940. (doi:10.1111/1365-2435.12400)
21. Berdahl A, Torney CJ, Ioannou CC, Faria JJ, Couzin ID. 2013 Emergent sensing of complex environments by mobile animal groups. Science 339, 574-576. (doi:10.1126/science.1225883)
22. Guttal V, Couzin ID. 2010 Social interactions, information use, and the evolution of collective migration. Proc. Natl Acad. Sci. USA 107, 16 172-16 177. (doi:10.1073/pnas. 10068 74107)
23. Devine BM, Munday PL, Jones GP. 2012 Homing ability of adult cardinalfish is affected by elevated carbon dioxide. Oecologia 168, 269-276. (doi:10.1007/s00442-011-2081-2)
24. Pope EC et al. 2014 European sea bass, Dicentrarchus labrax, in a changing ocean. Biogeosciences 11, 2519-2530. (doi:10.5194/bg-11-2519-2014)
25. Claireaux G, Couturier C, Groison AL. 2006 Effect of temperature on maximum swimming speed and cost of transport in juvenile European sea bass (Dicentrarchus labrax). J. Exp. Biol. 209, 3420-3428. (doi:10.1242/jeb.02346)
26. Jolliffe IT. 2002 Principal component analysis. New York, NY: Springer.
27. R Core Team. 2013 R v. 3.0.2 (2013-09-25).
28. Bates D, Maechler M, Bolker B, Walker S. 2013 Fitting linear mixed-effects models using Ime 4. (https://arxiv.org/abs/1406.5823)
29. Bignami S, Sponaugle S, Cowen RK. 2013 Response to ocean acidification in larvae of a large tropical marine fish, Rachycentron canadum. Glob. Change Biol. 19, 996-1006. (doi:10.1111/gcb. 12133)
30. Bignami S, Sponaugle S, Cowen RK. 2014 Effects of ocean acidification on the larvae of a high-value pelagic fisheries species, mahi-mahi Coryphaena hippurus. Aquat. Biol. 21, 249-260. (doi:10.3354 ab00598)
31. Jutfelt F, Hedgarde M. 2013 Atlantic cod actively avoid $\mathrm{CO}_{2}$ and predator odour, even after long-term $\mathrm{CO}_{2}$ exposure. Front. Zool. 10, 81. (doi:10.1186/1742-9994-10-81)
32. Maneja RH et al. 2015 The swimming kinematics and foraging behavior of larval Atlantic herring (Clupea harengus L.) are unaffected by elevated pCO2. J. Exp. Mar. Biol. Ecol. 466, 42-48 (doi:10.1016/j.jembe.2015.02.008)
33. Ioannou CC, Singh M, Couzin ID. 2015 Potential leaders trade off goal-oriented and socially oriented behavior in mobile animal groups. Am. Nat. 186, 284-293. (doi:10.1086/681988)
34. Conradt L, Krause J, Couzin ID, Roper TJ. 2009 'Leading according to need' in self-organizing groups. Am. Nat. 173, 304-312. (doi:10.1086/596532
35. Ward AJW, Herbert-Read JE, Jordan LA, James R, Krause J, Ma Q, Rubenstein DI, Sumpter DJT, Morrel LJ. 2013 Initiators, leaders, and recruitment mechanisms in the collective movements of damselfish. Am. Nat. 181, 748-760. (doi:10.1086 670242)
36. King AJ, Fürtbauer I, Mamuneas D, James C, Manica A. 2013 Sex-differences and temporal consistency in stickleback fish boldness. PLOS ONE 8,6. (doi:10.1371/journal.pone.0081116)
37. Munday PL, Dixson DL, McCormick MI, Meekan M, Ferrari MCO, Chivers DP. 2010 Replenishment of fish populations is threatened by ocean acidification. Proc. Natl Acad. Sci. USA 107, 12 930-12 934. (doi:10.1073/pnas.1004519107)
38. Sundin J, Jutfelt F. 2016 9-28 d of exposure to elevated $\mathrm{pCO}_{2}$ reduces avoidance of predator odour but had no effect on behavioural lateralization or swimming activity in a temperate wrasse (Ctenolabrus rupestris). ICES J. Mar. Sci. 73, 620-632 (doi:10.1093/icesjms/fsv101)
39. Duteil M, Pope EC, Pérez-Escudero A, de Polavieja GG, Fürtbauer I, Brown MR, King AJ. 2016 Data from: European sea bass show behavioural resilience to near-future ocean acidification. Dryad Digital Repository. (http://dx.doi.org/10.5061/dryad.2dc8k)

## REFERENCES

[1] M. Duteil et al., "European sea bass show behavioural resilience to near-future ocean acidification," R. Soc. Open Sci., vol. 3, no. 11, p. 160656, Nov. 2016.
[2] D. H. Meadows and D. Wright, Thinking in systems : a primer. Chelsea Green Publishing, 2008.
[3] K. E. Boulding, "General Systems Theory—The Skeleton of Science," Manage. Sci., vol. 2, no. 3, pp. 197-208, Apr. 1956.
[4] M. Levy, J. P. Perdew, and V. Sahni, "Exact differential equation for the density and ionization energy of a many-particle system," Phys. Rev. A, vol. 30, no. 5, pp. 2745-2748, Nov. 1984.
[5] H. Svensson, "Isoelectric fractionation, analysis, and characterization of ampholytes in natural pH gradients. I. The Differential Equation of Solute Concentrations at a Steady State and its Solution for Simple Cases," Acta chemica Scandinavica, vol. 15. pp. 325-341, 1961.
[6] J. Baranyi and T. A. Roberts, "A dynamic approach to predicting bacterial growth in food," Int. J. Food Microbiol., vol. 23, no. 3-4, pp. 277-294, Nov. 1994.
[7] R. J. H. Beverton and S. J. Holt, On the Dynamics of Exploited Fish Populations. Springer Netherlands, 1993.
[8] M. Fujita and T. Mori, "Structural stability and evolution of urban systems," Reg. Sci. Urban Econ., vol. 27, no. 4-5, pp. 399-442, Aug. 1997.
[9] B. S. Goh, G. Leitmann, and T. L. Vincent, "Optimal control of a prey-predator system," Math. Biosci., vol. 19, no. 3-4, pp. 263-286, Apr. 1974.
[10] D. Strömbom, "Attraction Based Models of Collective Motion," Department of Mathematics, 2013.
[11] J. Goldstein, "Emergence as a Construct: History and Issues," Emergence, vol. 1, no. 1, pp. 49-72, Mar. 1999.
[12] D. A. Kessler, J. Koplik, and H. Levine, "Numerical simulation of two-
dimensional snowflake growth," Phys. Rev. A, vol. 30, no. 5, pp. 2820-2823, Nov. 1984.
[13] T. Kawasaki and H. Tanaka, "Formation of a crystal nucleus from liquid.," Proc. Natl. Acad. Sci. U. S. A., vol. 107, no. 32, pp. 14036-41, Aug. 2010.
[14] M. D. Hollingsworth, "Crystal engineering: from structure to function.," Science, vol. 295, no. 5564, pp. 2410-3, Mar. 2002.
[15] E. Bonabeau et al., "A model for the emergence of pillars, walls and royal chambers in termite nests," Philos. Trans. R. Soc. B Biol. Sci., vol. 353, no. 1375, pp. 1561-1576, Oct. 1998.
[16] M. Bindschadler, J. L. McGrath, P. O. Tran, M. D. Sjaastad, and T. E. Machen, "Sheet migration by wounded monolayers as an emergent property of single-cell dynamics.," J. Cell Sci., vol. 120, no. Pt 5, pp. 876-84, Mar. 2007.
[17] M. Bar and N. Ori, "Leaf development and morphogenesis.," Development, vol. 141, no. 22, pp. 4219-30, Nov. 2014.
[18] E. Sharon, M. Marder, and H. L. Swinney, "Leaves, Flowers and Garbage Bags: Making Waves," Am. Sci., vol. 92, pp. 254-261, 2004.
[19] J. A. S. Kelso, Dynamic patterns : the self-organization of brain and behavior. MIT Press, 1997.
[20] N. J. Mlot, C. A. Tovey, and D. L. Hu, "Fire ants self-assemble into waterproof rafts to survive floods.," Proc. Natl. Acad. Sci. U. S. A., vol. 108, no. 19, pp. 7669-73, May 2011.
[21] C. K. Hemelrijk, H. Hildenbrandt, J. Reinders, and E. J. Stamhuis, "Emergence of Oblong School Shape: Models and Empirical Data of Fish," Ethology, vol. 116, no. 11, pp. 1099-1112, Nov. 2010.
[22] C. K. Hemelrijk and H. Hildenbrandt, "Self-Organized Shape and Frontal Density of Fish Schools," Ethology, vol. 114, no. 3, pp. 245-254, Mar. 2008.
[23] F. Cucker and S. Smale, "Emergent behavior in flocks," IEEE Trans. Automat. Contr., vol. 52, no. 5, pp. 852-862, May 2007.
[24] F. H. Heppner, "Avian Flight Formations," Bird-Banding, vol. 45, no. 2, p. 160,
1974.
[25] Guinness World Records Limited, "Largest Mexican Wave," Guinness World Records Limited, 2019. [Online]. Available: http://www.guinnessworldrecords.com/world-records/largest-mexican-wave.
[Accessed: 21-Jan-2019].
[26] I. Farkas, D. Helbing, and T. Vicsek, "Crowd behaves as excitable media during Mexican wave," Oct. 2002.
[27] T. C. Schelling, "Dynamic models of segregation†," J. Math. Sociol., vol. 1, no. 2, pp. 143-186, Jul. 1971.
[28] V. Hart and N. Case, "Parable of the Polygons - a playable post on the shape of society," 2014. [Online]. Available: https://ncase.me/polygons/. [Accessed: 21-Jan-2019].
[29] P. R. Krugman, The self-organizing economy. .
[30] "Wikimedia Commons," Wikimedia, 2001. [Online]. Available: https://commons.wikimedia.org/. [Accessed: 16-Aug-2018].
[31] C. Darwin, On the Origin of Species. Routledge, 1859.
[32] J. F. Wittenberger and R. L. Tilson, "The Evolution of Monogamy: Hypotheses and Evidence," Annu. Rev. Ecol. Syst., vol. 11, no. 1, pp. 197-232, Nov. 1980.
[33] F. B. M. de (Frans B. M. . Waal, Chimpanzee politics : power and sex among apes. Johns Hopkins University Press, 2007.
[34] F. A. Huntingford and A. K. Turner, Animal Conflict. Springer Netherlands, 1987.
[35] M. Woyciechowski and A. Łomnicki, "Multiple mating of queens and the sterility of workers among eusocial hymenoptera," J. Theor. Biol., vol. 128, no. 3, pp. 317-327, Oct. 1987.
[36] F. E. Rheindt, C. P. Strehl, and J. Gadau, "A genetic component in the determination of worker polymorphism in the Florida harvester ant Pogonomyrmex badius," Insectes Soc., vol. 52, no. 2, pp. 163-168, May 2005.
[37] N. L. Naeger, M. Peso, N. Even, A. B. Barron, and G. E. Robinson, "Altruistic

Behavior by Egg-Laying Worker Honeybees," Curr. Biol., vol. 23, no. 16, pp. 1574-1578, Aug. 2013.
[38] L. Straub, G. R. Williams, J. Pettis, I. Fries, and P. Neumann, "Superorganism resilience: eusociality and susceptibility of ecosystem service providing insects to stressors," Curr. Opin. Insect Sci., vol. 12, pp. 109-112, Dec. 2015.
[39] W. A. Foster and J. E. Treherne, "Evidence for the dilution effect in the selfish herd from fish predation on a marine insect," Nature, vol. 293, no. 5832, pp. 466-467, Oct. 1981.
[40] W. D. Hamilton, "Geometry for the selfish herd," J. Theor. Biol., vol. 31, no. 2, pp. 295-311, May 1971.
[41] L. S. Rayor and G. W. Uetz, "Ontogenetic shifts within the selfish herd: predation risk and foraging trade-offs change with age in colonial web-building spiders," Oecologia, vol. 95, no. 1, pp. 1-8, Mar. 1993.
[42] M. Milinski and R. Heller, "Influence of a predator on the optimal foraging behaviour of sticklebacks (Gasterosteus aculeatus L.)," Nature, vol. 275, no. 5681, pp. 642-644, Oct. 1978.
[43] G. Roberts, "Why individual vigilance declines as group size increases," Anim. Behav., vol. 51, no. 5, pp. 1077-1086, May 1996.
[44] J. Maynard Smith and D. Harper, Animal signals. Oxford University Press, 2003.
[45] B. Graw and M. B. Manser, "The function of mobbing in cooperative meerkats," Anim. Behav., vol. 74, no. 3, pp. 507-517, Sep. 2007.
[46] C. W. Clark and M. Mangel, "The evolutionary advantages of group foraging," Theor. Popul. Biol., vol. 30, no. 1, pp. 45-75, Aug. 1986.
[47] S. Creel and N. M. Creel, "Communal hunting and pack size in African wild dogs, Lycaon pictus," Anim. Behav., vol. 50, no. 5, pp. 1325-1339, Jan. 1995.
[48] M. Daghooghi and I. Borazjani, "The hydrodynamic advantages of synchronized swimming in a rectangular pattern," Bioinspir. Biomim., vol. 10, no. 5, p. 056018, Oct. 2015.
[49] S. J. Portugal et al., "Upwash exploitation and downwash avoidance by flap
phasing in ibis formation flight," Nature, vol. 505, no. 7483, pp. 399-402, Jan. 2014.
[50] S. R. X. Dall, A. M. Bell, D. I. Bolnick, and F. L. W. Ratnieks, "An evolutionary ecology of individual differences," Ecol. Lett., vol. 15, no. 10, pp. 1189-1198, Oct. 2012.
[51] J. H. Fewell, C. T. Holbrook, and P. F. Kukuk, "Increased group size promotes task specialization in a normally solitary halictine bee," Behaviour, vol. 150, no. 12, pp. 1-18, Jan. 2013.
[52] R. Bergmüller and M. Taborsky, "Animal personality due to social niche specialisation," Trends Ecol. Evol., vol. 25, no. 9, pp. 504-511, Sep. 2010.
[53] D. A. Harrison, K. H. Price, and M. P. Bell, "Beyond Relational Demography: Time and the Effects of Surface- and Deep-Level Diversity on Work Group Cohesion," Acad. Manag. J., vol. 41, no. 1, pp. 96-107, Feb. 1998.
[54] W. Wood, "Meta-analytic review of sex differences in group performance," Psychol. Bull., vol. 102, no. July, pp. 53-71, 1987.
[55] F. J. Milliken and L. L. Martins, "Searching for Common Threads: Understanding the Multiple Effects of Diversity in Organizational Groups," Acad. Manag. Rev., vol. 21, no. 2, pp. 402-433, Apr. 1996.
[56] G. M. Cooper, "The Origin and Evolution of Cells," 2000.
[57] Å. Brännström and D. J. T. Sumpter, "The role of competition and clustering in population dynamics," Proc. R. Soc. B Biol. Sci., vol. 272, no. 1576, pp. 20652072, 2005.
[58] W. Cresswell and J. L. Quinn, "Predicting the optimal prey group size from predator hunting behaviour," J. Anim. Ecol., vol. 80, no. 2, pp. 310-319, Mar. 2011.
[59] A. C. Markham, L. R. Gesquiere, S. C. Alberts, and J. Altmann, "Optimal group size in a highly social mammal.," Proc. Natl. Acad. Sci. U. S. A., vol. 112, no. 48, pp. 14882-7, Dec. 2015.
[60] M. J. Keeling, "The effects of local spatial structure on epidemiological
invasions," 1999.
[61] F. Liljeros, C. R. Edling, and L. A. N. Amaral, "Sexual networks: implications for the transmission of sexually transmitted infections," Microbes Infect., vol. 5, no. 2, pp. 189-196, Feb. 2003.
[62] F. Aureli et al., "Fission - Fusion Dynamics," Curr. Anthropol., vol. 49, no. 4, pp. 627-654, Aug. 2008.
[63] D. J. T. Sumpter, Collective animal behavior. Princeton University Press, 2010.
[64] I. L. Bajec and F. H. Heppner, "Organized flight in birds," Anim. Behav., vol. 78, no. 4, pp. 777-789, Oct. 2009.
[65] J. Gautrais et al., "Deciphering Interactions in Moving Animal Groups," PLoS Comput. Biol., vol. 8, no. 9, p. e1002678, Sep. 2012.
[66] E. Selous, Thought-transference (or What?) in Birds. London: Constable, 1931.
[67] H. Penrose, I Flew with the Birds. London: Country Life, 1949.
[68] R. W. Gerard, "Synchrony in flock wheeling," Science, vol. 97, no. 2511, pp. 160-1, Feb. 1943.
[69] Y. Hayakawa, "Spatiotemporal dynamics of skeins of wild geese," EPL (Europhysics Lett., vol. 89, no. 4, p. 48004, Feb. 2010.
[70] A. R. . Sinclair, The African buffalo: a study of resource limitation of populations. Chicago: University of Chicago Press, 1977.
[71] C. Becco, N. Vandewalle, J. Delcourt, and P. Poncin, "Experimental evidences of a structural and dynamical transition in fish school," Phys. A Stat. Mech. its Appl., vol. 367, pp. 487-493, Jul. 2006.
[72] J. K. Parrish and P. Turchin, "Individual decisions, traffic rules, and emergent pattern in schooling fish," in Animal Groups in Three Dimensions, J. K. Parrish and W. M. Hamner, Eds. Cambridge: Cambridge University Press, 1997, pp. 126-142.
[73] B. L. Partridge, T. Pitcher, J. M. Cullen, and J. Wilson, "The three-dimensional structure of fish schools," Behav. Ecol. Sociobiol., vol. 6, no. 4, pp. 277-288,

