Author:
Auletta, Fabrizia

Title:
Identifying and modelling decision making and collective behaviour in multi-agent human and artificial systems

Identifying and modelling decision making and collective behaviour in multi-agent human and artificial systems

By

FABRIZIA AULETTA

Department of Engineering Mathematics
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Engineering.

MARCH 30TH, 2022

Word count: 29854

ABSTRACT

The emergence of coordinated action between interacting individuals or agents is a common characteristic of everyday behaviour. Pivotal to the organisation of multiagent activity is the ability of agents to effectively decide how and when to act, with robust decision-making often differentiating expert from non-expert performance. In this thesis we investigated and modelled the behavioural coordination and decision-making behaviour of human and artificial agents completing various *herding tasks*. Herding tasks involve the interaction of two sets of autonomous agents – one or more *herder agents* are required to corral a set of heterogeneous *target agents*. Such activities are ubiquitous in daily life and provide a prototypical example of everyday multi-agent behaviour. We first propose a simple set of local control rules and target selection strategies that enable herder agents to collect and contain a herd of non-cooperative, non-flocking target agents. We then investigated the robustness of the proposed control process to variations in herd size and the strength of the repulsive force that herders imposed on targets. The effectiveness of the proposed approach was also confirmed via ROS simulations and experiments using real robots. We then employed supervised machine learning (SML) to predict the target selection decisions of human herders. The findings demonstrated that the decision-making behaviour of human actors can be effectively predicted using SML at both short (< 1 s) and long (> 10 s) timescales, and that the resultant models can be employed to endow artificial herders with "human-like" decision making capabilities. Finally, we employed explainable AI to understand the state information employed by human herders when making target selection decisions. The findings revealed differences in how expert and novice herders weight state information when making decisions and is the first study to highlight the potential utility of explainable AI techniques for understanding human decision-marking behaviour during multi-agent fast paced interactions.

DEDICATION AND ACKNOWLEDGEMENTS

To my family, for supporting me, especially on choices that would have led me far from home. Thank you, mum, dad, Francesca, for being the example that everything is possible, everything is achievable if done with passion, integrity and kindness.

To Roberto Pellungrini, we crossed paths by chance but remained on the same path by choice; the luckiest result of this doctorate and the one I am most grateful for.

To Martina, Maria and Lucia, for being the wonderful and unique human beings you are, you made even the most stressful times (and I know we are thinking about the same ones) funny memories to look back at. A thank also to all my friends and colleagues, sharing laughs, litres of coffee, late-night teas and video calls made everything easier.

A special thank to my supervisors, professors Mario di Bernardo and Michael J. Richardson. None of this would have been possible without you. I am extremely grateful to you for taking me on this journey, always believing in me and navigating together this innovative cotutelle experience. Your love for research and science inspired me to be a better scientist, your teachings and guidance made me a better scientist.

Finally, I would like to thank the staff of the University of Bristol and Macquarie University for making possible this work, conducted under the cotutelle agreement between the Department of Mathematical Engineering at the University of Bristol (Bristol, United Kingdom) and the School of Psychological Sciences at Macquarie University (Sydney, Australia).

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author. This thesis is being submitted to Macquarie University and University of Bristol in accordance with the Cotutelle agreement dated 10^{th} October 2018.

SIGNED: ..... DATE: .....

TABLE OF CONTENTS

List of Tables

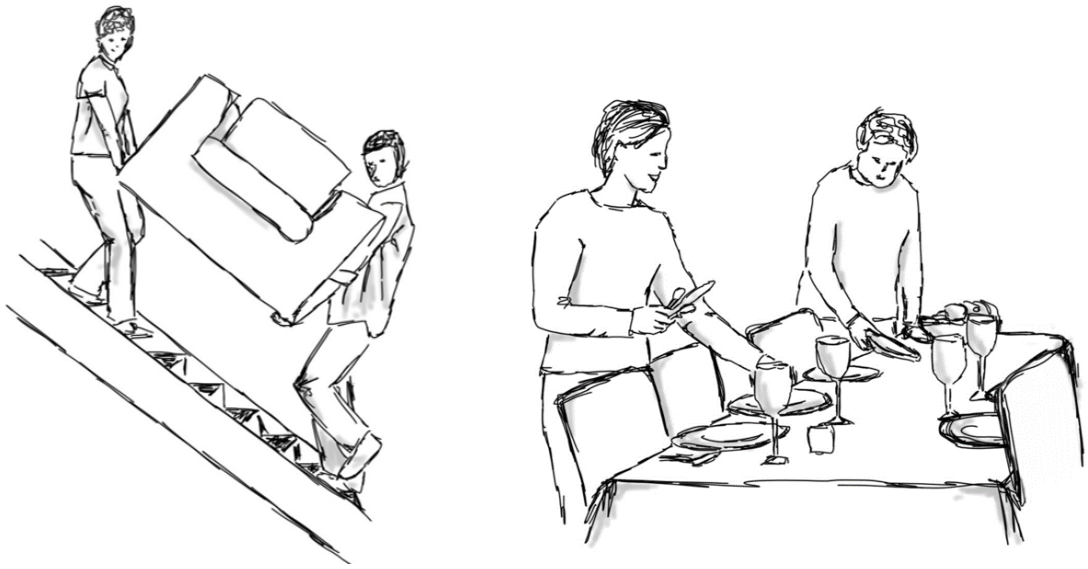
List of Figures

CHAPTER 1

INTRODUCTION

The emergence of coordinated action between groups of interacting individuals or agents is a common characteristic of *multi-agent systems*, e.g. fish schools, bird flocks, insect swarms, or human crowds [2, 36, 39, 130]. Commonly referred to as joint-action in the cognitive and psychological sciences, such behaviour requires that co-actors reciprocally coordinate and adjust their actions with respect to each other and to changing task demands [34, 127, 136, 139]. Thus, pivotal to the structural organisation of coordinated joint-action is the ability of co-actors to effectively decide how and when to act, with robust decision-making often differentiating expert from non-expert performance [35]. This is true whether one considers the simple activity of two or more family members moving a piece of furniture together or setting a table [128, 137, 161] (see Figure 1.1), or the more complex activities that elite athletes engage in during team sports [7, 174] or soldiers in military units perform during high-stakes operations [173].

Understanding the decision-making processes that lead to effective joint- (and individual-) action is therefore fundamental for developing a basic and applied understanding of task- and teamworking skills [97, 138], as well as human decision-making in general. However, identifying and modelling the decision-making processes that lead people to make effective action decisions remains a key challenge, particularly in complex multi-agent tasks that require a high level of situational awareness. This is because, in contrast to tasks that require practical reasoning or deliberative decision-making (i.e., where an actor extensively evaluates all possibilities to determine the optimal action), the decision-making that occurs during many multi-agent activities is fast-paced and highly context dependent [157, 172], with actors spontaneously adapting their actions to achieve task goals as "best as possible" (i.e., with little deliberative or conscious concern for what is actually optimal [25]). Indeed, the actions performed by human co-actors during multi-agent behavioural activities are essentially part of the decision-making process itself, and not simply a consequence



(a) (b)

Figure 1.1: Example of pairs performing coordinated joint actions in everyday life.

of decision-making events [14, 83, 85, 172]. Coherent with research on naturalistic decision making [73–75], the effectiveness of action decisions during joint-action or multi-agent behaviour is a function of an actor's level of situational control and awareness [24, 96]. That is, task competence reflects the trained attention or attunement of an actor to the information that specifies what action possibilities (affordances) may be enacted to ensure task completion [68, 128, 150, 158, 176].

Due to the rapid advances in interactive robotic or artificial agents (AAs) and their increasing presence within our social world, understanding and modelling human decision making during joint-action also has significant implications for the development of interactive AAs [41, 43]. Indeed, like human-human interaction, effective human-machine interaction rests on the ability of AAs to not only predict the future actions or action decisions of human actors [19, 43, 122], but, in many instances, also to enact human-understandable patterns of reciprocal behavioural action [15, 19, 62, 81, 109, 166]. One way of achieving this is to endow AAs with the same decision making processes employed by human actors.

In addition to the general challenges associated with identifying and understanding human decision making during joint-action, developing AAs capable of human-like decision making behaviour requires addressing a number of other key questions. First, how do different decision-making processes and levels of expertise influence the collective behaviour (e.g. an autonomous distribution of workload) and the overall performance of a multi-agent system? Second, how to best develop models of effective (and ineffective) joint-action behaviour and decision-making that cannot only effectively predict human behaviour, but can be readily implemented into the control architecture

of AAs? And, finally, determine what specific task information should human actors or AAs leverage to make effective decisions during multi-agent or joint-action behaviour.

In this Thesis, we begin to tackle these issues by understanding and modelling the coordinate movement and decision-making behaviour of humans and AAs completing various *herding tasks*. Herding tasks involve the interaction of two sets of autonomous agents – one or more *herder* agents are required to corral and contain a set of heterogeneous *target* agents. Such activities are ubiquitous in daily life and provide a prototypical example of everyday joint- or multi-agent behaviour. Indeed, while the most obvious examples involve farmers herding sheep or cattle (see Figure 1.2), similar task dynamics define teachers corralling a group of young children through a museum or firefighters evacuating a crowd of people from a building [94].

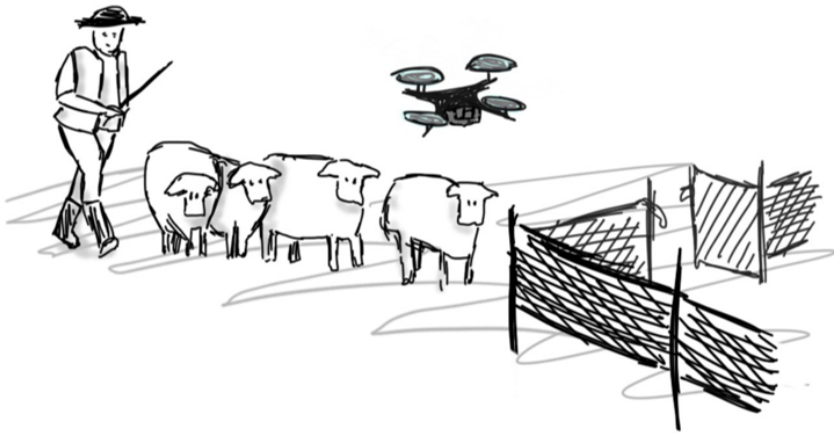


Figure 1.2: Example of one human herder and one artificial herder agent corralling a flock of sheep in a fenced arena.

The majority of research involving the herding task has focused on the modelling and design of autonomous "herder" agents able to successfully corral a second set of flocking "target" agents. Popular solutions, used to explore and design bio-inspired autonomous robotic or artificial systems capable of interacting with animal targets and other human (or animal) herders, span from using a combination of virtual attraction forces [84, 151, 160], to exploiting path planning techniques [67, 113], to non-linear feedback controllers [87, 120] and bio-inspired computation and dynamical models [53, 109]. In each case, herders' actions are both explicitly embedded in the herder dynamics and implicitly conveyed in how each herder selects which target to corral at each time step; i.e., herders decision making or *target selection strategy*. Indeed, fundamental to the success of such herding models is the target selection strategies adopted by herders, (i.e., which target or targets a herder or herders choose to corral at any point in time). This is particularly true when targets do not exhibit collective flocking behaviour, which (as detailed elsewhere in this Thesis) significantly simplifies the herding task (problem). However, despite the central role that target selection

strategies play in the success of herding, the rules that have been employed when modelling human performance or developing AA-herders are typically derived and implemented heuristically [67, 109, 129, 151], with almost no empirical research directed towards determining the validity of these heuristic selection strategies.

Accordingly, the work presented in this Thesis is one of the first attempts (i) to consider the target selection strategies (target action decision) of herders in non-flocking autonomous target herding tasks. In addition, the Thesis explores (ii) how to couple herders' local dynamics with target selection strategies that dynamically leverage computationally low-cost task information and (iii) how to best model and understand target selection strategies of pairs and teams of human agents acting as herders. The overall aim is to develop a deeper understanding of decision making during skilful multi-agent behaviour and to aid in the design of AAs that are not only able to successfully corral large or small herds of non-flocking targets but can work seamlessly with human co-actors (e.g., drones as in Figure 1.2).

1.1 Thesis outline

We start by providing an extensive overview of the herding literature in Chapter 2. Specifically, we provide a general framework for the herding task, currently missing in the literature. In doing so, we survey relevant state of the art herding solutions, highlighting how the limitations of these solutions are due to assuming target flocking behaviour, and to herder decision strategies being a combination of simple rules that exploit such modelling hypothesis of flocking target agents (e.g., each herder first corral the target farther from the herd to the herd).

We address the latter issue in Chapter 3 by designing artificial herder agents capable of successfully corralling a large non-flocking and freely roaming herd. The dynamics of each herder leads the agent to corral one specific target at a time, while dynamic target selection strategies shape the collective coordination between them. The local control laws describing herder motion were derived from the more complex and highly non-linear phenomenological model proposed in [107, 109], which replicates the emergent herding behaviour observed in human pairs. The derivation from the original model to a simplified version and the motivation for adopting this simplified model are presented in Appendix A. Innovatively, such local control laws and dynamic decision strategies exploit the spatial distribution of the herders without using formation control techniques or solving (on-line or off-line) optimisation problems. Consequently, such models result in herding solutions that are effective and can be easily implemented in physical systems, as demonstrated in Chapter 3 via simulations in ROS and experiments on real robots.

The possibility of solving the herding task by equipping AAs with the decision-making process employed by real human actors is explored in Chapter 4. More specifically, we exploit supervised machine learning (SML), and human herding data from [129], to train artificial neural networks (ANNs) to model and predict the target selection strategies of expert and non-expert pairs of hu-

man agents acting as herders. Extensive numerical simulations demonstrate how artificial herders employing these human-inspired decision making models are not only as successful as the artificial herders that employ the dynamic target selection strategies investigated in Chapter 3, but can do so with lower computational costs by requiring only short sequences of partial state information.