Mar. 1980.
[74] I. D. Couzin and N. R. Franks, "Self-organized lane formation and optimized traffic flow in army ants.," Proceedings. Biol. Sci., vol. 270, no. 1511, pp. 13946, Jan. 2003.
[75] J. Buhl et al., "From Disorder to Order in Marching Locusts," vol. 312, no. 5778, Jun. 2006.
[76] Y. Wu, A. D. Kaiser, Y. Jiang, and M. S. Alber, "Periodic reversal of direction allows Myxobacteria to swarm.," Proc. Natl. Acad. Sci. U. S. A., vol. 106, no. 4, pp. 1222-7, Jan. 2009.
[77] J. W. Jolles, A. J. King, A. Manica, and A. Thornton, "Heterogeneous structure in mixed-species corvid flocks in flight," Anim. Behav., vol. 85, no. 4, pp. 743750, Apr. 2013.
[78] J. A. Shapiro, "Organization of developing Escherichia coli colonies viewed by scanning electron microscopy.," J. Bacteriol., vol. 169, no. 1, pp. 142-56, Jan. 1987.
[79] T. Vicsek, M. Cserző, and V. K. Horváth, "Self-affine growth of bacterial colonies," Phys. A Stat. Mech. its Appl., vol. 167, no. 2, pp. 315-321, Aug. 1990.
[80] A. J. Engler, S. Sen, H. L. Sweeney, and D. E. Discher, "Matrix Elasticity Directs Stem Cell Lineage Specification," Cell, vol. 126, no. 4, pp. 677-689, Aug. 2006.
[81] A. Pérez-Escudero, J. Vicente-Page, R. C. Hinz, S. Arganda, and G. G. de Polavieja, "idTracker: tracking individuals in a group by automatic identification of unmarked animals," Nat. Methods, vol. 11, no. 7, pp. 743-748, Jul. 2014.
[82] R. Lukeman, Y.-X. Li, and L. Edelstein-Keshet, "Inferring individual rules from collective behavior.," Proc. Natl. Acad. Sci. U. S. A., vol. 107, no. 28, pp. 1257680, Jul. 2010.
[83] J. Delcourt, C. Becco, N. Vandewalle, and P. Poncin, "A video multitracking system for quantification of individual behavior in a large fish shoal: Advantages and limits," Behav. Res. Methods, vol. 41, no. 1, pp. 228-235, Feb. 2009.
[84] M. Ballerini et al., "Interaction ruling animal collective behavior depends on
topological rather than metric distance: evidence from a field study.," Proc. Natl. Acad. Sci. U. S. A., vol. 105, no. 4, pp. 1232-7, Jan. 2008.
[85] A. I. Dell et al., "Automated image-based tracking and its application in ecology," Trends Ecol. Evol., vol. 29, no. 7, pp. 417-428, 2014.
[86] O. Mirat, J. R. Sternberg, K. E. Severi, and C. Wyart, "ZebraZoom: an automated program for high-throughput behavioral analysis and categorization," Front. Neural Circuits, vol. 7, p. 107, Jun. 2013.
[87] A. . Spink, R. A. . Tegelenbosch, M. O. . Buma, and L. P. J. . Noldus, "The EthoVision video tracking system-A tool for behavioral phenotyping of transgenic mice," Physiol. Behav., vol. 73, no. 5, pp. 731-744, Aug. 2001.
[88] S. Narkilahti, K. Rajala, H. Pihlajamäki, R. Suuronen, O. Hovatta, and H. Skottman, "Monitoring and analysis of dynamic growth of human embryonic stem cells: comparison of automated instrumentation and conventional culturing methods," Biomed. Eng. Online, vol. 6, no. 1, p. 11, Apr. 2007.
[89] S. Ohayon, O. Avni, A. L. Taylor, P. Perona, and S. E. Roian Egnor, "Automated multi-day tracking of marked mice for the analysis of social behaviour," J. Neurosci. Methods, vol. 219, no. 1, pp. 10-19, Sep. 2013.
[90] M. J. Hansen, A. J. W. Ward, I. Fürtbauer, and A. J. King, "Environmental quality determines finder-joiner dynamics in socially foraging three-spined sticklebacks (Gasterosteus aculeatus)," Behav. Ecol. Sociobiol., vol. 70, no. 6, pp. 889-899, Jun. 2016.
[91] I. Fürtbauer, A. J. King, and M. Heistermann, "Visible implant elastomer (VIE) tagging and simulated predation risk elicit similar physiological stress responses in three-spined stickleback Gasterosteus aculeatus," J. Fish Biol., vol. 86, no. 5, pp. 1644-1649, May 2015.
[92] S. J. Dahlbom, D. Lagman, K. Lundstedt-Enkel, L. F. Sundström, and S. Winberg, "Boldness Predicts Social Status in Zebrafish (Danio rerio)," PLoS One, vol. 6, no. 8, p. e23565, Aug. 2011.
[93] R. L. Dennis, R. C. Newberry, H.-W. Cheng, and I. Estevez, "Appearance Matters: Artificial Marking Alters Aggression and Stress," Poult. Sci., vol. 87,
no. 10, pp. 1939-1946, Oct. 2008.
[94] H. D. Summers et al., "Statistical analysis of nanoparticle dosing in a dynamic cellular system," Nat. Nanotechnol., vol. 6, no. 3, pp. 170-174, Mar. 2011.
[95] D. A. Persons et al., "Use of the green fluorescent protein as a marker to identify and track genetically modified hematopoietic cells," Nat. Med., vol. 4, no. 10, pp. 1201-1205, Oct. 1998.
[96] J. M. Stanton, "ICAO and the Biometric RFID Passport: History and A nalysis," in Playing the Identity Card: Surveillance, Security and Identification in Global Perspective, C. J. Bennett and D. Lyon, Eds. Routledge, 2013, pp. 253-267.
[97] P. R. Lowe and D. G. J. Small, "Electronic anti-shoplifting system employing an RFID tag," US5874896A, 1996.
[98] V. Chawla and D. Ha, "An overview of passive RFID," IEEE Commun. Mag., vol. 45, no. 9, pp. 11-17, Sep. 2007.
[99] R. E. Floyd, "RFID in Animal-Tracking Applications," IEEE Potentials, vol. 34, no. 5, pp. 32-33, Sep. 2015.
[100] L. Catarinucci et al., "Near Field UHF RFID Antenna System Enabling the Tracking of Small Laboratory Animals," Int. J. Antennas Propag., vol. 2013, pp. 1-10, Jun. 2013.
[101] M. Kritzler, L. Lewejohann, A. Krger, M. Raubal, and N. Sachser, "An rfid based tracking system for laboratory mice in a semi-natural environment," Pervasive 2006 Work. Proceedings. Dublin, Irel., 2006.
[102] J. Aguzzi et al., "A New Laboratory Radio Frequency Identification (RFID) System for Behavioural Tracking of Marine Organisms," Sensors, vol. 11, no. 10, pp. 9532-9548, Oct. 2011.
[103] M. Wikelski et al., "Large-Range Movements of Neotropical Orchid Bees Observed via Radio Telemetry," PLoS One, vol. 5, no. 5, p. e10738, May 2010.
[104] L. Gygax, G. Neisen, and H. Bollhalder, "Accuracy and validation of a radarbased automatic local position measurement system for tracking dairy cows in free-stall barns," Comput. Electron. Agric., vol. 56, no. 1, pp. 23-33, Mar. 2007.
[105] J. Krause, S. Krause, R. Arlinghaus, I. Psorakis, S. Roberts, and C. Rutz, "Reality mining of animal social systems," Trends Ecol. Evol., vol. 28, no. 9, pp. 541551, Sep. 2013.
[106] C. Rutz and G. C. Hays, "New frontiers in biologging science.," Biol. Lett., vol. 5, no. 3, pp. 289-92, Jun. 2009.
[107] P. G. Ryan, S. L. Petersen, G. Peters, and D. Grémillet, "GPS tracking a marine predator: the effects of precision, resolution and sampling rate on foraging tracks of African Penguins," Mar. Biol., vol. 145, no. 2, pp. 215-223, Aug. 2004.
[108] G. Schofield et al., "Novel GPS tracking of sea turtles as a tool for conservation management," J. Exp. Mar. Bio. Ecol., vol. 347, no. 1-2, pp. 58-68, Aug. 2007.
[109] H. Kaartinen et al., "Accuracy of Kinematic Positioning Using Global Satellite Navigation Systems under Forest Canopies," Forests, vol. 6, no. 12, pp. 32183236, Sep. 2015.
[110] L. L. Arnold and P. A. Zandbergen, "Positional accuracy of the Wide Area Augmentation System in consumer-grade GPS units," Comput. Geosci., vol. 37, no. 7, pp. 883-892, Jul. 2011.
[111] H. Haddadi et al., "Determining association networks in social animals: choosing spatial-temporal criteria and sampling rates," Behav. Ecol. Sociobiol., vol. 65, no. 8, pp. 1659-1668, Aug. 2011.
[112] B. Van Moorter, C. M. Rolandsen, M. Basille, and J.-M. Gaillard, "Movement is the glue connecting home ranges and habitat selection," J. Anim. Ecol., vol. 85, no. 1, pp. 21-31, Jan. 2016.
[113] F. M. van Beest, L. E. Loe, A. Mysterud, and J. M. Milner, "Comparative Space Use and Habitat Selection of Moose Around Feeding Stations," J. Wildl. Manage., vol. 74, no. 2, pp. 219-227, Feb. 2010.
[114] S. B. Merrill and L. D. Mech, "The Usefulness of GPS Telemetry to Study Wolf Circadian and Social Activity," Wildlife Society Bulletin (1973-2006), vol. 31. WileyWildlife Society, pp. 947-960.
[115] C. Brooks, C. Bonyongo, and S. Harris, "Effects of Global Positioning System

Collar Weight on Zebra Behavior and Location Error," J. Wildl. Manage., vol. 72, no. 2, pp. 527-534, Feb. 2008.
[116] P. Berthold, W. v. d. Bossche, Z. Jakubiec, C. Kaatz, M. Kaatz, and U. Querner, "Long-term satellite tracking sheds light upon variable migration strategies of White Storks (Ciconia ciconia)," J. für Ornithol., vol. 143, no. 4, pp. 489-493, Oct. 2002.
[117] R. E. van Wijk, A. Kölzsch, H. Kruckenberg, B. S. Ebbinge, G. J. D. M. Müskens, and B. A. Nolet, "Individually tracked geese follow peaks of temperature acceleration during spring migration," Oikos, vol. 121, no. 5, pp. 655-664, May 2012.
[118] F. Bonadonna, M.-A. Lea, and C. Guinet, "Foraging routes of Antarctic fur seals ( Arctocephalus gazella ) investigated by the concurrent use of satellite tracking and time-depth recorders," Polar Biol., vol. 23, no. 3, pp. 149-159, Feb. 2000.
[119] Y. Naito, "New steps in bio-logging science," Mem. Natl. Inst. Polar Res., vol. 58, pp. 50-57, 2008.
[120] Y. Ropert-Coudert and R. P. Wilson, "Trends and perspectives in animalattached remote sensing," Front. Ecol. Environ., vol. 3, no. 8, pp. 437-444, Oct. 2005.
[121] N. Liebsch, R. P. Wilson, H. Bornemann, D. Adelung, and J. Plötz, "Mouthing off about fish capture: Jaw movement in pinnipeds reveals the real secrets of ingestion," Deep Sea Res. Part II Top. Stud. Oceanogr., vol. 54, no. 3-4, pp. 256-269, Feb. 2007.
[122] S. J. Cooke et al., "Biotelemetry: a mechanistic approach to ecology," Trends Ecol. Evol., vol. 19, no. 6, pp. 334-343, Jun. 2004.
[123] P. J. Ponganis, "Bio-logging of physiological parameters in higher marine vertebrates," Deep Sea Res. Part II Top. Stud. Oceanogr., vol. 54, no. 3-4, pp. 183-192, Feb. 2007.
[124] R. P. Wilson, J. Cooper, and J. Plötz, "Can we determine when marine endotherms feed? A case study with seabirds," J. Exp. Biol., vol. 167, no. September, pp. 267-275, 1992.
[125] G. Peters, "Measurement of digestive variables in free-living animals: gastric motility in penguins during foraging," Mem. Natl. Inst. Polar Res., vol. 58, pp. 203-209, 2004.
[126] P. F. Scholander, "Experimental investigations on the respiratory function in diving mammals and birds."
[127] M. Fedak, "Marine animals as platforms for oceanographic sampling: a 'win/win' situation for biology and operational oceanography," Mem. Natl. Inst. Polar Res., vol. 58, pp. 133-147, 2004.
[128] G. W. Boehlert et al., "Autonomous Pinniped Environmental Samplers: Using Instrumented Animals as Oceanographic Data Collectors," J. Atmos. Ocean. Technol., vol. 18, no. 11, pp. 1882-1893, Nov. 2001.
[129] J.-B. Charrassin, Y.-H. Park, Y. Le Maho, and C.-A. Bost, "Penguins as oceanographers unravel hidden mechanisms of marine productivity," Ecol. Lett., vol. 5, no. 3, pp. 317-319, May 2002.
[130] P. M. Webb, D. E. Crocker, S. B. Blackwell, D. P. Costa, and B. J. Boeuf, "Effects of buoyancy on the diving behavior of northern elephant seals.," J. Exp. Biol., vol. 201, no. 16, 1998.
[131] R. M. Casper, "Guidelines for the instrumentation of wild birds and mammals," Anim. Behav., vol. 78, no. 6, pp. 1477-1483, Dec. 2009.
[132] R. A. Phillips, J. C. Xavier, and J. P. Croxall, "Effects of satellite transmitters on albatrosses and petrels," Auk Ornithol. Adv., vol. 120, no. 4, pp. 1082-1090, 2003.
[133] Y. Ropert-Coudert, "Bio-logging: recording the ecophysiology and behaviour of animals moving freely in their environment," in Sensors for ecology: Towards integrated knowledge of ecosystems, Paris, 2012, pp. 17-41.
[134] R. Wilson, E. Shepard, and N. Liebsch, "Prying into the intimate details of animal lives: use of a daily diary on animals," Endanger. Species Res., vol. 4, no. 1-2, pp. 123-137, Jan. 2008.
[135] S. Rotics et al., "Wintering in Europe instead of Africa enhances juvenile
survival in a long-distance migrant," Anim. Behav., vol. 126, pp. 79-88, Apr. 2017.
[136] G. Fehlmann et al., "Identification of behaviours from accelerometer data in a wild social primate," Anim. Biotelemetry, vol. 5, no. 1, p. 6, Dec. 2017.
[137] N. C. Makris, P. Ratilal, D. T. Symonds, S. Jagannathan, S. Lee, and R. W. Nero, "Fish population and behavior revealed by instantaneous continental shelf-scale imaging.," Science, vol. 311, no. 5761, pp. 660-3, Feb. 2006.
[138] N. Kulkarni, "A life-saving invention that prevents human stampedes," TEDNYC, 2018.
[Online].
Available: https://www.ted.com/talks/nilay_kulkarni_a_life_saving_invention_that_prevents _human_stampedes. [Accessed: 01-Mar-2018].
[139] G. D. Ruxton, T. N. Sherratt, and M. P. Speed, Avoiding attack: the evolutionary ecology of crypsis, warning signals and mimicry. Oxford University Press, 2004.
[140] S. V. S. Viscido, J. K. J. Parrish, and D. Grünbaum, "Individual behavior and emergent properties of fish schools: a comparison of observation and theory," Mar. Ecol. Prog. Ser., vol. 273, pp. 239-249, Jun. 2004.
[141] N. Miller and R. Gerlai, "Quantification of shoaling behaviour in zebrafish (Danio rerio)," Behav. Brain Res., vol. 184, no. 2, pp. 157-166, Dec. 2007.
[142] S. S. Killen, S. Marras, M. R. Ryan, P. Domenici, and D. J. McKenzie, "A relationship between metabolic rate and risk-taking behaviour is revealed during hypoxia in juvenile European sea bass," Funct. Ecol., vol. 26, no. 1, pp. 134-143, Feb. 2012.
[143] G. Buscaino et al., "Impact of an acoustic stimulus on the motility and blood parameters of European sea bass (Dicentrarchus labrax L.) and gilthead sea bream (Sparus aurata L.)," Mar. Environ. Res., vol. 69, no. 3, pp. 136-142, Apr. 2010.
[144] A. Barbosa, L. Litman, and R. T. Hanlon, "Changeable cuttlefish camouflage is influenced by horizontal and vertical aspects of the visual background," J. Comp. Physiol. A, vol. 194, no. 4, pp. 405-413, Apr. 2008.
[145] G. Horneck et al., "Microbial Rock Inhabitants Survive Hypervelocity Impacts on Mars-Like Host Planets: First Phase of Lithopanspermia Experimentally Tested," Astrobiology, vol. 8, no. 1, pp. 17-44, Feb. 2008.
[146] M. A. Saito, D. M. Sigman, and F. M. . Morel, "The bioinorganic chemistry of the ancient ocean: the co-evolution of cyanobacterial metal requirements and biogeochemical cycles at the Archean-Proterozoic boundary?," Inorganica Chim. Acta, vol. 356, pp. 308-318, Dec. 2003.
[147] D. R. Farine, A. Strandburg-Peshkin, I. D. Couzin, T. Y. Berger-Wolf, and M. C. Crofoot, "Individual variation in local interaction rules can explain emergent patterns of spatial organization in wild baboons.," Proceedings. Biol. Sci., vol. 284, no. 1853, p. 20162243, Apr. 2017.
[148] J. R. G. Dyer, D. P. Croft, L. J. Morrell, and J. Krause, "Shoal composition determines foraging success in the guppy," Behav. Ecol., vol. 20, no. 1, pp. 165171, Jan. 2009.
[149] W. Schuett and S. R. X. Dall, "Sex differences, social context and personality in zebra finches, Taeniopygia guttata," Anim. Behav., vol. 77, no. 5, pp. 1041-1050, May 2009.
[150] A. J. King, I. Fürtbauer, D. Mamuneas, C. James, and A. Manica, "SexDifferences and Temporal Consistency in Stickleback Fish Boldness," PLoS One, vol. 8, no. 12, p. e81116, Dec. 2013.
[151] A. J. King, L. J. Williams, and C. Mettke-Hofmann, "The effects of social conformity on Gouldian finch personality," Anim. Behav., vol. 99, pp. 25-31, Jan. 2015.
[152] C. Brown and E. Irving, "Individual personality traits influence group exploration in a feral guppy population," Behav. Ecol., vol. 25, no. 1, pp. 95-101, Jan. 2014.
[153] L. M. Aplin, D. R. Farine, R. P. Mann, and B. C. Sheldon, "Individual-level personality influences social foraging and collective behaviour in wild birds.," Proceedings. Biol. Sci., vol. 281, no. 1789, p. 20141016, Aug. 2014.
[154] M. Nagy, Z. Ákos, D. Biro, and T. Vicsek, "Hierarchical group dynamics in pigeon flocks," Nature, vol. 464, no. 7290, pp. 890-893, Apr. 2010.
[155] A. J. King, C. M. S. Douglas, E. Huchard, N. J. B. Isaac, and G. Cowlishaw, "Dominance and Affiliation Mediate Despotism in a Social Primate," Curr. Biol., vol. 18, no. 23, pp. 1833-1838, Dec. 2008.
[156] A. Frohnwieser, T. W. Pike, J. C. Murray, and A. Wilkinson, "Perception of artificial conspecifics by bearded dragons ( Pogona vitticeps )," Integr. Zool., Jan. 2018.
[157] J. Halloy et al., "Social integration of robots into groups of cockroaches to control self-organized choices.," Science, vol. 318, no. 5853, pp. 1155-8, Nov. 2007.
[158] J. J. Faria et al., "A novel method for investigating the collective behaviour of fish: introducing 'Robofish,"" Behav. Ecol. Sociobiol., vol. 64, no. 8, pp. 12111218, Aug. 2010.
[159] J. R. Stowers et al., "Virtual reality for freely moving animals," Nat. Methods, vol. 14, no. 10, pp. 995-1002, Aug. 2017.
[160] R. M. Calisi and G. E. Bentley, "Lab and field experiments: Are they the same animal?," Horm. Behav., vol. 56, no. 1, pp. 1-10, Jun. 2009.
[161] T. Piersma and J. Drent, "Phenotypic flexibility and the evolution of organismal design," Trends Ecol. Evol., vol. 18, no. 5, pp. 228-233, May 2003.
[162] N. J. Réale, D. and Dingemanse, "Personality and individual social specialisation," in Social Behaviour: Genes, Ecology and Evolution, T. et al. Szekely, Ed. Cambridge University Press, 2010.
[163] N. J. Dingemanse, A. J. N. Kazem, D. Réale, and J. Wright, "Behavioural reaction norms: animal personality meets individual plasticity," Trends Ecol. Evol., vol. 25, no. 2, pp. 81-89, Feb. 2010.
[164] D. W. A. Noble, P. Carazo, and M. J. Whiting, "Learning outdoors: male lizards show flexible spatial learning under semi-natural conditions.," Biol. Lett., vol. 8, no. 6, pp. 946-8, Dec. 2012.
[165] A. Berdahl, C. J. Torney, C. C. Ioannou, J. J. Faria, and I. D. Couzin, "Emergent sensing of complex environments by mobile animal groups.," Science, vol. 339,
no. 6119, pp. 574-6, Feb. 2013.
[166] D. Huh, G. A. Hamilton, and D. E. Ingber, "From 3D cell culture to organs-onchips," Trends Cell Biol., vol. 21, no. 12, pp. 745-754, Dec. 2011.
[167] E. C. Butcher, E. L. Berg, and E. J. Kunkel, "Systems biology in drug discovery," Nat. Biotechnol., vol. 22, no. 10, pp. 1253-1259, Oct. 2004.
[168] A. T. Showler, "A summary of control strategies for the desert locust, Schistocerca gregaria (Forskål)," Agric. Ecosyst. Environ., vol. 90, no. 1, pp. 97103, Jun. 2002.
[169] D. Helbing, I. Farkas, and T. Vicsek, "Simulating dynamical features of escape panic," Nature, vol. 407, no. 6803, pp. 487-490, Sep. 2000.
[170] K. M. Ngai, F. M. Burkle, A. Hsu, and E. B. Hsu, "Human Stampedes: A Systematic Review of Historical and Peer-Reviewed Sources," Disaster Med. Public Health Prep., vol. 3, no. 04, pp. 191-195, Dec. 2009.
[171] F. T. Illiyas, S. K. Mani, A. P. Pradeepkumar, and K. Mohan, "Human stampedes during religious festivals: A comparative review of mass gathering emergencies in India," Int. J. Disaster Risk Reduct., vol. 5, pp. 10-18, Sep. 2013.
[172] H. Gayathri, P. M. Aparna, and A. Verma, "A review of studies on understanding crowd dynamics in the context of crowd safety in mass religious gatherings," Int. J. Disaster Risk Reduct., vol. 25, pp. 82-91, Oct. 2017.
[173] A. Mintz, "Non-adaptive group behavior," J. Abnorm. Soc. Psychol., vol. 46(2), pp. 150-159, 1951.
[174] Y. Zhang, X. Chen, and N. Li, "Experimental Research of Relationship between Pedestrian Stopping Distance and Walking Speed to Avoid Pedestrian Crowd Stampede Accident," in ICCTP 2010, 2010, pp. 839-850.
[175] I. Fürtbauer, A. Pond, M. Heistermann, and A. J. King, "Personality, plasticity and predation: linking endocrine and behavioural reaction norms in stickleback fish," Funct. Ecol., vol. 29, no. 7, pp. 931-940, Jul. 2015.
[176] M. A. Bell, "Behavioural differences between individuals and two populations of stickleback (Gasterosteus aculeatus)," J. Evol. Biol., vol. 18, no. 2, pp. 464-473,

Oct. 2004.
[177] D. Biro, T. Sasaki, and S. J. Portugal, "Bringing a Time-Depth Perspective to Collective Animal Behaviour," Trends Ecol. Evol., vol. 31, no. 7, pp. 550-562, Jul. 2016.
[178] A. J. King, G. Fehlmann, D. Biro, A. J. Ward, and I. Fürtbauer, "Re-wilding Collective Behaviour: An Ecological Perspective.," Trends Ecol. Evol., vol. 0, no. 0, Apr. 2018.
[179] A. J. W. Ward, T. M. Schaerf, J. E. Herbert-Read, L. Morrell, D. J. T. Sumpter, and M. M. Webster, "Local interactions and global properties of wild, freeranging stickleback shoals," R. Soc. Open Sci., vol. 4, no. 7, p. 170043, Jul. 2017.
[180] M. Niazi and A. Hussain, "Agent-based computing from multi-agent systems to agent-based models: a visual survey," Scientometrics, vol. 89, no. 2, pp. 479499, Nov. 2011.
[181] L. Tesfatsion, "Agent-based Computational Economics: A constructive approach to economic theory," Handb. Comput. Econ. Vol. 2 Agent-Based Comput. Econ., no. October 2003, p. 55, 2005.
[182] D. C. Parker, S. M. Manson, M. A. Janssen, M. J. Hoffmann, and P. Deadman, "Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review," Annals of the Association of American Geographers, vol. 93. Taylor \& Francis, Ltd.Association of American Geographers, pp. 314-337.
[183] J. L. Segovia-Juarez, S. Ganguli, and D. Kirschner, "Identifying control mechanisms of granuloma formation during M. tuberculosis infection using an agent-based model," J. Theor. Biol., vol. 231, no. 3, pp. 357-376, 2004.
[184] J. Toner and Y. Tu, "Long-Range Order in a Two-Dimensional Dynamical XY Model: How Birds Fly Together," Phys. Rev. Lett., vol. 75, no. 23, pp. 43264329, Dec. 1995.
[185] A. Mogilner and L. Edelstein-Keshet, "Spatio-angular order in populations of self-aligning objects: formation of oriented patches," Phys. D Nonlinear Phenom., vol. 89, no. 3-4, pp. 346-367, Jan. 1996.
[186] G. Flierl, D. Grünbaum, S. Levins, and D. Olson, "From Individuals to Aggregations: the Interplay between Behavior and Physics," J. Theor. Biol., vol. 196, no. 4, pp. 397-454, Feb. 1999.
[187] A. Mogilner and L. Edelstein-Keshet, "A non-local model for a swarm," J. Math. Biol., vol. 38, no. 6, pp. 534-570, Jun. 1999.
[188] H. Levine, W.-J. Rappel, and I. Cohen, "Self-organization in systems of selfpropelled particles," Phys. Rev. E, vol. 63, no. 1, p. 017101, Dec. 2000.
[189] C. M. Topaz and A. L. Bertozzi, "Swarming Patterns in a Two-Dimensional Kinematic Model for Biological Groups," SIAM J. Appl. Math., vol. 65, no. 1, pp. 152-174, Jan. 2004.
[190] C. M. Topaz, A. L. Bertozzi, and M. A. Lewis, "A Nonlocal Continuum Model for Biological Aggregation," Bull. Math. Biol., vol. 68, no. 7, pp. 1601-1623, Sep. 2006.
[191] I. Aoki, "A simulation study on the schooling mechanism in fish.," Nippon SUISAN GAKKAISHI, vol. 48, no. 8, pp. 1081-1088, Aug. 1982.
[192] A. Huth and C. Wissel, "The simulation of the movement of fish schools," J. Theor. Biol., vol. 156, no. 3, pp. 365-385, Jun. 1992.
[193] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, "Novel Type of Phase Transition in a System of Self-Driven Particles," Phys. Rev. Lett., vol. 75, no. 6, pp. 1226-1229, Aug. 1995.
[194] G. Grégoire, H. Chaté, and Y. Tu, "Moving and staying together without a leader," Phys. D Nonlinear Phenom., vol. 181, no. 3-4, pp. 157-170, Jul. 2003.
[195] I. D. Couzin, J. Krause, R. James, G. D. Ruxton, and N. R. Franks, "Collective Memory and Spatial Sorting in Animal Groups," J. Theor. Biol., vol. 218, no. 1, pp. 1-11, Sep. 2002.
[196] A. Czirók, M. Vicsek, and T. Vicsek, "Collective motion of organisms in three dimensions," Phys. A Stat. Mech. its Appl., vol. 264, no. 1-2, pp. 299-304, Feb. 1999.
[197] A. Czirók and T. Vicsek, "Collective behavior of interacting self-propelled
particles," Phys. A Stat. Mech. its Appl., vol. 281, no. 1-4, pp. 17-29, Jun. 2000.
[198] J. K. J. K. Parrish and S. V. S. V. Viscido, "Traffic rules of fish schools: a review of agent-based approaches," in Self-organisation and the evolution of social behaviour, C. Hemelrijk, Ed. Cambridge University Press, 2005, pp. 50-80.
[199] D. Strömbom, "Collective motion from local attraction," J. Theor. Biol., vol. 283, no. 1, pp. 145-151, Aug. 2011.
[200] T. Vicsek and A. Zafeiris, "Collective motion," Phys. Rep., vol. 517, no. 3-4, pp. 71-140, Aug. 2012.
[201] C. W. Reynolds, "Flocks, herds and schools: A distributed behavioral model," in Proceedings of the 14th annual conference on Computer graphics and interactive techniques - SIGGRAPH '87, 1987, vol. 21, no. 4, pp. 25-34.
[202] E. Shaw, "Schooling Fishes: The school, a truly egalitarian form of organization in which all members of the group are alike in influence, offers substantial benefits to its participants," American Scientist, vol. 66. Sigma Xi, The Scientific Research Society, pp. 166-175.
[203] N. R. Franks, N. Gomez, S. Goss, and J. L. Deneubourg, "The blind leading the blind in army ant raid patterns: Testing a model of self-organization (Hymenoptera: Formicidae)," J. Insect Behav., vol. 4, no. 5, pp. 583-607, Sep. 1991.
[204] J. Kennedy and R. Eberhart, "Particle swarm optimization," in Proceedings of ICNN'95 - International Conference on Neural Networks, vol. 4, pp. 1942-1948.
[205] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in 1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360), pp. 69-73.
[206] A. Carlisle and G. Dozier, "Adapting particle swarm optimization to dynamic environments," Int. Conf. Artif. Intell., pp. 429-434, 2000.
[207] I. C. Trelea, "The particle swarm optimization algorithm: convergence analysis and parameter selection," Inf. Process. Lett., vol. 85, no. 6, pp. 317-325, Mar. 2003.
[208] D. Helbing and I. Farkas, "Simulation of pedestrian crowds in normal and evacuation situations," Pedestr. evacuation Dyn., vol. 21, no. January, pp. 21-58, 2002.
[209] M. Moussaid, D. Helbing, and G. Theraulaz, "How simple rules determine pedestrian behavior and crowd disasters," Proc. Natl. Acad. Sci., vol. 108, no. 17, pp. 6884-6888, Apr. 2011.
[210] A. Czirók, E. Ben-Jacob, I. Cohen, and T. Vicsek, "Formation of complex bacterial colonies via self-generated vortices," Phys. Rev. E, vol. 54, no. 2, pp. 1791-1801, Aug. 1996.
[211] A. Czirók, M. Matsushita, and T. Vicsek, "Theory of periodic swarming of bacteria: Application to Proteus mirabilis," Phys. Rev. E, vol. 63, no. 3, p. 031915, Feb. 2001.
[212] C. A. Yates et al., "Inherent noise can facilitate coherence in collective swarm motion.," Proc. Natl. Acad. Sci. U. S. A., vol. 106, no. 14, pp. 5464-9, Apr. 2009.
[213] N. W. F. Bode, D. W. Franks, and A. J. Wood, "Limited interactions in flocks: relating model simulations to empirical data.," J. R. Soc. Interface, vol. 8, no. 55, pp. 301-4, Feb. 2011.
[214] H. Hildenbrandt, C. Carere, and C. K. Hemelrijk, "Self-organized aerial displays of thousands of starlings: a model," Behav. Ecol., vol. 21, no. 6, pp. 1349-1359, Nov. 2010.
[215] P. Romanczuk, I. D. Couzin, and L. Schimansky-Geier, "Collective Motion due to Individual Escape and Pursuit Response," Phys. Rev. Lett., vol. 102, no. 1, p. 010602, Jan. 2009.
[216] M. Camperi, A. Cavagna, I. Giardina, G. Parisi, and E. Silvestri, "Spatially balanced topological interaction grants optimal cohesion in flocking models.," Interface Focus, vol. 2, no. 6, pp. 715-25, Dec. 2012.
[217] G. D. Pickett and M. G. Pawson, Deep-sea research. Part B, Oceanographic literature review., vol. 9, no. 42. Pergamon Press, 1979.
[218] D. Melton, "Stemness": Definitions, Criteria, and Standards, Third edit. Elsevier

Inc., 2014.
[219] L. Shahriyari and N. L. Komarova, "Symmetric vs. Asymmetric Stem Cell Divisions: An Adaptation against Cancer?," PLoS One, vol. 8, no. 10, p. e76195, Oct. 2013.
[220] K. Hardy, A. H. Handyside, and R. M. Winston, "The human blastocyst: cell number, death and allocation during late preimplantation development in vitro," Development, vol. 107, no. 3, 1989.
[221] M. L. Condic, "Totipotency: what it is and what it is not.," Stem Cells Dev., vol. 23, no. 8, pp. 796-812, Apr. 2014.
[222] M. K. Juric et al., "Milestones of Hematopoietic Stem Cell Transplantation From First Human Studies to Current Developments.," Front. Immunol., vol. 7, p. 470, 2016.
[223] NIH, "NIH Stem Cell Information Home Page," Stem Cell Information, 2016. [Online]. Available: https://stemcells.nih.gov/info/basics/7.htm. [Accessed: 13-Jun-2018].
[224] J. P. Guyette et al., "Bioengineering Human Myocardium on Native Extracellular Matrix.," Circ. Res., vol. 118, no. 1, pp. 56-72, Jan. 2016.
[225] M. Takasato et al., "Directing human embryonic stem cell differentiation towards a renal lineage generates a self-organizing kidney," Nat. Cell Biol., vol. 16, no. 1, pp. 118-126, Jan. 2014.
[226] B. R. Dye et al., "In vitro generation of human pluripotent stem cell derived lung organoids," Elife, vol. 4, p. e05098, Mar. 2015.
[227] G. Levy et al., "Long-term culture and expansion of primary human hepatocytes," Nat. Biotechnol., vol. 33, no. 12, pp. 1264-1271, Dec. 2015.
[228] Y. S. Chun, K. Byun, and B. Lee, "Induced pluripotent stem cells and personalized medicine: current progress and future perspectives.," Anat. Cell Biol., vol. 44, no. 4, pp. 245-55, Dec. 2011.
[229] K. Takahashi and S. Yamanaka, "Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors," Cell, vol.

126, no. 4, pp. 663-676, Aug. 2006.
[230] K. Okita, M. Nakagawa, H. Hyenjong, T. Ichisaka, and S. Yamanaka, "Generation of mouse induced pluripotent stem cells without viral vectors.," Science, vol. 322, no. 5903, pp. 949-53, Nov. 2008.
[231] Y. Shi, C. Desponts, J. T. Do, H. S. Hahm, H. R. Schöler, and S. Ding, "Induction of Pluripotent Stem Cells from Mouse Embryonic Fibroblasts by Oct4 and Klf4 with Small-Molecule Compounds," Cell Stem Cell, vol. 3, no. 5, pp. 568-574, Nov. 2008.
[232] D. Huangfu et al., "Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds," Nat. Biotechnol., vol. 26, no. 7, pp. 795-797, Jul. 2008.
[233] U. Ben-David and N. Benvenisty, "The tumorigenicity of human embryonic and induced pluripotent stem cells," Nat. Rev. Cancer, vol. 11, no. 4, pp. 268-277, Apr. 2011.
[234] H. Zhou et al., "Generation of induced pluripotent stem cells using recombinant proteins.," Cell Stem Cell, vol. 4, no. 5, pp. 381-4, May 2009.
[235] T. K. Ng, D. Pelaez, V. R. Fortino, J. Greenberg, and H. S. Cheung, "Pluripotent Adult Stem Cells: A Potential Revolution in Regenerative Medicine and Tissue Engineering," Pluripotent Stem Cells, Aug. 2013.
[236] A. J. Sloan and R. J. Waddington, "Dental pulp stem cells: what, where, how?," Int. J. Paediatr. Dent., vol. 19, no. 1, pp. 61-70, Jan. 2009.
[237] P. D. Telles, M. A. de A. M. Machado, V. T. Sakai, and J. E. Nör, "Pulp tissue from primary teeth: new source of stem cells.," J. Appl. Oral Sci., vol. 19, no. 3, pp. 189-94, 2011.
[238] S. Gronthos, M. Mankani, J. Brahim, P. G. Robey, and S. Shi, "Postnatal human dental pulp stem cells (DPSCs) in vitro and in vivo.," Proc. Natl. Acad. Sci. U. S. A., vol. 97, no. 25, pp. 13625-30, Dec. 2000.
[239] A. Graziano, R. d'Aquino, G. Laino, and G. Papaccio, "Dental Pulp Stem Cells: A Promising Tool for Bone Regeneration," Stem Cell Rev., vol. 4, no. 1, pp. 65-

65, Mar. 2008.
[240] I. V. Nosrat, J. Widenfalk, L. Olson, and C. A. Nosrat, "Dental Pulp Cells Produce Neurotrophic Factors, Interact with Trigeminal Neurons in Vitro, and Rescue Motoneurons after Spinal Cord Injury," Dev. Biol., vol. 238, no. 1, pp. 120-132, Oct. 2001.
[241] F. Young, A. Sloan, and B. Song, "Dental pulp stem cells and their potential roles in central nervous system regeneration and repair," J. Neurosci. Res., vol. 91, no. 11, pp. 1383-1393, Nov. 2013.
[242] J. A. Knoblich, "Mechanisms of Asymmetric Stem Cell Division," Cell, vol. 132, no. 4, pp. 583-597, Feb. 2008.
[243] I. G. Macara and S. Mili, "Polarity and Differential Inheritance-Universal Attributes of Life?," Cell, vol. 135, no. 5, Nov. 2008.
[244] A. Singhvi and G. Garriga, "Asymmetric divisions, aggresomes and apoptosis," Trends Cell Biol., vol. 19, no. 1, pp. 1-7, Jan. 2009.
[245] D. Gerlich, J. Mattes, and R. Eils, "Quantitative motion analysis and visualization of cellular structures," Methods, vol. 29, no. 1, pp. 3-13, Jan. 2003.
[246] M. Prasad and D. J. Montell, "Cellular and Molecular Mechanisms of Border Cell Migration Analyzed Using Time-Lapse Live-Cell Imaging," Dev. Cell, vol. 12, no. 6, pp. 997-1005, Jun. 2007.
[247] Y.-Y. Hsu, Y.-N. Liu, W.-W. Lu, and S.-H. Kung, "Visualizing and quantifying the differential cleavages of the eukaryotic translation initiation factors eIF4GI and eIF4GII in the enterovirus-infected cell," Biotechnol. Bioeng., vol. 104, no. 6, pp. 1142-1152, Dec. 2009.
[248] K. Hanaki et al., "Semiconductor quantum dot/albumin complex is a long-life and highly photostable endosome marker," Biochem. Biophys. Res. Commun., vol. 302, no. 3, pp. 496-501, Mar. 2003.
[249] A. J. Pretorius, I. A. Khan, and R. J. Errington, "A Survey of Visualization for Live Cell Imaging," Comput. Graph. Forum, vol. 36, no. 1, pp. 46-63, Jan. 2017.
[250] G. G. Stokes, "On the Change of Refrangibility of Light," Philos. Trans. R. Soc.