The modelling approach presented in Chapter 4 represents an innovative "human-inspired" solution for the design of AAs. Nevertheless, as noted above, key to the development of effective human-machine interaction is developing a deeper understanding of human decision making behaviour, including identifying what differentiates expert from non-expert performance. Thus, in the following two Chapters we demonstrate how explainable AI tools can not only be employed to identify differences in the decision making processes of expert and novice pairs of human herders in a smaller-scale fast-paced herding game (Chapter 5), but also in human teams completing a larger-scale, slower-paced herding task (Chapter 6). More specifically, we leverage SHapley Additive exPlanation [92] to identify how ANN models that are able to accurately predict the target selection strategies of human actors weighted task information.

Finally, a summary of our results and a discussion of their implications, followed by directions for future work, are presented in Chapter 7. Here it is highlighted how the results of this Thesis benefited from the cotutelle agreement between the Department of Mathematical Engineering at the University of Bristol (Bristol, United Kingdom) and the School of Psychological Sciences at Macquarie University (Sydney, Australia), which provided the ideal context for this highly interdisciplinary research project to be conducted.

Note that in order to maximise the readability of the Thesis, some of the technical and methodologically details relating to model development and data analysis are provided in Appendix B. A significant amount of additional simulation and data analysis work was also conducted to validate the findings and conclusions detailed in Chapters 4-6, this supplementary material is reported in Appendix C-E.

1.2 Publications

Part of the work presented in this Thesis were the subject of the publications listed below:

- **Fabrizia Auletta**, Davide Fiore, Michael J. Richardson, and Mario di Bernardo. "*Herding stochastic autonomous agents via local control rules and online global target selection strategies*". Autonomous Robots, 1-13. 2022
  (Chapter 3)

- **Fabrizia Auletta**, Mario di Bernardo, and Michael J. Richardson. "*Human-inspired strategies to solve complex joint tasks in multi agent systems*". In Proc. of the 6th IFAC Hybrid Conference on Analysis and Control of Chaotic Systems (CHAOS), IFAC-PapersOnLine, 54(17), 105-110. 2021
  (Chapter 4)

- **Fabrizia Auletta**, Rachel W. Kallen, Mario di Bernardo, and Michael J. Richardson. "*Employing Supervised Machine Leaning and Explainable-AI to Model and Understand Human Decision Making During Skillful Joint-Action*". Prepared for submission. 2021
  (Chapter 5)

Other publications include:

- Francesco De Lellis, **Fabrizia Auletta**, Giovanni Russo, and Mario diBernardo. "*An Application of Control- Tutored Reinforcement Learning to the Herding Problem*". In Proc. of the 17th IEEE International Workshop on Cellular Nanoscale Networks and their Applications (CNNA) (pp. 1-4). 2021

CHAPTER 2

THE MULTI-AGENT HERDING TASK

Long *et al.* in [90] have recently surveyed the state-of-the-art of multi-agent herding systems, with a focus on the heuristic models and the simulation tools that have previously been used to solve the multi-agent herding task. Still missing in the literature, however, is a general mathematical formulation of the herding problem to help conceptualise and understand the task problem and potential solutions within a common framework. Accordingly, in this Chapter we first propose a general mathematical formulation of the herding task (Section 2.1). Using this formulation we then review the possible modelling approaches (Section 2.2) that have been adopted to solve the herding task, as well as survey the relevant herding literature with respect to potential solutions that involve multiple (Section 2.3) and single herders (Section 2.4). Finally, in Section 2.5, we briefly review the pursuit-evasion theory and encirclement problems related to solving the multi-agent herding task.

2.1 Mathematical formulation of the herding problem

A crowd strolling through a mall, a school of dolphins foraging, or a pack of predators hunting in the woods are all examples of multi-agent systems where the collective behaviour of the system emerges via local *agent* interactions. A sub-category of such systems is the *multi-agent herding system* where the individual and collective behaviour of a set of (passive) *target* agents is influenced via interactions with a second set of (active) *herder* agents. Synonymous with farmers herding sheep or cattle, or teachers herding students together during a field trip, such multi-agent herding scenario always entail the following (irrespective of the specific agents involved or the environmental task context):

(I) the interaction of two sets of heterogeneous dynamical agents, namely herders and targets,

(II) the possibility of directly controlling the dynamics of the herder agents,

(III) the possibility of influencing the dynamics of the target agents (i.e., the herd) only through their proximity to herder agents,

(IV) the existence of a goal region in space or herd density area in which the herder should contain the targets (i.e., where all targets are considered to be successfully herded) needs to be satisfied in finite time.

Here we will focus on the planar herding problem where agents move in a region $\mathrm{D} \subseteq \mathbb{R}^2$. Thus, more formally, using $\mathbf{y}_j$ to represent the position of the j-th herder (N_H) and $\mathbf{x}_i$ to represent the position of the i-th target agent (N_T) within the herding task space, the herding problem can be stated as follows.

Herding problem. *Given N_H autonomous herder agents*

$$\dot{\mathbf{y}}_j(t) = f(t, \mathbf{x}_1, \ldots, \mathbf{x}_{N_T}, \mathbf{y}_1, \ldots, \mathbf{y}_{N_H}) \qquad j = 1, ..., N_H \qquad , \tag{2.1}$$

able to influence $N_T \geq N_H$ target agents, whose dynamics is expressed by

$$\dot{\mathbf{x}}_i(t) = g(\mathbf{x}_i(t), \mathbf{y}_j(t)) \qquad i = 1, ..., N_T \qquad , \tag{2.2}$$

and a goal area $G \subseteq \mathbb{R}^2$ we want to design the herder dynamics

$$f : \mathbb{R}^{2N_H} \times \mathbb{R}^{2N_T} \times \mathbb{R} \rightarrow \mathbb{R}^{2N_H} \tag{2.3}$$

so as to guarantee that

$$\exists t_g \geq 0 : \mathbf{x}_i \in G \qquad \forall i, \forall t \geq t_g. \tag{2.4}$$

Here we choose G as the circular area of radius $r^\star$ and centred at $\mathbf{x}^\star$ so that the goal is assumed to be achieved if

$$\exists t_g \geq 0 : \|\mathbf{x}_i(t) - \mathbf{x}^\star(t)\| \leq r^\star(t) \qquad \forall i, \forall t \geq t_g \tag{2.5}$$

where $\|\cdot\|$ denotes the Euclidean norm and $\mathbf{x}^\star(t)$ and $r^\star(t)$ are the centre and radius.

A graphical representation of the herding task is depicted in Figure 2.1, and in Table 2.1 we provide a list of the task variables that will be used in this Chapter and in the rest of the Thesis.