London, vol. 142, no. 0, pp. 463-562, Jan. 1852.
[251] R. W. Wood, "The Time Interval between Absorption and Emission of Light in Fluorescence," Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, vol. 99. Royal Society, pp. 362-371.
[252] J. R. Lakowicz, "Introduction to Fluorescence," in Principles of Fluorescence Spectroscopy, MA: Springer US, 2006, pp. 1-26.
[253] T. Karstens and K. Kobs, "Rhodamine B and rhodamine 101 as reference substances for fluorescence quantum yield measurements," J. Phys. Chem., vol. 84, no. 14, pp. 1871-1872, Jul. 1980.
[254] K. Rurack, "Fluorescence Quantum Yields: Methods of Determination and Standards," in Standardization and Quality Assurance in Fluorescence Measurements I, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 101-145.
[255] "Fluorophore Photobleaching Literature References," MicroscopyU, 2018. [Online]. Available: https://www.microscopyu.com/references/fluorophorephotobleaching. [Accessed: 29-Jun-2018].
[256] B. S. Hulka, T. C. Wilcosky, and J. D. Griffith, Biological markers in epidemiology. Oxford University Press, 1990.
[257] D. A. Shagin et al., "GFP-like Proteins as Ubiquitous Metazoan Superfamily: Evolution of Functional Features and Structural Complexity," Mol. Biol. Evol., vol. 21, no. 5, pp. 841-850, May 2004.
[258] D. Glenn, R. L. Caldwell, and M. J. Pakes, "Fluorescence in arthropoda informs ecological studies in anchialine crustaceans, Remipedia, and Atyidae," J. Crustac. Biol., vol. 33, no. 5, pp. 620-626, Jan. 2013.
[259] K. Andrews, S. M. Reed, and S. E. Masta, "Spiders fluoresce variably across many taxa.," Biol. Lett., vol. 3, no. 3, pp. 265-7, Jun. 2007.
[260] P. Vukusic and I. Hooper, "Directionally controlled fluorescence emission in butterflies.," Science, vol. 310, no. 5751, p. 1151, Nov. 2005.
[261] K. E. Arnold, I. P. F. Owens, N. J. Marshall, and N. J. Marshall, "Fluorescent
signaling in parrots.," Science, vol. 295, no. 5552, p. 92, Jan. 2002.
[262] F. Gandía-Herrero, F. García-Carmona, and J. Escribano, "Floral fluorescence effect," Nature, vol. 437, no. 7057, pp. 334-334, Sep. 2005.
[263] D. F. Gruber, H.-T. Kao, S. Janoschka, J. Tsai, and V. A. Pieribone, "Patterns of fluorescent protein expression in Scleractinian corals.," Biol. Bull., vol. 215, no. 2, pp. 143-54, Oct. 2008.
[264] M. V. Matz, N. J. Marshall, and M. Vorobyev, "Are Corals Colorful?," Photochem. Photobiol., vol. 82, no. 2, p. 345, 2006.
[265] D. M. Chudakov, M. V. Matz, S. Lukyanov, and K. A. Lukyanov, "Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues," Physiol. Rev., vol. 90, no. 3, pp. 1103-1163, Jul. 2010.
[266] F. Bou-Abdallah, N. D. Chasteen, and M. P. Lesser, "Quenching of superoxide radicals by green fluorescent protein," Biochim. Biophys. Acta - Gen. Subj., vol. 1760, no. 11, pp. 1690-1695, Nov. 2006.
[267] A. Salih, A. Larkum, G. Cox, M. Kühl, and O. Hoegh-Guldberg, "Fluorescent pigments in corals are photoprotective," Nature, vol. 408, no. 6814, pp. 850-853, Dec. 2000.
[268] O. Shimomura, F. H. Johnson, and Y. Saiga, "Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan,Aequorea," J. Cell. Comp. Physiol., vol. 59, no. 3, pp. 223-239, Jun. 1962.
[269] D. C. Prasher, V. K. Eckenrode, W. W. Ward, F. G. Prendergast, and M. J. Cormier, "Primary structure of the Aequorea victoria green-fluorescent protein," Gene, vol. 111, no. 2, pp. 229-233, Feb. 1992.
[270] K. A. Lukyanov, D. M. Chudakov, S. Lukyanov, and V. V. Verkhusha, "Photoactivatable fluorescent proteins," Nat. Rev. Mol. Cell Biol., vol. 6, no. 11, pp. 885-890, Nov. 2005.
[271] A. H. Coons, H. J. Creech, and R. N. Jones, "Immunological Properties of an Antibody Containing a Fluorescent Group.," Exp. Biol. Med., vol. 47, no. 2, pp.

200-202, Jun. 1941.
[272] A. M. Peters, "A brief history of cell labelling.," Q. J. Nucl. Med. Mol. Imaging, vol. 49, no. 4, pp. 304-7, Dec. 2005.
[273] J. E. Aubin, "Autofluorescence of viable cultured mammalian cells.," J. Histochem. Cytochem., vol. 27, no. 1, pp. 36-43, Jan. 1979.
[274] R. C. Benson, R. A. Meyer, M. E. Zaruba, and G. M. McKhann, "Cellular autofluorescence--is it due to flavins?," J. Histochem. Cytochem., vol. 27, no. 1, pp. 44-8, Jan. 1979.
[275] M. Monici, "Cell and tissue autofluorescence research and diagnostic applications," Biotechnol. Annu. Rev., vol. 11, pp. 227-256, Jan. 2005.
[276] M. Fritzsche and C.-F. Mandenius, "Fluorescent cell-based sensing approaches for toxicity testing," Anal. Bioanal. Chem., vol. 398, no. 1, pp. 181-191, Sep. 2010.
[277] M. R. Brown et al., "Long-term time series analysis of quantum dot encoded cells by deconvolution of the autofluorescence signal," Cytom. Part A, vol. 77A, no. 10, pp. 925-932, Oct. 2010.
[278] A. I. Ekimov and A. A. Onushchenko, "Quantum size effect in three-dimensional microscopic semi-conductor crystals," ZhETF Pis ma Redaktsiiu, vol. 34, p. 363, 1981.
[279] M. Bruchez, M. Moronne, P. Gin, S. Weiss, and A. P. Alivisatos, "Semiconductor nanocrystals as fluorescent biological labels.," Science, vol. 281, no. 5385, pp. 2013-6, Sep. 1998.
[280] C. B. Murray, D. J. Norris, and M. G. Bawendi, "Synthesis and Characterization of Nearly Monodisperse CdE ( $\mathrm{E}=\mathrm{S}, \mathrm{Se}, \mathrm{Te}$ ) Semiconductor Nanocrystallites," J. Am. Chem. Soc., vol. 115, no. 19, pp. 8706-8715, 1993.
[281] Lianhua Qu, A. Z. Adam Peng, and X. Peng, "Alternative Routes toward High Quality CdSe Nanocrystals," 2001.
[282] A. P. Alivisatos, "Perspectives on the Physical Chemistry of Semiconductor Nanocrystals," 1996.
[283] A. Imamoglu, "Are quantum dots useful for quantum computation?," Phys. E Low-dimensional Syst. Nanostructures, vol. 16, no. 1, pp. 47-50, Jan. 2003.
[284] A. J. Nozik, "Quantum dot solar cells," Phys. E Low-dimensional Syst. Nanostructures, vol. 14, no. 1-2, pp. 115-120, Apr. 2002.
[285] J. M. Caruge, J. E. Halpert, V. Wood, V. Bulović, and M. G. Bawendi, "Colloidal quantum-dot light-emitting diodes with metal-oxide charge transport layers," Nat. Photonics, vol. 2, no. 4, pp. 247-250, Apr. 2008.
[286] Y. Shirasaki, G. J. Supran, M. G. Bawendi, and V. Bulović, "Emergence of colloidal quantum-dot light-emitting technologies," Nat. Photonics, vol. 7, no. 1, pp. 13-23, Jan. 2013.
[287] M. A. Walling, J. A. Novak, and J. R. E. Shepard, "Quantum dots for live cell and in vivo imaging.," Int. J. Mol. Sci., vol. 10, no. 2, pp. 441-91, Feb. 2009.
[288] B. A. Rzigalinski and J. S. Strobl, "Cadmium-containing nanoparticles: Perspectives on pharmacology and toxicology of quantum dots," Toxicol. Appl. Pharmacol., vol. 238, no. 3, pp. 280-288, Aug. 2009.
[289] I. Moreels et al., "Composition and Size-Dependent Extinction Coefficient of Colloidal PbSe Quantum Dots," Chem. Mater., vol. 19, no. 25, pp. 6101-6106, Dec. 2007.
[290] Q. Xia, H.-M. Hwang, P. C. Ray, and H. Yu, "Mechanisms of nanotoxicity: Generation of reactive oxygen species," J. Food Drug Anal., vol. 22, no. 1, pp. 64-75, Mar. 2014.
[291] B. Ballou, B. C. Lagerholm, L. A. Ernst, M. P. Bruchez, and A. S. Waggoner, "Noninvasive Imaging of Quantum Dots in Mice," 2003.
[292] J. P. Ryman-Rasmussen, J. E. Riviere, and N. A. Monteiro-Riviere, "Surface Coatings Determine Cytotoxicity and Irritation Potential of Quantum Dot Nanoparticles in Epidermal Keratinocytes," J. Invest. Dermatol., vol. 127, no. 1, pp. 143-153, Jan. 2007.
[293] F. Chen and D. Gerion, "Fluorescent CdSe/ZnS Nanocrystal-Peptide Conjugates for Long-term, Nontoxic Imaging and Nuclear Targeting in Living Cells," 2004.
[294] J. L. Pelley, A. S. Daar, and M. A. Saner, "State of academic knowledge on toxicity and biological fate of quantum dots.," Toxicol. Sci., vol. 112, no. 2, pp. 276-96, Dec. 2009.
[295] D. V. Talapin, A. L. Rogach, A. Kornowski, M. Haase, and H. Weller, "Highly Luminescent Monodisperse CdSe and CdSe/ZnS Nanocrystals Synthesized in a Hexadecylamine-Trioctylphosphine Oxide-Trioctylphospine Mixture," 2001.
[296] B. O. Dabbousi et al., "(CdSe)ZnS Core-Shell Quantum Dots: Synthesis and Characterization of a Size Series of Highly Luminescent Nanocrystallites," 1997.
[297] M. A. Hines and P. Guyot-Sionnest, "Synthesis and Characterization of Strongly Luminescing ZnS-Capped CdSe Nanocrystals," J. Phys. Chem., vol. 100, no. 2, pp. 468-471, Jan. 1996.
[298] M. E. Akerman, W. C. W. Chan, P. Laakkonen, S. N. Bhatia, and E. Ruoslahti, "Nanocrystal targeting in vivo.," Proc. Natl. Acad. Sci. U. S. A., vol. 99, no. 20, pp. 12617-21, Oct. 2002.
[299] S. Pathak, M. C. Davidson, and G. A. Silva, "Characterization of the Functional Binding Properties of Antibody Conjugated Quantum Dots," 2007.
[300] "Guide to safe handling and disposal of manufactured nanomaterials," London, 2007.
[301] D. A. Navarro, S. Banerjee, D. F. Watson, and D. S. Aga, "Differences in Soil Mobility and Degradability between Water-Dispersible CdSe and $\mathrm{CdSe} / \mathrm{ZnS}$ Quantum Dots," Environ. Sci. Technol., vol. 45, no. 15, pp. 6343-6349, Aug. 2011.
[302] C. A. E. Hauser and S. Zhang, "Nanotechnology: Peptides as biological semiconductors," Nature, vol. 468, no. 7323, pp. 516-517, Nov. 2010.
[303] M. Anees, M. I. Masood, M. Ilyas, T. Hussain, and M. Ammad, "Nanoparticles as a novel drug delivery system: a review," Pakistan J. Pharm. Res., vol. 2, no. 2, pp. 160-167, 2016.
[304] W. C. Chan and S. Nie, "Quantum dot bioconjugates for ultrasensitive nonisotopic detection.," Science, vol. 281, no. 5385, pp. 2016-8, Sep. 1998.
[305] C. B. Murray, S. Sun, W. Gaschler, H. Doyle, T. A. Betley, and C. R. Kagan, "Colloidal synthesis of nanocrystals and nanocrystal superlattices," IBM J. Res. Dev., vol. 45, no. 1, pp. 47-56, Jan. 2001.
[306] M. V. Kovalenko et al., "SnTe Nanocrystals: A New Example of Narrow-Gap Semiconductor Quantum Dots," 2007.
[307] T. D. Lacoste, X. Michalet, F. Pinaud, D. S. Chemla, A. P. Alivisatos, and S. Weiss, "Ultrahigh-resolution multicolor colocalization of single fluorescent probes.," Proc. Natl. Acad. Sci. U. S. A., vol. 97, no. 17, pp. 9461-6, Aug. 2000.
[308] H. Kobayashi et al., "Simultaneous Multicolor Imaging of Five Different Lymphatic Basins Using Quantum Dots," 2007.
[309] W. Mayer, A. Niveleau, J. Walter, R. Fundele, and T. Haaf, "Demethylation of the zygotic paternal genome," Nature, vol. 403, no. 6769, pp. 501-502, Feb. 2000.
[310] A. E. Karnoub et al., "Mesenchymal stem cells within tumour stroma promote breast cancer metastasis," Nature, vol. 449, no. 7162, pp. 557-563, Oct. 2007.
[311] R. J. Errington et al., "Single cell nanoparticle tracking to model cell cycle dynamics and compartmental inheritance," Cell Cycle, vol. 9, no. 1, pp. 121-130, Jan. 2010.
[312] M. Cremer et al., "Multicolor 3D Fluorescence In Situ Hybridization for Imaging Interphase Chromosomes," Humana Press, Totowa, NJ, 2012, pp. 205-239.
[313] X. Gao, Y. Cui, R. M. Levenson, L. W. K. Chung, and S. Nie, "In vivo cancer targeting and imaging with semiconductor quantum dots," Nat. Biotechnol., vol. 22, no. 8, pp. 969-976, Aug. 2004.
[314] H. Shapiro, Practical Flow Cytometry. Hoboken, NJ, USA, NJ, USA: John Wiley \& Sons, Inc., 2003.
[315] I. Schmid, C. H. Uittenbogaart, and J. V. Giorgi, "A gentle fixation and permeabilization method for combined cell surface and intracellular staining with improved precision in DNA quantification," Cytometry, vol. 12, no. 3, pp. 279285, 1991.
[316] E. A. Jares-Erijman and T. M. Jovin, "FRET imaging," Nat. Biotechnol., vol. 21, no. 11, pp. 1387-1395, Nov. 2003.
[317] K. Truong and M. Ikura, "The use of FRET imaging microscopy to detect protein-protein interactions and protein conformational changes in vivo.," Curr. Opin. Struct. Biol., vol. 11, no. 5, pp. 573-8, Oct. 2001.
[318] G. M. van Dam et al., "Intraoperative tumor-specific fluorescence imaging in ovarian cancer by folate receptor- $\alpha$ targeting: first in-human results," Nat. Med., vol. 17, no. 10, pp. 1315-1319, Oct. 2011.
[319] M. Baker, "Cellular imaging: Taking a long, hard look," Nature, vol. 466, no. 7310, pp. 1137-1138, Aug. 2010.
[320] R. Cole, "Live-cell imaging: The cell's perspective," Cell Adh. Migr., vol. 8, no. 5, p. 452, 2014.
[321] M. M. Frigault, J. Lacoste, J. L. Swift, and C. M. Brown, "Live-cell microscopy tips and tools.," J. Cell Sci., vol. 122, no. Pt 6, pp. 753-67, Mar. 2009.
[322] Lodish et al., "Growing and Studying Cells in Culture," in Molecular Cell Biology, 8th ed., New York, New York, USA, New York, USA: Macmillan Learning, 2016, pp. 129-138.
[323] M. A. Taylor et al., "Biological measurement beyond the quantum limit," Nat. Photonics, vol. 7, no. 3, pp. 229-233, Mar. 2013.
[324] L. Song, E. J. Hennink, I. T. Young, and H. J. Tanke, "Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy," Biophys. J., vol. 68, no. 6, pp. 2588-2600, Jun. 1995.
[325] G. H. Patterson and D. W. Piston, "Photobleaching in Two-Photon Excitation Microscopy," Biophys. J., vol. 78, no. 4, pp. 2159-2162, Apr. 2000.
[326] Wikipedia, "Plate Reader," Wikipedia, 2018. [Online]. Available: https://en.wikipedia.org/wiki/Plate_reader. [Accessed: 20-Jun-2018].
[327] H. Peng, "Bioimage informatics: a new area of engineering biology," Bioinformatics, vol. 24, no. 17, pp. 1827-1836, Sep. 2008.
[328] G. McConnell, J. Trägårdh, R. Amor, J. Dempster, E. Reid, and W. B. Amos, "A
novel optical microscope for imaging large embryos and tissue volumes with sub-cellular resolution throughout," Elife, vol. 5, Sep. 2016.
[329] E. Meijering, O. Dzyubachyk, and I. Smal, "Methods for Cell and Particle Tracking," Methods Enzymol., vol. 504, pp. 183-200, Jan. 2012.
[330] I. Seroussi, D. Veikherman, N. Ofer, S. Yehudai-Resheff, and K. Keren, "Segmentation and tracking of live cells in phase-contrast images using directional gradient vector flow for snakes," J. Microsc., vol. 247, no. 2, pp. 137146, Aug. 2012.
[331] Lodish et al., "Light Microscopy: Exploring Cell Structure and Visualising Proteins within Cells," in Molecular Cell Biology, 8th ed., New York, New York, USA, New York, USA: Macmillan Learning, 2016, pp. 139-156.
[332] D. Boal, Size Limits of Very Small Microorganisms : Proceedings of a Workshop. National Academy Press, 1999.
[333] J. Huotari and A. Helenius, "Endosome maturation.," EMBO J., vol. 30, no. 17, pp. 3481-500, Aug. 2011.
[334] N. Y. Ashri, S. A. Ajlan, and A. M. Aldahmash, "Dental pulp stem cells. Biology and use for periodontal tissue engineering.," Saudi Med. J., vol. 36, no. 12, pp. 1391-9, Dec. 2015.
[335] F. Zernike, "Phase contrast, a new method for the microscopic observation of transparent objects part II," Physica, vol. 9, no. 10, pp. 974-986, Dec. 1942.
[336] J. Icha, M. Weber, J. C. Waters, and C. Norden, "Phototoxicity in live fluorescence microscopy, and how to avoid it," BioEssays, vol. 39, no. 8, p. 1700003, Aug. 2017.
[337] E. C. Jensen, "Overview of Live-Cell Imaging: Requirements and Methods Used," Anat. Rec. Adv. Integr. Anat. Evol. Biol., vol. 296, no. 1, pp. 1-8, Jan. 2013.
[338] L. L. Vindeløv, I. J. Christensen, and N. I. Nissen, "A Detergent-trypsin method for the preparation of nuclei for flow cytometric DNA analysis," Cytometry, vol. 3, no. 5, pp. 323-327, Mar. 1983.
[339] S. Nomura, H. Nagata, K. Oda, T. Kokawa, and K. Yasunaga, "Effects of EDTA on the membrane glycoproteins IIb-IIIa complex--analysis using flow cytometry.," Thromb. Res., vol. 47, no. 1, pp. 47-58, Jul. 1987.
[340] M. Kaur and L. Esau, "Two-step protocol for preparing adherent cells for highthroughput flow cytometry," Biotechniques, vol. 59, no. 3, pp. 119-126, 2015.
[341] P. P. Austin Suthanthiraraj and S. W. Graves, "Fluidics.," Curr. Protoc. Cytom., vol. Chapter 1, p. Unit 1.2, Jul. 2013.
[342] "Fluidics of a Flow Cytometer," ThermoFisher scientific. [Online]. Available: https://www.thermofisher.com/uk/en/home/life-science/cell-analysis/cell-analysis-learning-center/molecular-probes-school-of-fluorescence/flow-cytometry-basics/flow-cytometry-fundamentals/fluidics-flow-cytometer.html. [Accessed: 07-Jul-2018].
[343] A. Longobardi Givan, Flow Cytometry First Principles. John Wiley \& Sons, 2013.
[344] H. Schlichting, K. Gersten, and K. Mayes, Boundary-layer theory. .
[345] M. Ward, P. Turner, M. DeJohn, and G. Kaduchak, "Fundamentals of Acoustic Cytometry," Curr. Protoc. Cytom., vol. 49, no. 1, p. 1.22.1-1.22.12, Jul. 2009.
[346] H. Yoneda et al., "Atomic inner-shell laser at 1.5-ångström wavelength pumped by an X-ray free-electron laser," Nature, vol. 524, no. 7566, pp. 446-449, Aug. 2015.
[347] H. M. Goldenberg, D. Kleppner, and N. F. Ramsey, "Atomic Hydrogen Maser," Phys. Rev. Lett., vol. 5, no. 8, pp. 361-362, Oct. 1960.
[348] V. M. Kuzmichev and S. N. Polhilko, "Measurement of the generalized laser beam cross section area," in Third International Kharkov Symposium "Physics and Engineering of Millimeter and Submillimeter Waves". MSMW'98. Symposium Proceedings (Cat. No.98EX119), vol. 2, pp. 722-724.
[349] A. Tzur, J. K. Moore, P. Jorgensen, H. M. Shapiro, and M. W. Kirschner, "Optimizing optical flow cytometry for cell volume-based sorting and analysis.," PLoS One, vol. 6, no. 1, p. e16053, Jan. 2011.
[350] M. Birtwistle, "Lecture 1 - Flow Cytometry 1 - Flow and Mass Cytometry for Single Cell Protein Levels and Cell Fate | Coursera," Coursera, Experimental Methods in Systems Biology, 2014. [Online]. Available: https://www.coursera.org/lecture/experimental-methods/lecture-1-flow-cytometry-1-paVmA?authMode=signup. [Accessed: 08-Jul-2018].
[351] J. M. Scherer, W. Stillwell, and L. J. Jenski, "Anomalous changes in forward scatter of lymphocytes with loosely packed membranes," Cytometry, vol. 37, no. 3, pp. 184-190, Nov. 1999.
[352] "Flow cytometry Core Facility," 2010. [Online]. Available: https://www.ess.uci.edu/group/fcm/home. [Accessed: 08-Jul-2018].
[353] W. H. Coulter, "High speed automatic blood cell counter and cell size analyzer," Proc. Natl. Electron. Conf., vol. 12, 1956.
[354] M. H. Julius, T. Masuda, and L. A. Herzenberg, "Demonstration that antigenbinding cells are precursors of antibody-producing cells after purification with a fluorescence-activated cell sorter.," Proc. Natl. Acad. Sci. U. S. A., vol. 69, no. 7, pp. 1934-8, Jul. 1972.
[355] G. Leslie, "Flow Cytometric Cell Sorting: A Basic Guide," https://www.sdu.dk/. [Online]. Available: https://www.sdu.dk//media/files/om_sdu/institutter/imb/flow/flow+cytometric+cell+sortingbasic+gui de.pdf. [Accessed: 09-Jul-2018].
[356] J. Picot, C. L. Guerin, C. Le Van Kim, and C. M. Boulanger, "Flow cytometry: retrospective, fundamentals and recent instrumentation.," Cytotechnology, vol. 64, no. 2, pp. 109-30, Mar. 2012.
[357] G. Brecher, E. F. Jakobek, M. A. Schneiderman, G. Z. Williams, and P. J. Schmidt, "Size Distribution Of Erythrocytes," Ann. N. Y. Acad. Sci., vol. 99, no. 2, pp. 242-261, Dec. 1962.
[358] M. J. Fulwyler, "Electronic separation of biological cells by volume.," Science, vol. 150, no. 3698, pp. 910-1, Nov. 1965.
[359] L. A. Johnson, G. R. Welch, and W. Rens, "The Beltsville sperm sexing technology: high-speed sperm sorting gives improved sperm output for in vitro
fertilization and AI," J. Anim. Sci., vol. 77, no. suppl_2, p. 213, Jan. 1999.
[360] D. Rath et al., "Sex selection of sperm in farm animals: status report and developmental prospects.," Reproduction, vol. 145, no. 1, pp. R15-30, Jan. 2013.
[361] M. Darnovsky, "Countries with laws or policies on sex selection," Cent. Genet. Soc., 2009.
[362] P. X. Petit, N. Glab, D. Marie, H. Kieffer, and P. Métézeau, "Discrimination of respiratory dysfunction in yeast mutants by confocal microscopy, image, and flow cytometry," Cytometry, vol. 23, no. 1, pp. 28-38, Jan. 1996.
[363] M. Valli, M. Sauer, P. Branduardi, N. Borth, D. Porro, and D. Mattanovich, "Improvement of Lactic Acid Production in Saccharomyces cerevisiae by Cell Sorting for High Intracellular pH," Appl. Environ. Microbiol., vol. 72, no. 8, pp. 5492-5499, Aug. 2006.
[364] E. Böhm, R. Voglauer, W. Steinfellner, R. Kunert, N. Borth, and H. Katinger, "Screening for improved cell performance: Selection of subclones with altered production kinetics or improved stability by cell sorting," Biotechnol. Bioeng., vol. 88, no. 6, pp. 699-706, Dec. 2004.
[365] D. Mattanovich and N. Borth, "Applications of cell sorting in biotechnology.," Microb. Cell Fact., vol. 5, p. 12, Mar. 2006.
[366] P. O. Krutzik and G. P. Nolan, "Fluorescent cell barcoding in flow cytometry allows high-throughput drug screening and signaling profiling," Nat. Methods, vol. 3, no. 5, pp. 361-368, 2006.
[367] A. M. Lahesmaa Korpinen et al., "FlowAnd: Comprehensive Computational Framework for Flow Cytometry Data Analysis," J. Proteomics Bioinform., vol. 4, no. 11, Nov. 2011.
[368] D. Wu, B. L. Wood, and J. R. Fromm, "Flow Cytometry for Non-Hodgkin and Classical Hodgkin Lymphoma," Humana Press, Totowa, NJ, 2013, pp. 27-47.
[369] C. S. Ferreira-Facio et al., "Contribution of multiparameter flow cytometry immunophenotyping to the diagnostic screening and classification of pediatric cancer.," PLoS One, vol. 8, no. 3, p. e55534, 2013.
[370] B. Brando et al., "Cytofluorometric methods for assessing absolute numbers of cell subsets in blood," Cytometry, vol. 42, no. 6, pp. 327-346, Dec. 2000.
[371] Z. Darzynkiewicz et al., "Features of apoptotic cells measured by flow cytometry," Cytometry, vol. 13, no. 8, pp. 795-808, 1992.
[372] I. Nicoletti, G. Migliorati, M. C. Pagliacci, F. Grignani, and C. Riccardi, "A rapid and simple method for measuring thymocyte apoptosis by propidium iodide staining and flow cytometry," J. Immunol. Methods, vol. 139, no. 2, pp. 271-279, Jun. 1991.
[373] D. W. Galbraith, K. R. Harkins, J. M. Maddox, N. M. Ayres, D. P. Sharma, and E. Firoozabady, "Rapid flow cytometric analysis of the cell cycle in intact plant tissues.," Science, vol. 220, no. 4601, pp. 1049-51, Jun. 1983.
[374] R. Nunez, "DNA measurement and cell cycle analysis by flow cytometry," Curr. Issues Mol. Biol., vol. 3, no. 3, pp. 67-70, 2001.
[375] Z. Darzynkiewicz, F. Traganos, and M. R. Melamed, "New cell cycle compartments identified by multiparameter flow cytometry," Cytometry, vol. 1, no. 2, pp. 98-108, Sep. 1980.
[376] M. A. Hoyt, L. Totis, and B. T. Roberts, "S. cerevisiae genes required for cell cycle arrest in response to loss of microtubule function," Cell, vol. 66, no. 3, pp. 507-517, Aug. 1991.
[377] J. Bender et al., "Identification and comparison of CD34-positive cells and their subpopulations from normal peripheral blood and bone marrow using multicolor flow cytometry," Blood, vol. 77, no. 12, 1991.
[378] M. J. Jaroszeski, R. Gilbert, and R. Heller, "Detection and Quantitation of CellCell Electrofusion Products by Flow Cytometry," Anal. Biochem., vol. 216, no. 2, pp. 271-275, Feb. 1994.
[379] K. Ben Amor et al., "Multiparametric flow cytometry and cell sorting for the assessment of viable, injured, and dead bifidobacterium cells during bile salt stress.," Appl. Environ. Microbiol., vol. 68, no. 11, pp. 5209-16, Nov. 2002.
[380] W. Kern et al., "Detection of minimal residual disease in unselected patients with
acute myeloid leukemia using multiparameter flow cytometry for definition of leukemia-associated immunophenotypes and determination of their frequencies in normal bone marrow.," Haematologica, vol. 88, no. 6, pp. 646-53, Jun. 2003.
[381] K. Matsumoto et al., "Preferential target is mitochondria in $\alpha$-mangostin-induced apoptosis in human leukemia HL60 cells," Bioorg. Med. Chem., vol. 12, no. 22, pp. 5799-5806, Nov. 2004.
[382] F. Hammes, M. Berney, Y. Wang, M. Vital, O. Köster, and T. Egli, "Flowcytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes," Water Res., vol. 42, no. 1-2, pp. 269277, Jan. 2008.
[383] I. A. Khan et al., "Interoperability of time series cytometric data: A cross platform approach for modeling tumor heterogeneity," Cytom. Part A, vol. 79A, no. 3, pp. 214-226, Mar. 2011.
[384] H. D. Summers et al., "Quantification of Nanoparticle Dose and Vesicular Inheritance in Proliferating Cells," ACS Nano, vol. 7, no. 7, pp. 6129-6137, Jul. 2013.
[385] J. S. Colombo, R. A. Howard-Jones, F. I. Young, R. J. Waddington, R. J. Errington, and A. J. Sloan, "A 3D ex vivo mandible slice system for longitudinal culturing of transplanted dental pulp progenitor cells," Cytom. Part A, vol. 87, no. 10, pp. 921-928, Oct. 2015.
[386] G. P. Dowthwaite et al., "The surface of articular cartilage contains a progenitor cell population," J. Cell Sci., vol. 117, no. 6, pp. 889-897, Feb. 2004.
[387] "Qtracker 705 Cell Labeling Kit," Thermo Fisher Scientific. [Online]. Available: https://www.thermofisher.com/order/catalog/product/Q25061MP. [Accessed: 18-Jul-2018].
[388] N. Schmidt, A. Mishra, G. H. Lai, and G. C. L. Wong, "Arginine-rich cellpenetrating peptides," FEBS Lett., vol. 584, no. 9, pp. 1806-1813, May 2010.
[389] A. Bevilacqua, F. Piccinini, and A. Gherardi, "Vignetting correction by exploiting an optical microscopy image sequence," in 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011, vol.

2011, pp. 6166-6169.
[390] N. S. WHITE, R. J. ERRINGTON, M. D. FRICKER, and J. L. WOOD, "Aberration control in quantitative imaging of botanical specimens by multidimensional fluorescence microscopy," J. Microsc., vol. 181, no. 2, pp. 99116, Feb. 1996.
[391] I. T. Young, "Shading Correction: Compensation for Illumination and Sensor Inhomogeneities," in Current Protocols in Cytometry, no. 1, Hoboken, NJ, USA: John Wiley \& Sons, Inc., 2001.
[392] M. Schwarzfischer, C. Marr, J. Krumsiek, P. S. Hoppe, T. Schroeder, and F. J. Theis, "Efficient fluorescence image normalization for time lapse movies," Proc. Microsc. Image Anal. with Appl. Biol., vol. 5, pp. 1-5, 2011.
[393] T. R. Jones, A. E. Carpenter, D. M. Sabatini, and P. Golland, "Methods for highcontent, high-throughput image-based cell screening," Proc. First MICCAI Work. Microsc. Image Anal. with Appl. Biol., pp. 65-72, 2006.
[394] M. A. Model and J. K. Burkhardt, "A standard for calibration and shading correction of a fluorescence microscope," Cytometry, vol. 44, no. 4, pp. 309-316, Aug. 2001.
[395] W. Yu, "Practical anti-vignetting methods for digital cameras," IEEE Trans. Consum. Electron., vol. 50, no. 4, pp. 975-983, Nov. 2004.
[396] D. S. GAREAU et al., "Confocal mosaicing microscopy in skin excisions: a demonstration of rapid surgical pathology," J. Microsc., vol. 233, no. 1, pp. 149159, Jan. 2009.
[397] C. Baravalle, N. R. Salvetti, G. A. Mira, N. Pezzone, and H. H. Ortega, "Microscopic Characterization of Follicular Structures in Letrozole-induced Polycystic Ovarian Syndrome in the Rat," Arch. Med. Res., vol. 37, no. 7, pp. 830-839, Oct. 2006.
[398] Yuanjie Zheng, Jingyi Yu, Sing Bing Kang, S. Lin, and C. Kambhamettu, "Single-image vignetting correction using radial gradient symmetry," in 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1-8.
[399] Yuanjie Zheng, S. Lin, C. Kambhamettu, Jingyi Yu, and Sing Bing Kang, "Single-Image Vignetting Correction," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 12, pp. 2243-2256, Dec. 2009.
[400] K. He, P.-F. Tang, and R. Liang, "Vignetting Image Correction Based on Gaussian Quadrics Fitting," in 2009 Fifth International Conference on Natural Computation, 2009, pp. 158-161.
[401] F. J. W.-M. Leong, M. Brady, and J. O. McGee, "Correction of uneven illumination (vignetting) in digital microscopy images.," J. Clin. Pathol., vol. 56, no. 8, pp. 619-21, Aug. 2003.
[402] D. Bradley and G. Roth, "Adaptive Thresholding using the Integral Image," $J$. Graph. Tools, vol. 12, no. 2, pp. 13-21, Jan. 2007.
[403] E. Meijering, O. Dzyubachyk, I. Smal, and W. A. van Cappellen, "Tracking in cell and developmental biology," Semin. Cell Dev. Biol., vol. 20, no. 8, pp. 894902, Oct. 2009.
[404] Z. Bao, J. I. Murray, T. Boyle, S. L. Ooi, M. J. Sandel, and R. H. Waterston, "Automated cell lineage tracing in Caenorhabditis elegans.," Proc. Natl. Acad. Sci. U. S. A., vol. 103, no. 8, pp. 2707-12, Feb. 2006.
[405] G. Lin et al., "A multi-model approach to simultaneous segmentation and classification of heterogeneous populations of cell nuclei in 3D confocal microscope images," Cytom. Part A, vol. 71A, no. 9, pp. 724-736, Sep. 2007.
[406] N. N. Kachouie et al., "Probabilistic Model-Based Cell Tracking," Int. J. Biomed. Imaging, vol. 2006, pp. 1-10, Jul. 2006.
[407] Kang Li, E. D. Miller, L. E. Weiss, P. G. Campbell, and T. Kanade, "Online Tracking of Migrating and Proliferating Cells Imaged with Phase-Contrast Microscopy," in 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06), 2006, pp. 65-65.
[408] C. Zimmer, E. Labruyere, V. Meas-Yedid, N. Guillen, and J. Olivo-Marin, "Segmentation and tracking of migrating cells in videomicroscopy with parametric active contours: a tool for cell-based drug testing," IEEE Trans. Med. Imaging, vol. 21, no. 10, pp. 1212-1221, Oct. 2002.
[409] Chipman, Cell-iq analyser manual. Chip-Man Technologies Ltd, 2010.
[410] M. R. Brown, N. Hondow, R. Brydson, P. Rees, A. P. Brown, and H. D. Summers, "Statistical prediction of nanoparticle delivery: From culture media to cell," Nanotechnology, vol. 26, no. 15, 2015.
[411] F. Yang, M. A. Mackey, F. Ianzini, G. Gallardo, and M. Sonka, "Cell Segmentation, Tracking, and Mitosis Detection Using Temporal Context," Springer, Berlin, Heidelberg, 2005, pp. 302-309.
[412] R. A. Rensink, "Change Detection," Annu. Rev. Psychol., vol. 53, no. 1, pp. 245277, Feb. 2002.
[413] T. Bergeland, J. Widerberg, O. Bakke, and T. W. Nordeng, "Mitotic partitioning of endosomes and lysosomes," Curr. Biol., vol. 11, no. 9, pp. 644-651, May 2001.
[414] Y. H. Sun et al., "Photostability and pH sensitivity of $\mathrm{CdSe} / \mathrm{ZnSe} / \mathrm{ZnS}$ quantum dots in living cells," Nanotechnology, vol. 17, no. 17, pp. 4469-4476, Sep. 2006.
[415] H. D. Summers et al., "Cell-population tracking using quantum dots in flow cytometry," 2008, vol. 6865, p. 68650L.
[416] A. Higuchi, H. Ling-Yi, L. Y. Huang, H. Chen, Y.-J. Chen, and Q.-D. Ling, "Measurements of Movement and Diffusion Coefficients of Single Cells on Polymeric Surface from Image Analysis," J. Biomater. Sci. Polym. Ed., vol. 21, no. 12, pp. 1545-1558, Jan. 2010.
[417] E. Hellinger, "Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen," J. für die reine und Angew. Math., vol. 136, pp. 210-271, 1909.
[418] Sung-Hyuk Cha, "Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions," Int. J. Math. Model. METHODS Appl. Sci., vol. 1, no. 4, pp. 300-307, 2007.
[419] E. J. Gumbel, Statistics of extremes. Dover Publications, 2004.
[420] G. T.-J. Huang, S. Gronthos, and S. Shi, "Mesenchymal stem cells derived from dental tissues vs. those from other sources: their biology and role in regenerative
medicine.," J. Dent. Res., vol. 88, no. 9, pp. 792-806, Sep. 2009.
[421] R. Storn and K. Price, "Differential Evolution - A Simple and Efficient Heuristic for global Optimization over Continuous Spaces," J. Glob. Optim., vol. 11, no. 4, pp. 341-359, 1997.
[422] R. Storn, "On the usage of differential evolution for function optimization," in Proceedings of North American Fuzzy Information Processing, 1996, pp. 519523.
[423] K. Price, R. Storn, and J. Lampinen, "Benchmarking Differential Evolution," in Differential Evolution, Berlin/Heidelberg: Springer-Verlag, 2005, pp. 135-187.
[424] K. Zielinski, P. Weitkemper, R. Laur, and K.-D. Kammeyer, "Examination of stopping criteria for differential evolution based on a power allocation problem," Int. Conf. Optim. Electr. Electron. Equip., vol. 3, pp. 149-156, 2006.
[425] S. Gronthos et al., "Molecular and cellular characterisation of highly purified stromal stem cells derived from human bone marrow.," J. Cell Sci., vol. 116, no. Pt 9, pp. 1827-35, May 2003.
[426] K. Dunster, B. H. Toh, and W. Sentry, John, "Early endosomes , late endosomes , and lysosomes display distinct partitioning strategies of inheritance with similarities to Golgi-derived membranes," 2002.
[427] Y. M. Yamashita and M. T. Fuller, "Asymmetric centrosome behavior and the mechanisms of stem cell division.," J. Cell Biol., vol. 180, no. 2, pp. 261-6, Jan. 2008.
[428] B. Giebel and J. Beckmann, "Asymmetric Cell Divisions of Human Hematopoietic Stem and Progenitor Cells Meet Endosomes," Cell Cycle, vol. 6, no. 18, pp. 2201-2204, Sep. 2007.
[429] M. A. Rujano et al., "Polarised Asymmetric Inheritance of Accumulated Protein Damage in Higher Eukaryotes," PLoS Biol., vol. 4, no. 12, p. e417, Dec. 2006.
[430] A. Wodarz and C. Gonzalez, "Connecting Cancer to the Asymmetric Division of Stem Cells," Cell, vol. 124, no. 6, pp. 1121-1123, Mar. 2006.
[431] K. J. Cha et al., "Cell density-dependent differential proliferation of neural stem
cells on omnidirectional nanopore-arrayed surface," Sci. Rep., vol. 7, no. 1, p. 13077, Dec. 2017.
[432] M. M. Najafabadi, V. Bayati, M. Orazizadeh, M. Hashemitabar, and F. Absalan, "Impact of Cell Density on Differentiation Efficiency of Rat Adipose-derived Stem Cells into Schwann-like Cells.," Int. J. stem cells, vol. 9, no. 2, pp. 213220, Nov. 2016.
[433] J. Morgan and E. Anders, "Chemical composition of Earth, Venus, and Mercury.," Proc. Natl. Acad. Sci. U. S. A., vol. 77, no. 12, pp. 6973-7, Dec. 1980.
[434] B. Marty, "The origins and concentrations of water, carbon, nitrogen and noble gases on Earth," Earth Planet. Sci. Lett., vol. 313-314, pp. 56-66, Jan. 2012.
[435] Y. Zhang and A. Zindler, "Distribution and evolution of carbon and nitrogen in Earth," Earth Planet. Sci. Lett., vol. 117, no. 3-4, pp. 331-345, Jun. 1993.
[436] J. M. Hunt, "Distribution of Carbon in Crust of Earth: Geological Notes," Am. Assoc. Pet. Geol. Bull., vol. 56, no. 11, pp. 2273-2277, 1974.
[437] D. J. Des Marais, H. Strauss, R. E. Summons, and J. M. Hayes, "Carbon isotope evidence for the stepwise oxidation of the Proterozoic environment," Nature, vol. 359, no. 6396, pp. 605-609, Oct. 1992.
[438] Y.-H. Li, "Geochemical mass balance among lithosphere, hydrosphere, and atmosphere," Am. J. Sci., vol. 272, no. 2, pp. 119-137, Feb. 1972.
[439] J. S. Lipp, Y. Morono, F. Inagaki, and K.-U. Hinrichs, "Significant contribution of Archaea to extant biomass in marine subsurface sediments," Nature, vol. 454, no. 7207, pp. 991-994, Aug. 2008.
[440] P. Falkowski et al., "The global carbon cycle: a test of our knowledge of earth as a system.," Science, vol. 290, no. 5490, pp. 291-6, Oct. 2000.
[441] "Deep Carbon Science," Deep Carbon Observatory, 2018. [Online]. Available: https://deepcarbon.net/dco/deep-carbon-science. [Accessed: 06-Sep-2018].
[442] D. M. Kerrick, "Present and past nonanthropogenic $\mathrm{CO}_{2}$ degassing from the solid earth," Rev. Geophys., vol. 39, no. 4, pp. 565-585, Nov. 2001.
[443] C. Di-Giovanni, J. R. Disnar, and J. J. Macaire, "Estimation of the annual yield of organic carbon released from carbonates and shales by chemical weathering," Glob. Planet. Change, vol. 32, no. 2-3, pp. 195-210, Apr. 2002.
[444] Y. Copard, P. Amiotte-Suchet, and C. Di-Giovanni, "Storage and release of fossil organic carbon related to weathering of sedimentary rocks," Earth Planet. Sci. Lett., vol. 258, no. 1-2, pp. 345-357, Jun. 2007.
[445] J. C. Alt et al., "Recycling of water, carbon, and sulfur during subduction of serpentinites: A stable isotope study of Cerro del Almirez, Spain," Earth Planet. Sci. Lett., vol. 327-328, pp. 50-60, Apr. 2012.
[446] H. H. Schobert, Chemistry of fossil fuels and biofuels. Cambridge [England] ;;New York: Cambridge University Press, 2013.
[447] R. Dasgupta and M. M. Hirschmann, "The deep carbon cycle and melting in Earth's interior," Earth Planet. Sci. Lett., vol. 298, no. 1-2, pp. 1-13, Sep. 2010.
[448] J. S. Olson, J. A. Watts, and L. J. Allison, "Carbon in Live Vegetation of Major World Ecosystems," Springfield, 1983.
[449] J. Roy, B. Saugier, and H. A. Mooney, Terrestrial global productivity. .
[450] W. H. Schlesinger and E. S. Bernhardt, Biogeochemistry: an analysis of global change. Academic Press, 2013.
[451] S. Kellomäki and T. Karjalainen, "Sequestration of carbon in the Finnish boreal forest ecosystem managed for timber production," in Forest Ecosystems, Forest Management and the Global Carbon Cycle, Heidelberg: Springer Berlin Heidelberg, 1996, pp. 59-68.
[452] C. B. Niel, "On the morphology and physiology of the purple and green sulphur bacteria," Arch. Mikrobiol., vol. 3, no. 1, pp. 1-112, 1932.
[453] E. L. Singsaas, D. R. Ort, and E. H. DeLucia, "Variation in measured values of photosynthetic quantum yield in ecophysiological studies," Oecologia, vol. 128, no. 1, pp. 15-23, Jun. 2001.
[454] S. R. Mishra, Photosynthesis in plants. Discovery Publishing House, 2004.
[455] J. F. Farrar, "The respiratory source of CO2," Plant, Cell Environ., vol. 8, no. 6,
pp. 427-438, Aug. 1985.
[456] A. Münster, Classical Thermodynamics. London, 1970.
[457] F. Horne, "How are seashells created? Or any other shell, such as a snail's or a turtle's?," Scientific American, Oct-2006.
[458] J.-P. Gattuso, M. Frankignoulle, I. Bourge, S. Romaine, and R. W. Buddemeier, "Effect of calcium carbonate saturation of seawater on coral calcification," Glob. Planet. Change, vol. 18, no. 1-2, pp. 37-46, Jul. 1998.
[459] J. D. Milliman, "Production and accumulation of calcium carbonate in the ocean: Budget of a nonsteady state," Global Biogeochem. Cycles, vol. 7, no. 4, pp. 927957, Dec. 1993.
[460] P. D. Quay, C. Peacock, K. Björkman, and D. M. Karl, "Measuring primary production rates in the ocean: Enigmatic results between incubation and nonincubation methods at Station ALOHA," Global Biogeochem. Cycles, vol. 24, no. 3, p. n/a-n/a, Sep. 2010.
[461] W. S. Broecker et al., "Fate of fossil fuel carbon dioxide and the global carbon budget.," Science, vol. 206, no. 4417, pp. 409-18, Oct. 1979.
[462] "Frequently Asked Global Change Questions," Carbon Dioxide Information Analysis Center, 2015. [Online]. Available: http://cdiac.essdive.lbl.gov/faq.html\#Q6. [Accessed: 12-Sep-2018].
[463] J. Perlin, A forest journey: the story of wood and civilization. .
[464] C. V. Cole, K. Flach, J. Lee, D. Sauerbeck, and B. Stewart, "Agricultural Sources and Sinks of Carbon," in Terrestrial Biospheric Carbon Fluxes:, Dordrecht: Springer Netherlands, 1993, pp. 111-122.
[465] J. Olofsson and T. Hickler, "Effects of human land-use on the global carbon cycle during the last 6,000 years," Veg. Hist. Archaeobot., vol. 17, no. 5, pp. 605-615, Sep. 2008.
[466] E. T. Sundquist, "The Global Carbon Dioxide Budget," Science, vol. 259. American Association for the Advancement of Science, pp. 934-941, 1993.
[467] T. A. Boden, R. J. Andres, and G. Marland, "Global, Regional, and National