2.1.1 Target agent modelling

Before reviewing the potential modelling solutions for herder agents, it is important to first briefly detail how the behaviour of target agents has been modelled within the multiagent herding task literature. This is relevant, as further detailed later on in this Chapter and within the following Chapters, because the ability of herders to corral and confine targets is highly dependent on both the type of influence they exert on target agents and the manner by which target agents influence each other.

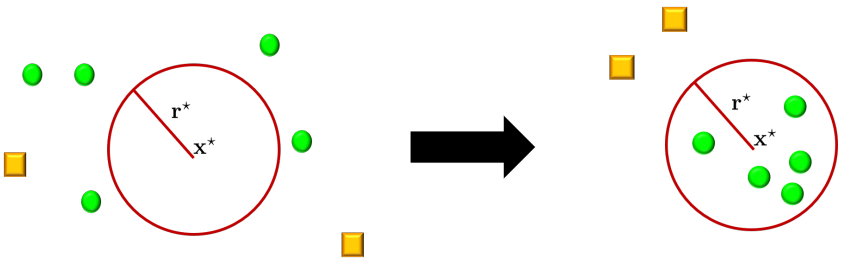


Figure 2.1: Graphical illustration of the herding task (problem) and its successful completion, with regard to $N_H = 2$ herders (yellow squares) tasked with corralling $N_T = 5$ targets (green circles) in a circular area of centre $\mathbf{x}^\star$ and radius $r^\star$.

Variable	Value	Description
$\mathbf{x}_i$		Cartesian position of target agent i
$\mathbf{x}_{gcm}$		Cartesian position of herd centre of mass
$\mathbf{x}^\star$		Cartesian position of goal position
$\tilde{\mathbf{x}}$		Cartesian position of a targeted agent
$\mathbf{y}_j^\star$		Desired trajectory of herder j
$\mathbf{d}_e$	$= \mathbf{y}_j^\star - \mathbf{y}_j$	Distance between herder j desired and current trajectory
$\mathbf{d}_{j,\star}$	$= \mathbf{y}_j - \mathbf{x}^\star$	Distance between herder j and goal $\mathbf{x}^\star$
$\mathbf{d}_{i,gcm}$	$= \mathbf{x}_i - \mathbf{x}_{gcm}$	Distance between target i and herd centre of mass
$\mathbf{d}_{j,gcm}$	$= \mathbf{y}_j - \mathbf{x}_{gcm}$	Distance between herder j and herd centre of mass
$K_{y1\rightarrow yN}$		Gains of herder dynamics (2.1)

Table 2.1: Notation and description of the main variables used to model multi-agent herding systems.

In most applications or proposed modelling approaches, the motion of targets (2.2) is the effect of interacting potential-fields centred around each herder and target agent, and environmental obstacles if present. These potential-fields then regulate the degree to which target agents are repelled from herders and environmental obstacles, as well as the degree of attraction between targets, if flocking is present. Importantly, this latter attractive force, independent of how it is actually implemented or modelled, captures the assumed tendency of target herds to converge into flocks, based on the theory of the selfish herd introduced by Hamilton [51], and later studied by King *et al.* [72]. Accordingly, such attraction is often hypothesised (explicitly or implicitly) to be fundamental to the herding problem and/or necessary for the success of herders. For instance, this is explicitly stated in [53, 67, 154], and implicitly conveyed elsewhere (e.g., [120, 151, 160]).

However, although this "flocking" assumption is representative of target herds in many real world contexts (either animals or people), assuming that it is fundamental or necessary limits the generalised applicability of many herding solutions. Indeed, the general "herding problem" is not constrained to only those situations were flocking occurs within natural systems. For instance, teachers who guide students on a field trip must compensate for their tendency to wander in a disor-

derly fashion, a drone could be tasked to herd unknown aerial vehicles outside an unsafe region, or robotic skimmers could be used to clean an oil spill. In such cases, the targets would not necessarily exhibit the tendency to flock and the unknown, and/or diffusive stochastic, dynamics exhibited should be integrated into the model (2.2).

A further limitation of assuming "flocking" behaviour has also been revealed from experiments conducted on the interaction of a flock of sheep with a robot sheepdog by Evered *et al.* [42]. Experimental results showed that, over repeated trials, the repulsive action of the robot herder on the animal targets (as consequently targets' tendency to aggregate in flocks) gradually dropped to the point sheep became accustomed to the robot herder.

While we highlight the limitations of assuming flocking behaviour, it is important to stress, though, that modelling the dynamics of targets is not the focus of this Thesis, and to overcome the above limitations, in Chapters 3-6 we will consider the most general case of targets (i) diffusing randomly in the environment and (ii) being repulsed from herder agents with a strength inversely proportional to their distance from them.

2.2 Overview of modelling approaches

Multi-agent herding problems can be classified based on a combination of the herding task goal, how target agents are modelled, and the relative number of herder and target agents. Importantly, the relative number of herder and target agents can affect the type of herding goal being achievable. For example, when herder agents are equal in number to the target agents (i.e., $N_H = N_T = 1$ or $N_H = N_T > 1$) the problem can be conceptualised as a pursuit-evasion problem; the pursuing (herder) agent attempts to track the evading (target) agent, with the goal being achieved once the evader is captured [26, 38, 69, 142]. In the case of herders exceeding the number of target agents, the task can be conceptualised as an encircling task. That is, the goal of herder agents is to spread around the target herd in order to build a repulsive "wall" around the herd, constraining the herd in an region within the circle of herders [22, 95, 102, 141, 159].

A possible way to classify herding tasks is via the choice of containment region variables; namely $\mathbf{x}^\star(t)$ and $r^\star(t)$ in (2.5). To illustrate this possible classification method, the mathematical formulation of a subset of possible task scenarios is reported in Table 2.2, with a summary of some possible corresponding classification parameters detailed in Table 2.3. In what follows, we will refer indifferently to the *herding task* or *problem* as the case of $N_H \geq 1$ herder agents corralling an equal or larger number of targets ($N_T \geq N_H$) to a fixed containment region ($\mathbf{x}^\star(t) = \mathbf{x}^\star$ and $r^\star(t) = r^\star$) and review the related relevant literature for multiple herders ($N_H \geq 2$) in Section 2.3 and for a single herder ($N_H = 1$) in Section 2.4. Relevant solutions of the herding task cast as a *pursuit-evasion problem* or an *encirclement problem* are further discussed in Section 2.5.

Task	Goal (2.5) definition	Task description
Containment	$\|\mathbf{x}_i(t) - \mathbf{x}^\star(t, \mathbf{x}_1, \ldots, \mathbf{x}_{N_T})\| \leq r^\star$	To corral N_T targets in a convex hull (e.g., around their centre of mass)
Herding	$\|\mathbf{x}_i(t) - \mathbf{x}^\star\| \leq r^\star$	To corral N_T targets around a fixed location
Navigation	$\|\mathbf{x}_i(t) - \mathbf{x}^\star(t)\| \leq r^\star$	To corral and drive N_T targets through around a moving location
Encirclement	$\|\mathbf{x}_i(t) - x^\star(t, \mathbf{x}_1, \ldots, \mathbf{x}_{N_T}, \mathbf{y}_1, \ldots, \mathbf{y}_{N_H})\| \leq r^\star$	To maintain corralled N_T targets in a region bounded by herder agents
Pursuit	$\|\mathbf{x}(t) - \mathbf{y}(t)\| \leq r^\star$	To capture evading target agents so that they remain within a set distance from herders

Table 2.2: Classification of tasks based on the definition of the desired final condition (2.5) for N_H herders and $N_T \geq N_H$ targets.