Fossil-Fuel CO2 Emissions (1751-2010)," Jan. 2013.
[468] IEA, "CO2 Emissions from Fuel Combustion 2017 - Highlights," 2017.
[469] C. Le Quéré et al., "The global carbon budget 1959-2011," Earth Syst. Sci. Data, vol. 5, no. 1, pp. 165-185, May 2013.
[470] M. R. Raupach et al., "Global and regional drivers of accelerating CO2 emissions.," Proc. Natl. Acad. Sci. U. S. A., vol. 104, no. 24, pp. 10288-93, Jun. 2007.
[471] C. Le Quéré et al., "Global Carbon Budget 2017," Earth Syst. Sci. Data, vol. 10, no. 1, pp. 405-448, Mar. 2018.
[472] C. Le Quéré et al., "Global Carbon Budget 2018," Earth Syst. Sci. Data, vol. 10, no. 4, pp. 2141-2194, Dec. 2018.
[473] F. M. Lea and P. C. Hewlett, Lea's chemistry of cement and concrete. Elservier Butterworth- Heinemann, 2001.
[474] I. Marlowe and D. Mansfield, "Toward a Sustainable Cement Industry: Environment, Health \& Safety Performance Improvement," 2002.
[475] S. Brown, C. A. S. Hall, W. Knabe, J. Raich, M. C. Trexler, and P. Woomer, "Tropical forests: Their past, present, and potential future role in the terrestrial carbon budget," Water, Air, Soil Pollut., vol. 70, no. 1-4, pp. 71-94, Oct. 1993.
[476] C. Le Quéré et al., "Global Carbon Budget 2016," Earth Syst. Sci. Data, vol. 8, no. 2, pp. 605-649, Nov. 2016.
[477] S. Solomon, D. Qin, M. Manning, K. Averyt, and M. Marquis, Climate change 2007: the physical science basis : contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, 2007.
[478] P. N. Pearson and M. R. Palmer, "Atmospheric carbon dioxide concentrations over the past 60 million years," Nature, vol. 406, no. 6797, pp. 695-699, Aug. 2000.
[479] G. L. Stephens et al., "An update on Earth's energy balance in light of the latest global observations," Nat. Geosci., vol. 5, no. 10, pp. 691-696, Oct. 2012.
[480] J. F. B. Mitchell, "The 'Greenhouse’ effect and climate change," Rev. Geophys., vol. 27, no. 1, p. 115, Feb. 1989.
[481] J. Hansen, M. Sato, R. Ruedy, K. Lo, D. W. Lea, and M. Medina-Elizade, "Global temperature change.," Proc. Natl. Acad. Sci. U. S. A., vol. 103, no. 39, pp. 14288-93, Sep. 2006.
[482] A. E. Dessler, Z. Zhang, and P. Yang, "Water-vapor climate feedback inferred from climate fluctuations, 2003-2008," Geophys. Res. Lett., vol. 35, no. 20, p. L20704, Oct. 2008.
[483] C. C. Ummenhofer and G. A. Meehl, "Extreme weather and climate events with ecological relevance: a review.," Philos. Trans. R. Soc. Lond. B. Biol. Sci., vol. 372, no. 1723, p. 20160135, Jun. 2017.
[484] S. Palacios and R. Zimmerman, "Response of eelgrass Zostera marina to CO2 enrichment: possible impacts of climate change and potential for remediation of coastal habitats," Mar. Ecol. Prog. Ser., vol. 344, pp. 1-13, Aug. 2007.
[485] A. P. Trujillo and H. V. Thurman, Essentials of Oceanography. Pearson Education, Inc., 2008.
[486] S. A. Zimov, E. A. G. Schuur, and F. S. Chapin III, "Permafrost and the Global Carbon Budget," Science (80-. )., vol. 312, no. 5780, pp. 1612-1613, 2006.
[487] J. Hansen, "Tipping point: Perspective of a climatologist," State Wild 2008-2009 A Glob. Portrait Wildlife, Wildlands, Ocean., pp. 6-15, 2008.
[488] T. Tyrrell, J. G. Shepherd, and S. Castle, "The long-term legacy of fossil fuels," Tellus B Chem. Phys. Meteorol., vol. 59, no. 4, pp. 664-672, Jan. 2007.
[489] M. Eby et al., "Lifetime of Anthropogenic Climate Change: Millennial Time Scales of Potential CO $2_{2}$ and Surface Temperature Perturbations," J. Clim., vol. 22, no. 10, pp. 2501-2511, May 2009.
[490] T. F. Stocker et al., "Technical Summary," Clim. Chang. 2013 Phys. Sci. Basis. Contrib. Work. Gr. I to Fifth Assess. Rep. Intergov. Panel Clim. Chang., pp. 33115, 2013.
[491] IPCC, Climate Change 2013: The Physical Science Basis. Contribution of

Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press, 2013.
[492] M. Aral and J. Guan, "Global Sea Surface Temperature and Sea Level Rise Estimation with Optimal Historical Time Lag Data," Water, vol. 8, no. 11, p. 519, 2016.
[493] R. A. Feely et al., "Impact of anthropogenic CO2 on the CaCO 3 system in the oceans.," Science, vol. 305, no. 5682, pp. 362-6, Jul. 2004.
[494] J. Raven et al., "Ocean acidification due to increasing atmospheric carbon dioxide," Coral Reefs, vol. 5, no. June, p. 60, 2005.
[495] C. L. Sabine and R. A. Feely, "The Oceanic Sink for Carbon Dioxide," in Greenhouse gas sinks, D. Reay, Ed. CABI, 2007, pp. 31-48.
[496] C. L. Sabine et al., "The Oceanic Sink for Anthropogenic CO2," Science (80-. )., vol. 305, no. 5682, 2004.
[497] J. M. Santana-Casiano, M. González-Dávila, M.-J. Rueda, O. Llinás, and E.-F. González-Dávila, "The interannual variability of oceanic CO2 parameters in the northeast Atlantic subtropical gyre at the ESTOC site," Global Biogeochem. Cycles, vol. 21, no. 1, Mar. 2007.
[498] M. González-Dávila, J. M. Santana-Casiano, M. J. Rueda, and O. Llinás, "The water column distribution of carbonate system variables at the ESTOC site from 1995 to 2004," Biogeosciences, vol. 7, no. 10, pp. 3067-3081, Oct. 2010.
[499] R. M. Key et al., "A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP)," Global Biogeochem. Cycles, vol. 18, no. 4, p. n/an/a, Dec. 2004.
[500] R. Eisler, Oceanic acidification : a comprehensive overview. CRC Press, 2012.
[501] Y. Luo, B. P. Boudreau, and A. Mucci, "Disparate acidification and calcium carbonate desaturation of deep and shallow waters of the Arctic Ocean," Nat. Comтип., vol. 7, p. 12821, Sep. 2016.
[502] J. C. Orr et al., "Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms," Nature, vol. 437, no. 7059, pp. 681-686,

Sep. 2005.
[503] S. C. Doney, V. J. Fabry, R. A. Feely, and J. A. Kleypas, "Ocean Acidification: The Other $\mathrm{CO}_{2}$ Problem," Ann. Rev. Mar. Sci., vol. 1, no. 1, pp. 169-192, Jan. 2009.
[504] G. A. McKinley, A. R. Fay, T. Takahashi, and N. Metzl, "Convergence of atmospheric and North Atlantic carbon dioxide trends on multidecadal timescales," Nat. Geosci., vol. 4, no. 9, pp. 606-610, Sep. 2011.
[505] M. K. Gagan et al., "New views of tropical paleoclimates from corals," Quat. Sci. Rev., vol. 19, no. 1-5, pp. 45-64, Jan. 2000.
[506] O. Hoegh-Guldberg et al., "Coral reefs under rapid climate change and ocean acidification.," Science, vol. 318, no. 5857, pp. 1737-42, Dec. 2007.
[507] A. M. Szmant and N. J. Gassman, "The effects of prolonged 'bleaching' on the tissue biomass and reproduction of the reef coral Montastrea annularis," Coral Reefs, vol. 8, no. 4, pp. 217-224, Apr. 1990.
[508] R. Stone, "A world without corals?," Science, vol. 316, no. May, pp. 678-682, 2007.
[509] J. C. Orr et al., "Research Priorities for Understanding Ocean Acidification: Summary From the Second Symposium on the Ocean in a High- $\mathrm{CO}_{2}$ World," Oceanography, vol. 22. Oceanography Society, pp. 182-189, 2015.
[510] L. S. Evans, P. Fidelman, C. Hicks, C. Morgan, A. Perry, and R. Tobin, Limits to climate change adaptation in the Great Barrier Reef: scoping ecological and social limits. 2012.
[511] J. B. Ries, A. L. Cohen, and D. C. McCorkle, "Marine calcifiers exhibit mixed responses to CO2-induced ocean acidification," Geology, vol. 37, no. 12, pp. 1131-1134, Dec. 2009.
[512] K. L. de la Haye, J. I. Spicer, S. Widdicombe, and M. Briffa, "Reduced sea water pH disrupts resource assessment and decision making in the hermit crab Pagurus bernhardus," Anim. Behav., vol. 82, no. 3, pp. 495-501, Sep. 2011.
[513] F. Gazeau, C. Quiblier, J. M. Jansen, J.-P. Gattuso, J. J. Middelburg, and C. H. R.

Heip, "Impact of elevated CO2 on shellfish calcification," Geophys. Res. Lett., vol. 34, no. 7, p. L07603, Apr. 2007.
[514] T. Kikkawa, A. Ishimatsu, and J. Kita, "Acute CO2 tolerance during the early developmental stages of four marine teleosts," Environ. Toxicol., vol. 18, no. 6, pp. 375-382, Dec. 2003.
[515] K.-S. Lee, J. Kita, and A. Ishimatsu, "Effects of Lethal Levels of Environmental Hypercapnia on Cardiovascular and Blood-Gas Status in Yellowtail, Seriola quinqueradiata," Zoolog. Sci., vol. 20, no. 4, pp. 417-422, Apr. 2003.
[516] D. Moran, R. M. G. Wells, and S. J. Pether, "Low stress response exhibited by juvenile yellowtail kingfish ( Seriola lalandi Valenciennes) exposed to hypercapnic conditions associated with transportation," Aquac. Res., vol. 39, no. 13, pp. 1399-1407, Sep. 2008.
[517] F. Melzner et al., "Physiological basis for high CO\<sub\>2\</sub\> tolerance in marine ectothermic animals: pre-adaptation through lifestyle and ontogeny?," Biogeosciences, vol. 6, no. 10, pp. 2313-2331, Oct. 2009.
[518] J.-P. Gattuso and L. Hansson, Ocean acidification. Oxford University Press, 2011.
[519] B. K. Larsen, H.-O. Pörtner, and F. B. Jensen, "Extra- and intracellular acid-base balance and ionic regulation in cod (Gadus morhua ) during combined and isolated exposures to hypercapnia and copper," Mar. Biol., vol. 128, no. 2, pp. 337-346, May 1997.
[520] E. K. Hoffmann and L. O. Simonsen, "Membrane mechanisms in volume and pH regulation in vertebrate cells.," Physiol. Rev., vol. 69, no. 2, pp. 315-82, Apr. 1989.
[521] C. Turley and K. Boot, "Environmental consequences of ocean acidification: A threat to food security," UNEP Emerg. Issues, p. 12, 2010.
[522] D. L. Dixson, P. L. Munday, and G. P. Jones, "Ocean acidification disrupts the innate ability of fish to detect predator olfactory cues," Ecol. Lett., vol. 13, no. 1, pp. 68-75, Jan. 2010.
[523] P. L. Munday et al., "Ocean acidification impairs olfactory discrimination and homing ability of a marine fish.," Proc. Natl. Acad. Sci. U. S. A., vol. 106, no. 6, pp. 1848-52, Feb. 2009.
[524] W.-S. Chung, N. J. Marshall, S.-A. Watson, P. L. Munday, and G. E. Nilsson, "Ocean acidification slows retinal function in a damselfish through interference with GABAA receptors," J. Exp. Biol., vol. 217, no. 3, pp. 323-326, Feb. 2014.
[525] M. C. O. Ferrari et al., "Effects of ocean acidification on visual risk assessment in coral reef fishes," Funct. Ecol., vol. 26, no. 3, pp. 553-558, Jun. 2012.
[526] S. D. Simpson et al., "Ocean acidification erodes crucial auditory behaviour in a marine fish.," Biol. Lett., vol. 7, no. 6, pp. 917-20, Dec. 2011.
[527] G. E. Brown, J. C. Adrian, Jr., M. G. Lewis, and J. M. Tower, "The effects of reduced pH on chemical alarm signalling in ostariophysan fishes," Can. J. Fish. Aquat. Sci., vol. 59, no. 8, pp. 1331-1338, Aug. 2002.
[528] A. O. H. C. Leduc, J. M. Kelly, and G. E. Brown, "Detection of conspecific alarm cues by juvenile salmonids under neutral and weakly acidic conditions: laboratory and field tests," Oecologia, vol. 139, no. 2, pp. 318-324, Apr. 2004.
[529] J. Caprio, M. Shimohara, T. Marui, S. Harada, and S. Kiyohara, "Marine teleost locates live prey through pH sensing.," Science, vol. 344, no. 6188, pp. 1154-6, Jun. 2014.
[530] I. L. Cripps, P. L. Munday, and M. I. McCormick, "Ocean Acidification Affects Prey Detection by a Predatory Reef Fish," PLoS One, vol. 6, no. 7, p. e22736, Jul. 2011.
[531] D. L. Dixson, A. R. Jennings, J. Atema, and P. L. Munday, "Odor tracking in sharks is reduced under future ocean acidification conditions," Glob. Chang. Biol., vol. 21, no. 4, pp. 1454-1462, Apr. 2015.
[532] T. J. Hamilton, A. Holcombe, and M. Tresguerres, "CO2-induced ocean acidification increases anxiety in rockfish via alteration of GABAA receptor functioning.," Proceedings. Biol. Sci., vol. 281, no. 1775, p. 20132509, Jan. 2014.
[533] D. Weetman, D. Atkinson, and J. C. Chubb, "Effects of temperature on antipredator behaviour in the guppy,Poecilia reticulata," Anim. Behav., vol. 55, no. 5, pp. 1361-1372, May 1998.
[534] G. E. Nilsson et al., "Near-future carbon dioxide levels alter fish behaviour by interfering with neurotransmitter function," Nat. Clim. Chang., vol. 2, no. 3, pp. 201-204, Mar. 2012.
[535] P. Domenici, B. Allan, M. I. McCormick, and P. L. Munday, "Elevated carbon dioxide affects behavioural lateralization in a coral reef fish.," Biol. Lett., vol. 8, no. 1, pp. 78-81, Feb. 2012.
[536] B. M. Devine, P. L. Munday, and G. P. Jones, "Homing ability of adult cardinalfish is affected by elevated carbon dioxide," Oecologia, vol. 168, no. 1, pp. 269-276, Jan. 2012.
[537] F. Melzner, S. Göbel, M. Langenbuch, M. A. Gutowska, H.-O. Pörtner, and M. Lucassen, "Swimming performance in Atlantic Cod (Gadus morhua) following long-term (4-12 months) acclimation to elevated seawater PCO2," Aquat. Toxicol., vol. 92, no. 1, pp. 30-37, Apr. 2009.
[538] J. P. Nowicki, G. M. Miller, and P. L. Munday, "Interactive effects of elevated temperature and CO2 on foraging behavior of juvenile coral reef fish," J. Exp. Mar. Bio. Ecol., vol. 412, pp. 46-51, Jan. 2012.
[539] J. C. A. Pistevos, I. Nagelkerken, T. Rossi, M. Olmos, and S. D. Connell, "Ocean acidification and global warming impair shark hunting behaviour and growth," Sci. Rep., vol. 5, no. 1, p. 16293, Dec. 2015.
[540] B. M. Devine, P. L. Munday, and G. P. Jones, "Rising CO2 concentrations affect settlement behaviour of larval damselfishes," Coral Reefs, vol. 31, no. 1, pp. 229-238, Mar. 2012.
[541] B. M. Devine and P. L. Munday, "Habitat preferences of coral-associated fishes are altered by short-term exposure to elevated CO2," Mar. Biol., vol. 160, no. 8, pp. 1955-1962, Aug. 2013.
[542] K. Hopkins, B. R. Moss, and A. B. Gill, "Increased ambient temperature alters the parental care behaviour and reproductive success of the three-spined
stickleback (Gasterosteus aculeatus)," Environ. Biol. Fishes, vol. 90, no. 2, pp. 121-129, Feb. 2011.
[543] M. Milazzo, S. Mirto, P. Domenici, and M. Gristina, "Climate change exacerbates interspecific interactions in sympatric coastal fishes," J. Anim. Ecol., vol. 82, no. 2, pp. 468-477, Mar. 2013.
[544] A. F. Lopes et al., "Behavioural lateralization and shoaling cohesion of fish larvae altered under ocean acidification," Mar. Biol., vol. 163, no. 12, p. 243, Dec. 2016.
[545] J. Freyhof and M. Kottelat, "Dicentrarchus labrax," IUCN Red List Threat. Species, 2008.
[546] G. D. Pickett and M. G. Pawson, Sea bass: biology, exploitation, and conservation. Chapman \& Hall, 1994.
[547] D. F. Kelley, "The importance of estuaries for sea-bass, Dicentrarchus labrax (L.)," J. Fish Biol., vol. 33, no. sa, pp. 25-33, Dec. 1988.
[548] J. Chervinski, "Sea bass, Dicentrarchus labrax, Linne (Pisces, Serranidae) a 'police-fish' in freshwater ponds and its adaptability to various saline conditions.," Bamidgeh, vol. 26, no. 4, pp. 110-113, 1974.
[549] J. Lancaster, "The feeding ecology of juvenile sea bass, Dicentrarchus Labrax," University of Wales, 1991.
[550] G. Barnabé, "Contribution à la connaissance de la biologie du loup, Dicentrarchus labrax (L.) (poisson Serranidae) de la region de Sète," Université des Sciences et Techniques du Languedoc fre, 1976.
[551] H. Peres and A. Oliva-Teles, "Effect of dietary lipid level on growth performance and feed utilization by European sea bass juveniles (Dicentrarchus labrax)," Aquaculture, vol. 179, no. 1-4, pp. 325-334, Sep. 1999.
[552] D. F. Kelley, "Age determination in bass and assessment of growth and yearclass strength," J. Mar. Biol. Assoc. United Kingdom, vol. 68, no. 01, p. 179, Feb. 1988.
[553] M. W. Aprahamian and C. D. Barr, "The Growth, Abundance and Diet of O-