2.3 Herding solution for multiple herders ($N_H > 1$)

Of particular relevance in this Thesis is the case of *multiple* herders ($N_H > 1$) tasked to corral $N_T > N_H$ targets. In this complex task scenario, individual herder agents need to be designed so as to satisfy (2.5); at the same time, in order to achieve effective task outcomes, they must do so by coordinating their behavioural actions with that of the other herders. In other words, herders must implement decision-making strategies that consider the state of the targets, that of the other herders, and environmental constraints.

An intuitive solution is to rely on formation control and path planning techniques [66, 67, 120, 162]. This is the case when herder agents are considered as a platoon of robots (e.g., aerial, ground) and each robot follows a computed (either on-line or off-line) trajectory that will push targets in the predefined containment region as it will be discussed in Section 2.3.1.

Taking inspiration from natural systems, instead, a different and more cumbersome approach is to model each herder as a rigid body subject to virtual attractive and repulsive forces. The idea is that rigid bodies can repel each other as celestial bodies do, making it possible to translate behaviours like "obstacle avoidance" or "target chasing" through forces proportional to the relative distance between the herder and the agent (or object) of interest. Results reported in [53, 84, 107, 109, 129] all start from this modelling approach, with herders replicating dolphins, sheepdogs or pairs of people as discussed more in detail in Section 2.3.2.

2.3.1 Path planning and formation control solutions

Path planning solutions were one of the earliest proposed solutions to the herding problem with multiple herders. Motivated by a desire to better understand and solve the herding problem within realistic environments, Lien *et al.* ([67]) adopted a global rule-based roadmap approach, where

Total number of herder and target agents	
$N_H = 1$, $N_T > N_H$	[27, 38, 44, 65, 80, 87, 88, 113, 151, 154, 160, 164]
$N_H > 1$, $N_T > N_H$	[53, 66, 67, 84, 106, 107, 109, 120, 129]
$N_H = N_T$	[38, 69, 142]
$N_H > N_T$	[22, 23, 95, 102, 140, 141, 149, 159]
Task	
Containment	[53]
Herding	[65, 67, 80, 86–88, 113, 146, 151, 154, 160]
Navigation	
Encirclement	[22, 23, 95, 102, 140, 141, 149, 159]
Pursuit	[38, 69, 142]
Behaviour of target agents	
Flocking	[65, 67, 80, 113, 146, 151, 154, 160]
Velocity alignment	[53, 53, 67, 80, 113, 154]
Repulsion from herder	[53, 65, 67, 80, 86–88, 113, 146, 151, 154, 160]
Diffusive motion	[107, 109]
Herder knowledge of the environment	
Global	[86–88, 113, 146, 151, 160]
Partial	[65, 67, 154]
Modelling techniques	
Bio-inspired	[113, 151, 154, 160]
Path planning	[65–67]
Machine learning	[27, 44]
Genetic algorithm	[80, 146]
(Optimal/Non-linear) control design	[86–88, 120]
Validation	
Computational	[11, 53, 65, 67, 80, 99, 113, 146, 151, 154]
Experimental	[66, 87, 107, 109, 113, 120, 151, 160]
Formal	[38, 86–88, 120]

Table 2.3: Taxonomy of literature on the multi-agent herding problem for single and multiple herders.

walkable paths are represented as edges of a directed graph [170], to model the trajectories of herder and target agents in an environment that contains obstacles. The roadmaps associated with the target agents were coupled through repulsive and attractive forces to the herders and to neighbouring targets, respectively. For herder agents, instead, the solution entailed three different strategies to approach the back of the herd, i.e., for them to reach one of the available nodes in the roadmap; moving (i) towards the first available node, or (ii) towards the first node both available and closest, or (iii) pairing herders and available nodes by solving a global minimisation problem with respect to the distance to be travelled by the herders (see Figure 2.2).

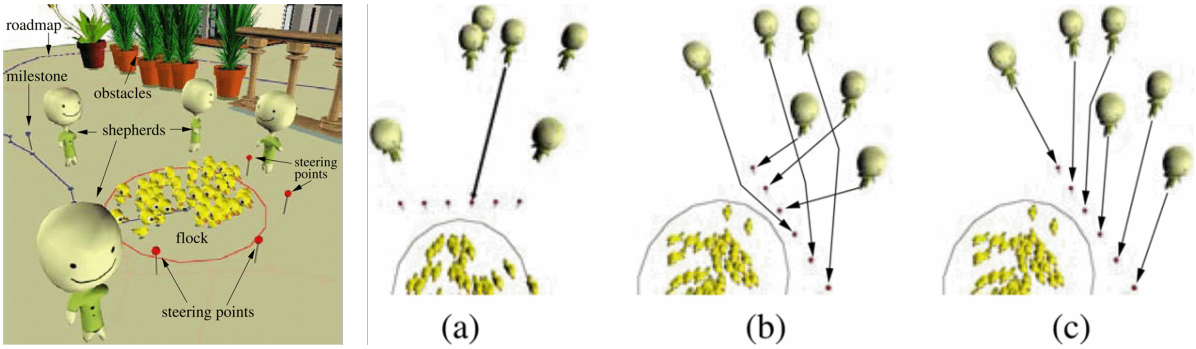


Figure 2.2: Herding environment considered by Lien *et al.* with multiple herders choosing (a) the first available node in the dynamic roadmap, (b) the first both available and closest and (c) by solving a global minimisation problem. Image reproduced from [67].

As expected, numerical experiments proved that multiple herders (up to five) were more successful than $N_H = 1$ herder in coping with both increasing herd sizes and, of particular importance, when environmental obstacles are present.

However, when, to test the robustness of the herding system developed, herders were tasked to herd targets of varying flocking tendency, even the best combination of number of herders and approaching strategy failed if faced with low flocking targets (similarly to [154]). To overcome such correlation of herding effectiveness with targets flocking behaviour, in later work, two different approaches to design dynamic roadmaps were pursued; "human-in-the-loop" design [66] and probabilistic motion planning techniques [162]. In the first case, human actors, equipped with laser pointers and a projected image of the environment, were asked to interactively modify the roadmap, generated by the algorithm, to help herders solve the task. In the second case, roadmaps were designed by training Rapidly Exploring Random Trees (RRTs) and Expansive-Spaces Trees (ESTs) to achieve a more extensive exploration of all the possible walkable paths. Numerical experiments showed that herders following an interactive roadmap [66] performed better when the number of targets exceeded $N_T = 20$ agents but the more extensive exploration of the environment, achieved through the probabilistic roadmaps [162], resulted in higher success rates.

The research conducted by Lien *et al.* proved the effectiveness of adopting path planning techniques to find the most suitable trajectory for the herders. However, the need for a priori knowledge

of the environment (e.g., to train RRTs and ESTs) and the computational cost of solving online optimisation problems opened the path to explore design approaches less constrained from a priori-knowledge and requiring a smaller computational load.

Recently, Pierson *et al.* proposed an arc-based formation control strategy for N_H herders to surround and drive N_T targets towards a desired goal region [120]. Each herder regulates its position to a desired one ($\mathbf{y}_j^\star(t)$) in the arc behind the herd at a distance r from the centre of mass of the herd $\mathbf{x}_{gcm}$ (Figure 2.3) by means of the following non-linear feedback controller,

$$\dot{\mathbf{y}}_j(t) = -K_{y1}(\mathbf{y}_j^\star(t) - \mathbf{y}_j(t)) \tag{2.6}$$

with

$$\mathbf{y}_j^\star(t) = \mathbf{x}_{gcm}(t) + r\begin{bmatrix}\cos(\psi+\Delta) & -\sin(\psi+\Delta)\end{bmatrix}^{\mathrm{T}} \tag{2.7}$$

The inter-herders angular distance Δ and the rotation ψ of the arc behind the herd were regulated by means of an additional proportional feedback controller.

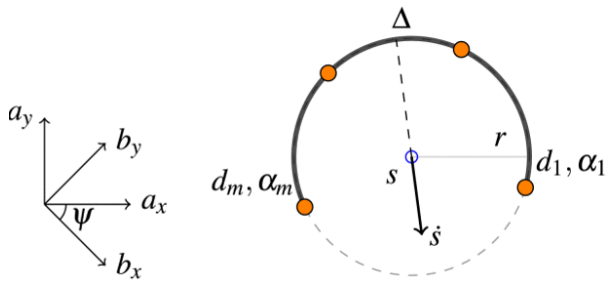


Figure 2.3: Multi agent herding system studied in [120]. $N_H = m$ herder agents are designed to position themselves at a distance r behind the flock (identified with s) with an angular displacement Δ. Image reproduced from [120].

The achievement of the herding task (2.5) was then demonstrated formally, numerically and experimentally. Indeed, a Lyapunov-based analysis was used to prove that the trajectory of the herders was able to corral the herd into the containment region (as long as the herd could be assumed to be all concentrated around their centre of mass). Such assumption clearly translates in considering, again, only targets that flock. The approach by Pierson *et al.* was one of the first attempts at adopting feedback control strategies to the herding problem, and more importantly to formally prove its convergence. Nevertheless, the behaviour of such N_H herders was tested (i) numerically, against $N_T < N_H$ targets and (ii) experimentally, against a single target ($N_T = 1$). That is, in the best case scenarios of herders outnumbering target agents.

Similar feedback control strategies were proposed in [88] to control a single herder tasked to corral $N_T > N_H$ targets. The benefits of this approach is that a formal proof of convergence (in both cases a Lyapunov stability analysis) of the overall system is achievable but strictly depends on the model of targets dynamics (2.2). Later, in Chapter 3, we will test the herding models proposed by [88,

120] against targets that wander randomly and compare herding performance with herders driven by local control rules. Comparison results will further highlight how target modelling is key to the success of the majority of herding solution, also when success (i.e., satisfaction of (2.5)) is formally proven.

2.3.2 Bio-inspired models

Inspired by the limited visual field of real sheepdogs and the absence of centralised coordination among them, Lee and Kim proposed a herding algorithm based entirely on local attraction rules [84]. Here, $N_H \geq 1$ herders are required to both corral and contain a group of $N_T > N_H$ target agents. Herders and targets are wheeled robots equipped with proximity sensors of limited range, with their dynamics modelled as the linear combination of potential field-like forces within a sensing area [125]. Both herders and targets were attracted to neighbouring agents, tended to match the motion of the agent (herder or target) closest to them, and were repelled away from obstacles. In addition to this, targets were also subject to a repulsive force away from the herders. Moreover, herders were subject to a control input force that is a function of both their distance from the position of the nearest target agent ($\mathbf{x}_t$) and their distance from a desired point ($\tilde{\mathbf{x}}$). More formally, the dynamics of the herders is given by:

$$\dot{\mathbf{y}}_j(t) = f(t, \mathbf{x}_1, \ldots, \mathbf{x}_{N_T}, \mathbf{y}_1, \ldots, \mathbf{y}_{N_H}) + K_{y1} \frac{(\mathbf{x}_t(t) - \mathbf{y}_j(t)) - (\tilde{\mathbf{x}}(t) - \mathbf{x}_t(t))}{\|(\mathbf{x}_t(t) - \mathbf{y}_j(t)) - (\tilde{\mathbf{x}}(t) - \mathbf{x}_t(t)\|} \tag{2.8}$$

Depending on heuristics based on the relative distance between each herder and the targets within its sensing area, $\tilde{\mathbf{x}}(t)$ is selected as the position of the centre of the flock ($\tilde{\mathbf{x}}(t) = \mathbf{x}_{gcm}(t)$), the position of the chased target ($\tilde{\mathbf{x}}(t) = \mathbf{x}_t(t)$) or a fixed position ($\tilde{\mathbf{x}}(t) = \mathbf{x}^\star$). Based on the choice of $\tilde{\mathbf{x}}(t)$, Lee *et al.* defined the task to solve as patrolling, collecting or herding respectively.

Interestingly, the same target agent can be targeted by multiple herders, meaning that the decision making of each herder is independent from that of the others and can overlap, leaving the cooperation among herders being only a consequence of their modelling tendency to match velocity with the closest neighbouring herder agent. The result is the emergence of an arc formation among the herders with herders able to corral and herd multiple sub-flocks without any explicit coordination rule – a similar formation instead had to be hard-coded into the algorithm presented in [67] and was the desired formation control goal of Pierson *et al.* in [120].