Group Sea Bass, Dicentrarchus Labrax, From the Severn Estuary," J. Mar. Biol. Assoc. United Kingdom, vol. 65, no. 01, p. 169, Feb. 1985.
[554] R. Nystøyl, Seabass \& Seabream, no. February. 2017.
[555] "European seabass | Fisheries," European commission, 2012. [Online]. Available: https://ec.europa.eu/fisheries/marine_species/farmed_fish_and_shellfish/seabass_ en. [Accessed: 04-Sep-2018].
[556] G. Barnabé, "Rearing bass and gilthead bream," Aquaculture, vol. 2, pp. 647686, 1990.
[557] C. Vinagre, D. Madeira, L. Narciso, H. Cabral, and M. Diniz, "Impact of climate change on coastal versus estuarine nursery areas: cellular and whole-animal indicators in juvenile seabass Dicentrarchus labrax," Mar. Ecol. Prog. Ser., vol. 464, pp. 237-243, Sep. 2012.
[558] S. Malavasi et al., "Effects of Temperature on the Antipredator Behaviour and on the Cholinergic Expression in the European Sea Bass (Dicentrarchus labrax L.) Juveniles," Ethology, vol. 119, no. 7, pp. 592-604, Jul. 2013.
[559] J. A. Grøttum and T. Sigholt, "Acute toxicity of carbon dioxide on European seabass (Dicentrarchus labrax): Mortality and effects on plasma ions," Comp. Biochem. Physiol. Part A Physiol., vol. 115, no. 4, pp. 323-327, Dec. 1996.
[560] S. Cecchini, M. Saroglia, G. Caricato, G. Terova, and L. Sileo, "Effects of graded environmental hypercapnia on sea bass (Dicentrarchus labrax L.) feed intake and acid-base balance," Aquac. Res., vol. 32, no. 6, pp. 499-502, Jun. 2001.
[561] A. Crespel et al., "The development of contemporary European sea bass larvae (Dicentrarchus labrax) is not affected by projected ocean acidification scenarios," Mar. Biol., vol. 164, no. 7, p. 155, Jul. 2017.
[562] S. V. Budaev, "'Personality’ in the guppy (Poecilia reticulata): A correlational study of exploratory behavior and social tendency.," J. Comp. Psychol., vol. 111, no. 4, pp. 399-411, 1997.
[563] E. C. Pope et al., "European sea bass, Dicentrarchus labrax, in a changing ocean," Biogeosciences, vol. 11, no. 9, pp. 2519-2530, May 2014.
[564] "Practical Salinity Scale - 1978," Salinometry, 2018. [Online]. Available: http://salinometry.com/pss-78/. [Accessed: 05-Sep-2018].
[565] Freemake. Digital Wave LTD, 2014.
[566] G. Claireaux, C. Couturier, and A.-L. Groison, "Effect of temperature on maximum swimming speed and cost of transport in juvenile European sea bass (Dicentrarchus labrax).," J. Exp. Biol., vol. 209, no. Pt 17, pp. 3420-8, Sep. 2006.
[567] S. Shams, D. Chatterjee, and R. Gerlai, "Chronic social isolation affects thigmotaxis and whole-brain serotonin levels in adult zebrafish," Behav. Brain Res., vol. 292, pp. 283-287, Oct. 2015.
[568] D. Benhaïm, D. D. Akian, M. Ramos, S. Ferrari, K. Yao, and M.-L. Bégout, "Self-feeding behaviour and personality traits in tilapia: A comparative study between Oreochromis niloticus and Sarotherodon melanotheron," Appl. Anim. Behav. Sci., vol. 187, pp. 85-92, Feb. 2017.
[569] C. Maximino, T. Marques de Brito, C. A. G. de M. Dias, A. Gouveia, and S. Morato, "Scototaxis as anxiety-like behavior in fish," Nat. Protoc., vol. 5, no. 2, pp. 209-216, Feb. 2010.
[570] S. Millot, M.-L. Bégout, and B. Chatain, "Exploration behaviour and flight response toward a stimulus in three sea bass strains (Dicentrarchus labrax L.)," Appl. Anim. Behav. Sci., vol. 119, no. 1-2, pp. 108-114, Jun. 2009.
[571] J. C. Redfern, S. J. Cooke, R. J. Lennox, M. A. Nannini, D. H. Wahl, and K. M. Gilmour, "Effects of maternal cortisol treatment on offspring size, responses to stress, and anxiety-related behavior in wild largemouth bass (Micropterus salmoides)," Physiol. Behav., vol. 180, pp. 15-24, Oct. 2017.
[572] S. Ferrari, D. Benhaïm, T. Colchen, B. Chatain, and M.-L. Bégout, "First links between self-feeding behaviour and personality traits in European seabass, Dicentrarchus labrax," Appl. Anim. Behav. Sci., vol. 161, pp. 131-141, Dec. 2014.
[573] I. T. Jolliffe, Principal component analysis. Springer, 2002.
[574] H. Abdi and L. J. Williams, "Principal component analysis," Wiley Interdiscip.

Rev. Comput. Stat., vol. 2, no. 4, pp. 433-459, Jul. 2010.
[575] H. F. Kaiser, "The varimax criterion for analytic rotation in factor analysis," Psychometrika, vol. 23, no. 3, pp. 187-200, Sep. 1958.
[576] R-Core-Team, "R." 2013.
[577] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models using lme4," Jun. 2014.
[578] R. Bellman, Dynamic programming. Princeton University Press, 1972.
[579] M. C. O. Ferrari et al., "Intrageneric variation in antipredator responses of coral reef fishes affected by ocean acidification: implications for climate change projections on marine communities," Glob. Chang. Biol., vol. 17, no. 9, pp. 2980-2986, Sep. 2011.
[580] U. Riebesell and P. D. Tortell, "Effects of ocean acidification on pelagic organisms and ecosystems," in Ocean acidification, 2011, p. 115.
[581] I. D. Couzin and J. Krause, "Self-Organization and Collective Behavior in Vertebrates," Adv. Study Behav., vol. 32, pp. 1-75, Jan. 2003.
[582] M. Holyoak, R. Casagrandi, R. Nathan, E. Revilla, and O. Spiegel, "Trends and missing parts in the study of movement ecology.," Proc. Natl. Acad. Sci. U. S. A., vol. 105, no. 49, pp. 19060-5, Dec. 2008.
[583] L. Conradt, "Models in animal collective decision-making: information uncertainty and conflicting preferences Models in animal collective decisionmaking : information uncertainty and conflicting preferences," Interface Focus, no. December 2011, 2012.
[584] T. Matsuyama and M. Matsushita, "Fractal Morphogenesis by a Bacterial Cell Population," Crit. Rev. Microbiol., vol. 19, no. 2, pp. 117-135, Jan. 1993.
[585] B. Szabó, G. J. Szöllösi, B. Gönci, Z. Jurányi, D. Selmeczi, and T. Vicsek, "Phase transition in the collective migration of tissue cells: Experiment and model," Phys. Rev. E, vol. 74, no. 6, p. 061908, Dec. 2006.
[586] H. P. Zhang, A. Be'Er, E.-L. Florin, and H. L. Swinney, "Collective motion and density fluctuations in bacterial colonies.," Proc. Natl. Acad. Sci. U. S. A., vol.

107, no. 31, pp. 13626-30, Aug. 2010.
[587] D. A. I. Mavridou, D. Gonzalez, W. Kim, S. A. West, and K. R. Foster, "Bacteria Use Collective Behavior to Generate Diverse Combat Strategies," Curr. Biol., vol. 0, no. 0, Jan. 2018.
[588] X. Tu and D. Terzopoulos, "Artificial fishes: Physics, Locomotion, Perception, Behavior," in Proceedings of the 21st annual conference on Computer graphics and interactive techniques - SIGGRAPH '94, 1994, pp. 43-50.
[589] U. Heppner, F.H., Grenander, "A Stochastic Nonlinear Model for Bird Flocking," in The ubiquity of chaos, 1st ed., Saul Krasner, Ed. American Assoc. Advance. Sci., 1990, pp. 233-238.
[590] M. Ballerini et al., "Empirical investigation of starling flocks: a benchmark study in collective animal behaviour," Anim. Behav., vol. 76, no. 1, pp. 201-215, Jul. 2008.
[591] B. Dumont and D. R. C. Hill, "Multi-agent simulation of group foraging in sheep: effects of spatial memory, conspecific attraction and plot size," Ecol. Modell., vol. 141, no. 1-3, pp. 201-215, Jul. 2001.
[592] D. L. Swain, M. R. Hutchings, and G. Marion, "Using a spatially explicit model to understand the impact of search rate and search distance on spatial heterogeneity within an herbivore grazing system," Ecol. Modell., vol. 203, no. 3-4, pp. 319-326, May 2007.
[593] Y. Guo, G. Poulton, P. Corke, G. J. Bishop-Hurley, T. Wark, and D. L. Swain, "Using accelerometer, high sample rate GPS and magnetometer data to develop a cattle movement and behaviour model," Ecol. Modell., vol. 220, no. 17, pp. 2068-2075, Sep. 2009.
[594] A. J. King et al., "Selfish-herd behaviour of sheep under threat," Curr. Biol., vol. 22, no. 14, pp. R561-R562, Jul. 2012.
[595] H. Fang, D. Dai, S. Li, G. He, and L. Huang, "Forecasting Yangtze finless porpoise movement behavior using an Eulerian-Eulerian-diffusion method (EEDM)," Ecol. Eng., vol. 88, pp. 39-52, Mar. 2016.
[596] A. Serani and M. Diez, "Dolphin Pod Optimization," Springer, Cham, 2018, pp. 50-62.
[597] W. H. Warren, "Collective Motion in Human Crowds," Curr. Dir. Psychol. Sci., vol. 27, no. 4, pp. 232-240, Aug. 2018.
[598] K. Kohlstedt, "Least Resistance: How Desire Paths Can Lead to Better Design 99\% Invisible," $2016 . \quad$ [Online]. Available: https://99percentinvisible.org/article/least-resistance-desire-paths-can-lead-betterdesign/. [Accessed: 22-Aug-2018].
[599] Y. Sugiyama et al., "Traffic jams without bottlenecks-experimental evidence for the physical mechanism of the formation of a jam," New J. Phys., vol. 10, no. 3, p. 033001, Mar. 2008.
[600] S. Zhou et al., "Crowd modeling and simulation technologies," ACM Trans. Model. Comput. Simul., vol. 20, no. 4, pp. 1-35, Oct. 2010.
[601] M. R. Hill, Walking, Crossing Streets and Choosing Pedestrian Routes: A Survey of Recent Insights from the Social/Behavioral Sciences. Lincoln, 1984.
[602] M. Moussaïd, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz, "The Walking Behaviour of Pedestrian Social Groups and Its Impact on Crowd Dynamics," PLoS One, vol. 5, no. 4, p. e10047, Apr. 2010.
[603] U. Blanke, G. Troster, T. Franke, and P. Lukowicz, "Capturing crowd dynamics at large scale events using participatory GPS-localization," in 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2014, pp. 1-7.
[604] M. Muramatsu, T. Irie, and T. Nagatani, "Jamming transition in pedestrian counter flow," Phys. A Stat. Mech. its Appl., vol. 267, no. 3-4, pp. 487-498, May 1999.
[605] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," Phys. Rev. E, vol. 51, no. 5, pp. 4282-4286, May 1995.
[606] N. Pelechano, J. Allbeck, and N. Badler, "Controlling Individual Agents in HighDensity Crowd Simulation," in Proceedings of the 2007 ACM

SIGGRAPH/Eurographics symposium on Computer animation, 2007.
[607] S. Gwynne, E. R. Galea, M. Owen, P. J. Lawrence, and L. Filippidis, "A review of the methodologies used in evacuation modelling," Fire Mater., vol. 23, no. 6, pp. 383-388, Nov. 1999.
[608] S. Ahrndt, A. Aria, J. Fähndrich, and S. Albayrak, "Ants in the OCEAN: Modulating Agents with Personality for Planning with Humans," Springer, Cham, 2015, pp. 3-18.
[609] A. Jadbabaie, Jie Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," IEEE Trans. Automat. Contr., vol. 48, no. 6, pp. 988-1001, Jun. 2003.
[610] E. Bonabeau, "Agent-based modeling: methods and techniques for simulating human systems.," Proc. Natl. Acad. Sci. U. S. A., vol. 99 Suppl 3, no. suppl 3, pp. 7280-7, May 2002.
[611] S. Gueron, S. A. Levin, and D. I. Rubenstein, "The Dynamics of Herds: From Individuals to Aggregations," J. Theor. Biol., vol. 182, no. 1, pp. 85-98, Sep. 1996.
[612] C. K. Hemelrijk, "An individual-orientated model of the emergence of despotic and egalitarian societies," Proc. R. Soc. B Biol. Sci., vol. 266, no. 1417, pp. 361369, Feb. 1999.
[613] C. Castelfranchi, F. De Rosis, R. Falcone, and S. Pizzutilo, "Personality Traits And Social Attitudes In Multiagent Cooperation," Appl. Artif. Intell., vol. 12, no. 7-8, pp. 649-675, Oct. 1998.
[614] S. J. Guy, S. Kim, M. C. Lin, and D. Manocha, "Simulating heterogeneous crowd behaviors using personality trait theory," in Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation - SCA '11, 2011, p. 43.
[615] P. Michelena, R. Jeanson, J.-L. Deneubourg, and A. M. Sibbald, "Personality and collective decision-making in foraging herbivores.," Proceedings. Biol. Sci., vol. 277, no. 1684, pp. 1093-9, Apr. 2010.
[616] A. J. W. Ward, P. Thomas, P. J. B. Hart, and J. Krause, "Correlates of boldness in three-spined sticklebacks ( Gasterosteus aculeatus )," Behav. Ecol. Sociobiol., vol. 55, no. 6, pp. 561-568, Apr. 2004.
[617] J. Krause, "Differential Fitness Returns In Relation To Spatial Position In Groups," Biol. Rev., vol. 69, no. 2, pp. 187-206, May 1994.
[618] A. J. King, D. D. P. Johnson, and M. Van Vugt, "The Origins and Evolution of Leadership," Curr. Biol., vol. 19, no. 19, pp. R911-R916, Oct. 2009.
[619] S. A. Rands, G. Cowlishaw, R. A. Pettifor, J. M. Rowcliffe, and R. A. Johnstone, "Spontaneous emergence of leaders and followers in foraging pairs," Nature, vol. 423, no. 6938, pp. 432-434, May 2003.
[620] P. K. Visscher, "Group Decision Making in Nest-Site Selection Among Social Insects," Annu. Rev. Entomol., vol. 52, no. 1, pp. 255-275, Jan. 2007.
[621] M. H. Pillot, J. Gautrais, J. Gouello, P. Michelena, A. Sibbald, and R. Bon, "Moving together: Incidental leaders and naïve followers," Behav. Processes, vol. 83, no. 3, pp. 235-241, Mar. 2010.
[622] R. H. J. M. Kurvers et al., "Personality differences explain leadership in barnacle geese," Anim. Behav., vol. 78, no. 2, pp. 447-453, Aug. 2009.
[623] L. Conradt and T. J. Roper, "Consensus decision making in animals," Trends Ecol. Evol., vol. 20, no. 8, pp. 449-456, Aug. 2005.
[624] R. O. Deaner, A. V. Khera, and M. L. Platt, "Monkeys Pay Per View: Adaptive Valuation of Social Images by Rhesus Macaques," Curr. Biol., vol. 15, no. 6, pp. 543-548, Mar. 2005.
[625] J. A. Johnson, "Dominance rank in juvenile olive baboons, Papio anubis: the influence of gender, size, maternal rank and orphaning," Anim. Behav., vol. 35, no. 6, pp. 1694-1708, Dec. 1987.
[626] R. I. M. Dunbar, "Structure of Gelada Baboon Reproductive Units," Z. Tierpsychol., vol. 63, no. 4, pp. 265-282, Apr. 2010.
[627] J. Salvit and E. Sklar, "Modulating Agent Behavior using Human Personality Type," in Workshop on Human-Agent Interaction Design and Models (HAIDM)
at Autonomous Agents and MultiAgent Systems (AAMAS), 2012.
[628] F. Durupinar, N. Pelechano, J. M. Allbeck, U. Gudukbay, and N. I. Badler, "How the Ocean Personality Model Affects the Perception of Crowds," IEEE Comput. Graph. Appl., vol. 31, no. 3, pp. 22-31, May 2011.
[629] S. L. Koole, W. Jager, A. E. van den Berg, C. A. J. Vlek, and W. K. B. Hofstee, "On the Social Nature of Personality: Effects of Extraversion, Agreeableness, and Feedback about Collective Resource Use on Cooperation in a Resource Dilemma," Personal. Soc. Psychol. Bull., vol. 27, no. 3, pp. 289-301, Mar. 2001.
[630] R. Kurzban and D. Houser, "Individual differences in cooperation in a circular public goods game," Eur. J. Pers., vol. 15, no. S1, pp. S37-S52, Nov. 2001.
[631] J. A. Chatman and F. J. Flynn, "The Influence Of Demographic Heterogeneity On The Emergence And Consequences Of Cooperative Norms In Work Teams," Acad. Manag. J., vol. 44, no. 5, pp. 956-974, Oct. 2001.
[632] E. Lindqvist, "Height and Leadership," Rev. Econ. Stat., vol. 94, no. 4, pp. 11911196, Nov. 2012.
[633] M. Lindeman and L. Sundvik, "Impact of Height on Assessments of Finnish Female Job Applicants’ Managerial Abilities," J. Soc. Psychol., vol. 134, no. 2, pp. 169-174, Apr. 1994.
[634] N. M. Blaker, I. Rompa, I. H. Dessing, A. F. Vriend, C. Herschberg, and M. van Vugt, "The height leadership advantage in men and women: Testing evolutionary psychology predictions about the perceptions of tall leaders," Gr. Process. Intergr. Relations, vol. 16, no. 1, pp. 17-27, Jan. 2013.
[635] B. Winterhalder, "Optimal foraging strategies and hunter-gatherer research in anthropology: Theory and models," in Hunter-gatherer foraging strategies: Ethnographic and archaeological analyses, 1981, pp. 13-35.
[636] K. Hill, H. Kaplan, K. Hawkes, and A. M. Hurtado, "Foraging decisions among Aché hunter-gatherers: New data and implications for optimal foraging models," Ethol. Sociobiol., vol. 8, no. 1, pp. 1-36, Jan. 1987.
[637] T. T. Hills, C. Kalff, and J. M. Wiener, "Adaptive Lévy Processes and Area-