Shepherds and sheepdogs corralling sheep is not the only natural system example of herding. A similar collective behaviour is observed when dolphins forage. In this case, dolphins (the herders) move to entrap target agents in a region from which they can not escape. Indeed, taking inspiration from such technique, Haque *et al.* designed herder agents with some pre-assigned region of influence [52, 53], assuming that targets' motion was only influenced by a specific herder if they happened to be within its region of influence (similar to [151]). When not influenced by a herder, targets travelled at a constant speed, with a heading direction aligned to that of their neighbouring

targets. The velocities of the herders were regulated according to the estimated time $t^\star$ necessary for the targets to escape the region of influence, arranging themselves in two opposite rows or in a carousel as in Figure 2.4. For example, when in carousel arrangement (Figure 2.4(b)), the j-th herder dynamics (2.1) is modelled as a function of (i) the radius of the containment region $r^\star$, (ii) the desired position $\Theta^\star_{i,j}$ where herder j should be at time $t^\star_{i,j}$ in order to corral target i, and (iii) the heading θ_i of target i:

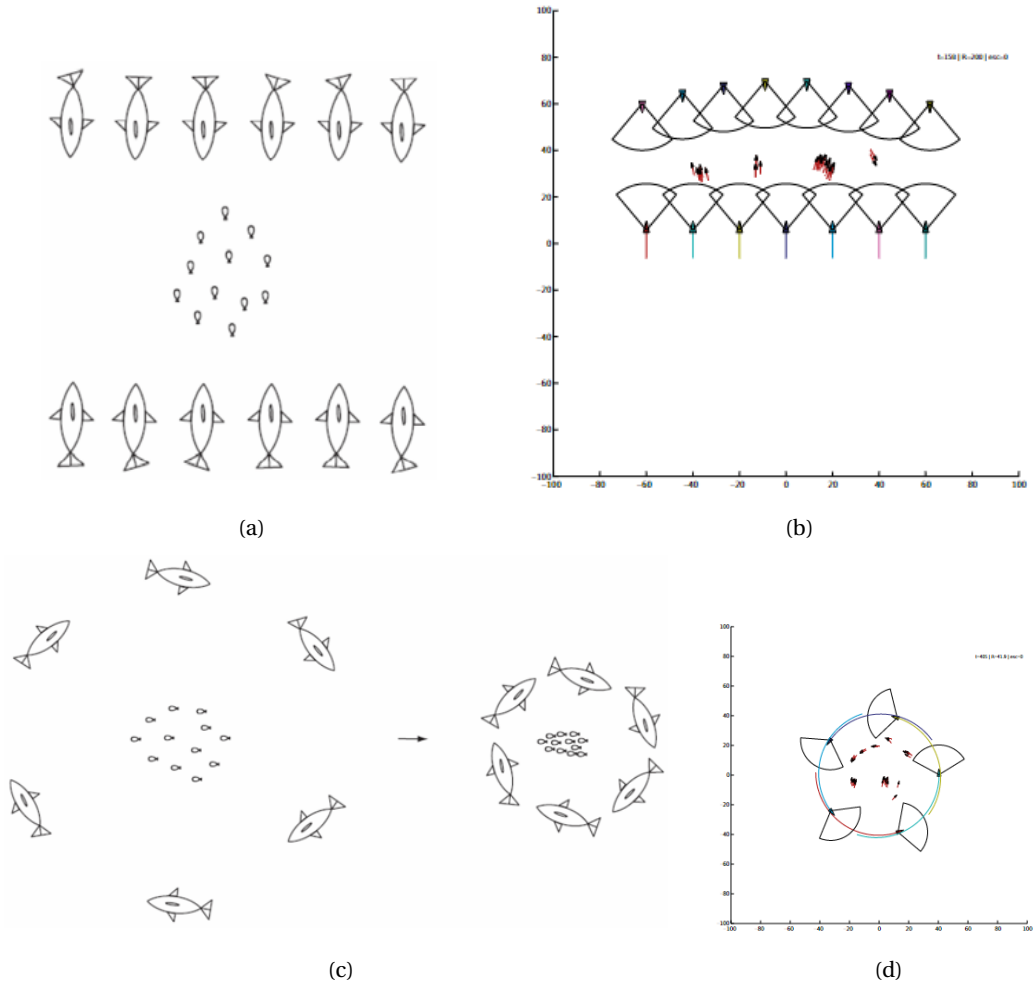
$$\dot{\mathbf{y}}_j(t) = \begin{cases} 2r^\star \dfrac{\Theta^\star_{i,j} - \theta_i}{t^\star_{i,j} - t} & \text{if } \Theta^\star - \theta > 0 \\ \\ 2r^\star \dfrac{2\pi + \Theta^\star_{i,j} - \theta_i}{t^\star_{i,j} - t} & \text{otherwise} \end{cases} \tag{2.9}$$

Note that, when a herder reaches target i, a new to-be-corralled target is selected, ensuring $t^\star_{i,j} \neq t$. The regions of successful containment were defined by optimising herders velocity on targets escaping time and distance from a possible exit point. That is, for which initial positions of targets and herders, the latter were able to contain the former.

Innovatively, the desired herding trajectory was not only designed assuming targets would escape from herders, but would do so adopting the most efficient escape strategy. Indeed, to the best of our knowledge, the work from Haque *et al.* is still the main research work considering targets able to perform "smart" manoeuvres against the herders.

Recently, herding tasks have also received growing attention within the cognitive and psychological sciences [106, 107, 109, 110, 123, 129] with multi-player herding tasks being used as a representative task for exploring the behavioural dynamics that underlie complex joint-action and team perceptual-motor coordination. Most relevant here is the herding task explored in [106, 109, 129], in which pairs of human players control virtual herder agents or avatars to corral a small herd of virtual sheep or cattle (targets) into a specified containment area within a large game field. The task can be played on a large tabletop display screen (as in Figure 2.5(a)) or in an immersive 3D virtual-reality environment. The human-controlled herder agents essentially act as 'sheepdogs', with the target agents repelled away from the herder agents when the targets come within a certain distance of a herder. When not influenced by a herder agent, the target agents exhibit Brownian motion, diffusely wandering around the game field.

To complete the task successfully, human herders learned to adopt two modes of behavioural coordination. The first, referred to as search-and-recover or S&R behaviour, involves human herders dynamically dividing the game area (more-or-less) into two regions of responsibility and then corralling the targets that were approximately furthest from the containment area within their current "sphere of responsibility". In other words, during S&R behaviour human herders essentially moved from "furthest target" to "furthest target within their current region of responsibility", coordinating their actions until all the targets were within the containment area. Once all targets were con-



(a) (b) (c) (d)

Figure 2.4: Multi agent herding system inspired by foraging techniques of dolphins consisting in (a) forming two opposite rows or (c) creating a carousel around the herd. Panels (b)-(d) are an example of how simulated herders replicates dolphins herding behaviour. Images reproduced from [53].