Restricted Search in Human Foraging," PLoS One, vol. 8, no. 4, p. e60488, Apr. 2013.
[638] D. A. Raichlen, B. M. Wood, A. D. Gordon, A. Z. P. Mabulla, F. W. Marlowe, and H. Pontzer, "Evidence of Levy walk foraging patterns in human huntergatherers.," Proc. Natl. Acad. Sci. U. S. A., vol. 111, no. 2, pp. 728-33, Jan. 2014.
[639] C. T. Brown, L. S. Liebovitch, and R. Glendon, "Lévy Flights in Dobe Ju/'hoansi Foraging Patterns," Hum. Ecol., vol. 35, no. 1, pp. 129-138, Feb. 2007.
[640] F. W. Marlowe, "Hunting and Gathering: The Human Sexual Division of Foraging Labor," Cross-Cultural Res., vol. 41, no. 2, pp. 170-195, May 2007.
[641] M. J. Hamilton, B. T. Milne, R. S. Walker, O. Burger, and J. H. Brown, "The complex structure of hunter-gatherer social networks.," Proceedings. Biol. Sci., vol. 274, no. 1622, pp. 2195-202, Sep. 2007.
[642] A. J. King et al., "Social density processes regulate the functioning and performance of foraging human teams," Sci. Rep., vol. 5, no. 1, p. 18260, Nov. 2015.
[643] E. A. Berg, "A Simple Objective Technique for Measuring Flexibility in Thinking," J. Gen. Psychol., vol. 39, no. 1, pp. 15-22, Jul. 1948.
[644] O. Monchi, M. Petrides, V. Petre, K. Worsley, and A. Dagher, "Wisconsin Card Sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging.," J. Neurosci., vol. 21, no. 19, pp. 7733-41, Oct. 2001.
[645] L. A. Jensen-Campbell, M. Rosselli, K. A. Workman, M. Santisi, J. D. Rios, and D. Bojan, "Agreeableness, conscientiousness, and effortful control processes," J. Res. Pers., vol. 36, no. 5, pp. 476-489, Oct. 2002.
[646] A. H. B. De Hoogh, D. N. Den Hartog, and P. L. Koopman, "Linking the Big Five-Factors of personality to charismatic and transactional leadership; perceived dynamic work environment as a moderator," J. Organ. Behav., vol. 26, no. 7, pp. 839-865, Nov. 2005.
[647] T. A. Judge and J. E. Bono, "Five - Factor Model of Personality and

Transformational Leadership," J. Appl. Psychol., vol. 85, no. 5, pp. 751-765, 2000.
[648] L. R. Goldberg, "An alternative \"description of personality\": the bigfive factor structure.," J. Pers. Soc. Psychol., vol. 59, no. 6, pp. 1216-29, Dec. 1990.
[649] R. R. McCrae and P. T. Costa, "Validation of the five-factor model of personality across instruments and observers.," J. Pers. Soc. Psychol., vol. 52, no. 1, pp. 8190, 1987.
[650] M. R. Barrick and M. K. Mount, "The Big Five Personality Dimensions and Job Performance: A Meta-Analysis," Pers. Psychol., vol. 44, no. 1, pp. 1-26, Mar. 1991.
[651] T. A. Judge, C. A. Higgins, C. J. Thoresen, and M. R. Barrick, "The Big Five personality traits, general mental ability, and career success across the life span," Pers. Psychol., vol. 52, no. 3, pp. 621-652, Sep. 1999.
[652] H. Zhao and S. E. Seibert, "The Big Five personality dimensions and entrepreneurial status: A meta-analytical review.," J. Appl. Psychol., vol. 91, no. 2, pp. 259-271, Mar. 2006.
[653] M. C. O'Connor and S. V. Paunonen, "Big Five personality predictors of postsecondary academic performance," Pers. Individ. Dif., vol. 43, no. 5, pp. 971990, Oct. 2007.
[654] I. B. Myers, M. H. McCaulley, N. L. Quenk, and A. L. Hammer, MBTI manual: A guide to the development and use of the Myers-Briggs Type Indicator (Vol. 3). Palo Alto, CA: Consulting Psychologists Press, 1998.
[655] R. M. Capraro and M. M. Capraro, "Myers-Briggs Type Indicator Score Reliability Across: Studies a Meta-Analytic Reliability Generalization Study," Educ. Psychol. Meas., vol. 62, no. 4, pp. 590-602, Aug. 2002.
[656] A. Thorne and H. G. Gough, Portraits of type: An MBTI research compendium. Davies-Black Publishing, 1991.
[657] R. R. McCrae and P. T. Costa, "Reinterpreting the Myers-Briggs Type Indicator

From the Perspective of the Five-Factor Model of Personality," J. Pers., vol. 57, no. 1, pp. 17-40, Mar. 1989.
[658] A. Furnham, "The big five versus the big four: the relationship between the Myers-Briggs Type Indicator (MBTI) and NEO-PI five factor model of personality," Pers. Individ. Dif., vol. 21, no. 2, pp. 303-307, Aug. 1996.
[659] P. T. Costa and R. R. McCrae, Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI): Professional manual. Psychological Assessment Resources, Incorporated, 1992.
[660] R. R. McCrae, "NEO-PI-R Data from 36 Cultures," in The Five-Factor Model of Personality Across Cultures, Boston, MA: Springer US, 2002, pp. 105-125.
[661] J. L. Holland, J. A. Johnston, K. F. Hughey, and N. F. Asama, "Some Explorations of a Theory of Careers: VII. A Replication and Some Possible Extensions," J. Career Dev., vol. 18, no. 2, pp. 91-100, Dec. 1991.
[662] I. Montag and J. Levin, "The five-factor personality model in applied settings," Eur. J. Pers., vol. 8, no. 1, pp. 1-11, Mar. 1994.
[663] R. L. Piedmont and H. P. Weinstein, "Predicting Supervisor Ratings of Job Performance Using the NEO Personality Inventory," J. Psychol., vol. 128, no. 3, pp. 255-265, May 1994.
[664] J. G. Rosse, J. L. Miller, and M. D. Stecher, "A field study of job applicants' reactions to personality and cognitive ability testing.," J. Appl. Psychol., vol. 79, no. 6, pp. 987-992, 1994.
[665] J. F. Salgado and A. Rumbo, "Personality and Job Performance in Financial Services Managers," Int. J. Sel. Assess., vol. 5, no. 2, pp. 91-100, Apr. 1997.
[666] B. Rammstedt and O. P. John, "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German," J. Res. Pers., vol. 41, no. 1, pp. 203-212, Feb. 2007.
[667] A. Polus, J. L. Schofer, and A. Ushpiz, "Pedestrian Flow and Level of Service," J. Transp. Eng., vol. 109, no. 1, pp. 46-56, Jan. 1983.
[668] Y. Tanaboriboon, S. S. Hwa, and C. H. Chor, "Pedestrian Characteristics Study
in Singapore," J. Transp. Eng., vol. 112, no. 3, pp. 229-235, May 1986.
[669] M. R. Vikler, "Prediction and Measurement of Travel Time Along Pedestrian Routes," Transp. Res. Rec. J. Transp. Res. Board, no. 1636, pp. 37-42, 1998.
[670] R. W. Bohannon, "Comfortable and maximum walking speed of adults aged 20-79 years: reference values and determinants," Age Ageing, vol. 26, no. 1, pp. 15-19, Jan. 1997.
[671] A. Kuznetsova, P. Brockhoff, and R. Christensen, "lmerTest Package: Tests in Linear Mixed Effects Models.," J. Stat. Softw., vol. 82, no. 13, pp. 1-26, 2017.
[672] T. Offerman and A. Schotter, "Sampling for information or sampling for imitation: An experiment on social learning with information on ranks," New York Univ., 2005.
[673] M. J. Salganik, P. S. Dodds, and D. J. Watts, "Experimental study of inequality and unpredictability in an artificial cultural market.," Science, vol. 311, no. 5762, pp. 854-6, Feb. 2006.
[674] J.-M. Amé, J. Halloy, C. Rivault, C. Detrain, and J. L. Deneubourg, "Collegial decision making based on social amplification leads to optimal group formation.," Proc. Natl. Acad. Sci. U. S. A., vol. 103, no. 15, pp. 5835-40, Apr. 2006.
[675] D. Sivers, "How to start a movement," Ted2010, 2010. [Online]. Available: https://www.ted.com/talks/derek_sivers_how_to_start_a_movement\#t-19592. [Accessed: 24-Aug-2018].
[676] G. I. Wilson et al., "A new perspective on how humans assess their surroundings; derivation of head orientation and its role in 'framing' the environment," PeerJ, vol. 3, 2015.
[677] W. F. Lloyd, Two lectures on the checks to population. 1833.
[678] P. Mannion, S. Devlin, J. Duggan, and E. Howley, "Avoiding the Tragedy of the Commons using Reward Shaping," 2016.
[679] O. P. John and S. Srivastava, "The Big Five trait taxonomy: History, measurement, and theoretical perspectives," Handb. Personal. Theory Res., vol.

2, no. 510, pp. 102-138, 1999.
[680] M. A. Girshick, "An Elementary Survey of Statistical Decision Theory," Rev. Educ. Res., vol. 24, no. 5, pp. 448-466, Dec. 1954.
[681] S. Bubeck and N. Cesa-Bianchi, "Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems," Found. Trends® Mach. Learn., vol. 5, no. 1, pp. 1-122, Dec. 2012.
[682] E. Brochu, M. W. Hoffman, and N. de Freitas, "Portfolio Allocation for Bayesian Optimization," eprint arXiv:1009.5419, Sep. 2010.
[683] L. Li, W. Chu, J. Langford, and R. E. Schapire, "A contextual-bandit approach to personalized news article recommendation," in Proceedings of the 19th international conference on World wide web - WWW '10, 2010, p. 661.
[684] J. R. Krebs, A. Kacelnik, and P. Taylor, "Test of optimal sampling by foraging great tits," Nature, vol. 275, no. 5675, pp. 27-31, Sep. 1978.
[685] I. Rojas-Ferrer, "Predicting Choices in Bumblebees (Bombus impatiens): Learning Rules and The Two-Armed Bandit," Theses, Jul. 2016.
[686] P. Auer, N. Cesa-bianchi, and P. Fischer, "Finite-time Analysis of the Multiarmed Bandit Problem," Mach. Learn., vol. 47, no. 2-3, pp. 235-256, 2002.
[687] W. R. Thompson, "On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples," Biometrika, vol. 25, no. 3/4, p. 285, Dec. 1933.
[688] S. Agrawal and N. Goyal, "Analysis of Thompson Sampling for the Multi-armed Bandit Problem." p. 39.1-39.26, 16-Jun-2012.
[689] O. Chapelle and L. Li, "An Empirical Evaluation of Thompson Sampling." pp. 2249-2257, 2011.
[690] L. Kocsis and C. Szepesvári, "Bandit Based Monte-Carlo Planning," Springer, Berlin, Heidelberg, 2006, pp. 282-293.
[691] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "Gambling in a rigged casino: The adversarial multi-armed bandit problem," in Proceedings of IEEE 36th Annual Foundations of Computer Science, 1995, pp. 322-331.
[692] K. Liu and Q. Zhao, "Distributed Learning in Multi-Armed Bandit With Multiple Players," IEEE Trans. Signal Process., vol. 58, no. 11, pp. 5667-5681, Nov. 2010.
[693] K. H. Schlag, "Why Imitate, and If So, How?: A Boundedly Rational Approach to Multi-armed Bandits," J. Econ. Theory, vol. 78, no. 1, pp. 130-156, Jan. 1998.
[694] L. Rendell, R. Boyd, M. Enquist, M. W. Feldman, L. Fogarty, and K. N. Laland, "How copying affects the amount, evenness and persistence of cultural knowledge: insights from the social learning strategies tournament.," Philos. Trans. R. Soc. Lond. B. Biol. Sci., vol. 366, no. 1567, pp. 1118-28, Apr. 2011.
[695] P. B. Reverdy, V. Srivastava, and N. E. Leonard, "Modeling Human Decision Making in Generalized Gaussian Multiarmed Bandits," Proc. IEEE, vol. 102, no. 4, pp. 544-571, Apr. 2014.
[696] D. Racey, M. E. Young, D. Garlick, J. Ngoc-Minh Pham, and A. P. Blaisdell, "Pigeon and human performance in a multi-armed bandit task in response to changes in variable interval schedules," Learn. Behav., vol. 39, no. 3, pp. 245258, Sep. 2011.
[697] R. D. Luce, "Detection and recognition." 1963.
[698] P. Eklöv, "Group foraging versus solitary foraging efficiency in piscivorous predators: the perch, Perca fluviatilis, and pike, Esox lucius, patterns," Anim. Behav., vol. 44, pp. 313-326, Aug. 1992.
[699] A. le Roux, M. I. Cherry, L. Gygax, and M. B. Manser, "Vigilance behaviour and fitness consequences: comparing a solitary foraging and an obligate groupforaging mammal," Behav. Ecol. Sociobiol., vol. 63, no. 8, pp. 1097-1107, Jun. 2009.
[700] A. L. Rypstra, "Foraging success of solitary and aggregated spiders: insights into flock formation," Anim. Behav., vol. 37, pp. 274-281, Feb. 1989.
[701] R. Boyd and P. J. Richerson, "Culture and the evolution of human cooperation.," Philos. Trans. R. Soc. Lond. B. Biol. Sci., vol. 364, no. 1533, pp. 3281-8, Nov. 2009.
[702] R. W. Byrne, A. Whiten, and S. P. Henzi, "Social relationships of mountain baboons: Leadership and affiliation in a non-female-bonded monkey," Am. J. Primatol., vol. 20, no. 4, pp. 313-329, Jan. 1990.
[703] S. Finkelstein, D. C. Hambrick, and A. A. Cannella, Strategic leadership : theory and research on executives, top management teams, and boards. Oxford University Press, 2009.
[704] R. M. (Ralph M. Stogdill and B. M. Bass, Stogdill's handbook of leadership : a survey of theory and research. Free Press, 1981.
[705] V. Schaller, C. Weber, C. Semmrich, E. Frey, and A. R. Bausch, "Polar patterns of driven filaments," Nature, vol. 467, pp. 73-77, 2010.
[706] T. Vicsek, "Universal patterns of collective motion from minimal models of flocking," Proc. - 2nd IEEE Int. Conf. Self-Adaptive Self-Organizing Syst. SASO 2008, pp. 3-11, 2008.
[707] A. Eriksson, M. Nilsson Jacobi, J. Nyström, and K. Tunstrøm, "Determining interaction rules in animal swarms," Behav. Ecol., vol. 21, no. 5, pp. 1106-1111, 2010.
[708] S. Weitz, S. Blanco, R. Fournier, J. Gautrais, C. Jost, and G. Theraulaz, "Modeling collective animal behavior with a cognitive perspective: A methodological framework," PLoS One, vol. 7, no. 6, 2012.
[709] H. Xiong, G. Pandey, M. Steinbach, and V. Kumar, "Enhancing data analysis with noise removal," IEEE Trans. Knowl. Data Eng., vol. 18, no. 3, pp. 304-319, 2006.
[710] N. M. Nasrabadi, "Pattern Recognition and Machine Learning," Electron. Imaging, vol. 16, no. 4, 2007.
[711] L. Cazenille et al., "How mimetic should a robotic fish be to socially integrate into zebrafish groups?," Bioinspiration and Biomimetics, vol. 13, no. 2, 2018.
[712] Q. Shi et al., "Modulation of rat behaviour by using a rat-like robot," Bioinspiration and Biomimetics, vol. 8, no. 4, 2013.
[713] D. Strömbom and A. J. King, "Robot Collection and Transport of Objects: A

Biomimetic Process," Front. Robot. AI, vol. 5, no. May, p. 48, 2018.
[714] Domo, "Data Never Sleeps 5.0," www.Domo.com, 2018. [Online]. Available: https://www.domo.com/learn/data-never-sleeps-5.
[715] T. Terada et al., "Design of a car navigation system that predicts user destination," Proc. - IEEE Int. Conf. Mob. Data Manag., vol. 2006, 2006.
[716] R. Claes and T. Holvoet, "Traffic coordination using aggregation-based traffic predictions," IEEE Intell. Syst., vol. 29, no. 4, pp. 96-100, 2014.
[717] M. Swan, "The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery," Surveill. Soc., vol. 11, no. 1-2, pp. 85-99, 2013.
[718] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock, "The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery," Proc. Natl. Acad. Sci., vol. 111, no. 29, pp. 10779-10779, 2014.
[719] E. Dixon, E. Enos, and S. Brodmerkle, "A/B testing of a webpage," US7975000B2, 2005.
[720] E. Ferrara, O. Varol, C. Davis, ... F. M.-C. of the, and U. 2016, "The rise of social bots," Dl.Acm.Org, 2016.
[721] R. Kohavi and R. Longbotham, "Online Controlled Experiments and A/B Testing," Encycl. Mach. Learn. Data Min., no. Ries 2011, pp. 922-929, 2017.
[722] Amazon.com, "The Math Behind A/B Testing," Developer.amazon.com, 2015. [Online]. Available: https://developer.amazon.com/public/apis/manage/ab-testing/doc/math-behind-ab-testing. [Accessed: 21-Sep-2015].
[723] P. N. Howard and B. Kollanyi, "Bots , \# StrongerIn , and \# Brexit: Computational Propaganda during the UK-EU Referendum," COMPROP Res. Note, vol. 1, 2016.