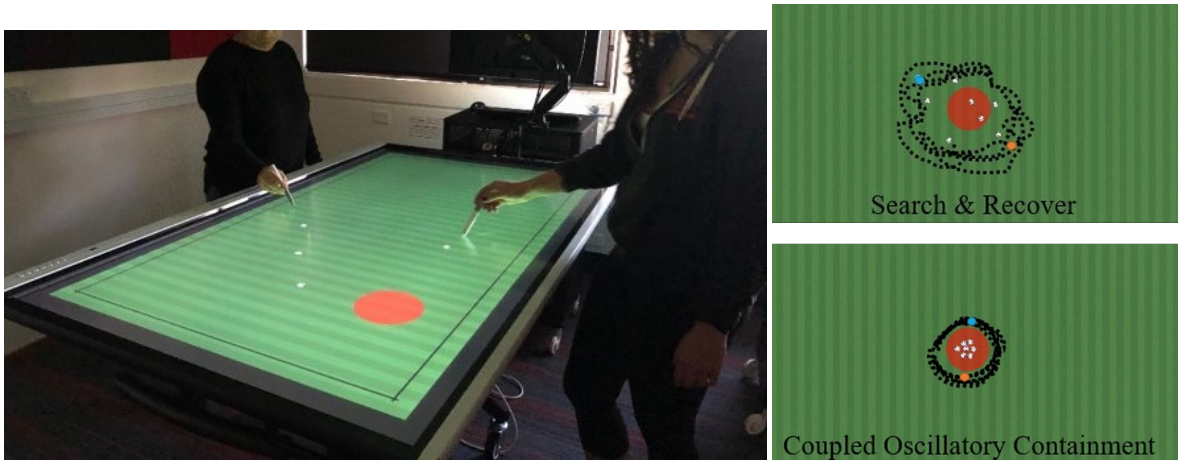
tained, some human teams also learned to switch to an oscillatory containment mode of behaviour, in which the players perform oscillatory movements around the herd, forming a "repulsive wall" around the herd to keep it contained (see Figure 2.5(b) for a graphical example of these emerging behaviours).

Of particular significance here, is that research exploring this joint-action task has demonstrated how the movement dynamics exhibited by human herders can be modelled using a simple, environmentally coupled, non-linear dynamical model of the general form

$$r_j(t) = -b_r \dot{r}_j(t) - R(\tilde{\mathbf{x}}_{k,j}, t), \tag{2.10}$$

$$\theta_j(t) = -b_{\theta,j}(t)\dot{\theta}_j(t) - T(\tilde{\mathbf{x}}_{k,j}, t) + O_{\mathrm{R,VP}}(t) + (1-\xi_j), \tag{2.11}$$

Here, the planar movements of the human-controlled herder agents are defined in polar coordi-



(a) (b)

Figure 2.5: Herding task and experimental set-up (a) used in [106, 109, 129] to model the S&R and oscillating behaviour (b) observed in human teams.

nates, such that the S&R behaviour is modelled using non-linear mass-spring terms $R(\tilde{\mathbf{x}}_{k,j}, t)$ and $T(\tilde{\mathbf{x}}_{k,j}, t)$ that attract (move) the herders radial and angular position $\mathbf{y}_j(t) = r_j\, e^{\mathrm{i}\theta_j}$ to the location of a specified target $\tilde{\mathbf{x}}_{k,j}(t) = \tilde{\rho}_{k,j}\, e^{\mathrm{i}\tilde{\phi}_{k,j}}$ (plus a corralling offset or preferred corralling distance). Once the targets are within the containment region, the non-linear Rayleigh-like and/or Van der Pol-like oscillator terms $O_{\mathrm{R,VP}}(t)$ in (2.11) and a parametric Hopf-bifurcation [55] process causes the model to produce oscillatory movements around the targeted herd (see e.g., [108, 109, 117]). Note that the model (2.10)-(2.11) will be further analysed in Chapter 3 and Appendix A; there, each of the terms $R(\tilde{\mathbf{x}}_{k,j}, t)$, $T(\tilde{\mathbf{x}}_{k,j}, t)$, $b_{\theta,j}(t)$ and $O_{\mathrm{R,VP}}(t)$ is individually defined and discussed.

In addition to producing behavioural movements comparable to human-controlled herders, research demonstrated how this dynamical model can also be used to control the behaviour of artificial herder agents that are not only able to effectively complete the task with human co-actors, but can also be used to train novice human players to achieve expert task performance at a rate similar to that observed when novice players are trained by an expert human trainer [109, 129].

2.4 Herding solutions for single herders ($N_H = 1$)

The multi-herder herding task is of particular relevance when designing or modelling agents able to coordinate in skilful joint task but it is important to report how some of the earliest solutions to the herding task [65, 151, 160] focused on the problem of a *single* herder ($N_H = 1$) needing to corral either a small or larger herd (or flock) of target agents. In this scenario, the main design concern was the herder's ability to leverage the targets' intrinsic characteristics when corralling the herd, either using attractive-repulsive forces [151, 154, 160] or path-planning techniques [65, 113] (that were later adopted in the case of multiple herders as discussed in Section 2.3 above), evolutionary

algorithms [80, 146], or via switching time models [88] that regulated herder engagement with each targeted agent and guaranteeing the satisfaction of (2.5). Here, we review each approach in turn.

2.4.1 Bio-inspired attraction force models

In one of the first approaches to solve the herding problem, Vaughan *et al.* aimed in [160] at designing a robot-dog able to interact with target animals that exhibit flocking tendencies by exploiting targets reaction to external agents. Targets' reaction to the robot, and their flocking tendency, were modelled using potential field forces ([125]) as

$$\dot{\mathbf{x}}_i = \sum_{k=1, n\neq i, k\neq i}^{N_T} \left[\left(\frac{K_{x1}}{(\|\mathbf{d}_{i,k}\| - R_n)^2} - \frac{K_{x2}}{\|\mathbf{x}_i\|^2} \right) \widehat{\mathbf{d}}_{i,k} \right] - \frac{K_{x3}}{\|\mathbf{d}_{i,w}\|^2} \widehat{\mathbf{d}}_{i,w} - \frac{K_{x4}}{\|\mathbf{d}_{i,j}\|^2} \widehat{\mathbf{d}}_{i,j} \tag{2.12}$$

where, $K_{x1}, K_{x2}, K_{x3}, K_{x4}$ are non-negative gains, $R_n > 0$ is a repulsion radius to avoid collision with neighbouring targets and $\mathbf{d}_{i,w}$ is the distance between target i and the nearest point on the wall ($\mathbf{w}$) of the arena in which herder and targets are assumed to move. Note that $\widehat{\mathbf{a}}$ denotes the unit vector of $\mathbf{a}$. Letting $\mathbf{x}_{gcm}$ to represent the centre of mass of the flock, the herder dynamics was defined as

$$\begin{aligned} \dot{\mathbf{y}}(t) &= f(t, \mathbf{x}_1, \ldots, \mathbf{x}_{N_T}, \mathbf{y}) \\ &= \left(K_{y1} \|\mathbf{x}_{gcm}(t) - \mathbf{x}^\star\| \right) \widehat{\mathbf{d}}_{j,gcm}(t) - K_{y2} \widehat{\mathbf{d}}_{j,\star}(t) \, . \end{aligned} \tag{2.13}$$

That is, a feedback proportional controller regulated a herder's position, with control gains (K_{y1}, K_{y2}) modulating how far the herder should be from the flock ($\mathbf{d}_{j,gcm}(t) = \mathbf{y}(t) - \mathbf{x}_{gcm}(t)$) and how close the latter was kept from the desired herding goal position ($\mathbf{d}_{j,\star}(t) = \mathbf{y}(t) - \mathbf{x}^\star$).

The algorithm, tested in both simulation and real world experiments, was used to successfully drive a mobile robot to corral a small flock of geese in a fenced circular arena as can be seen in Figure 2.6. The robot-herder (which could move faster than the geese) collected information about

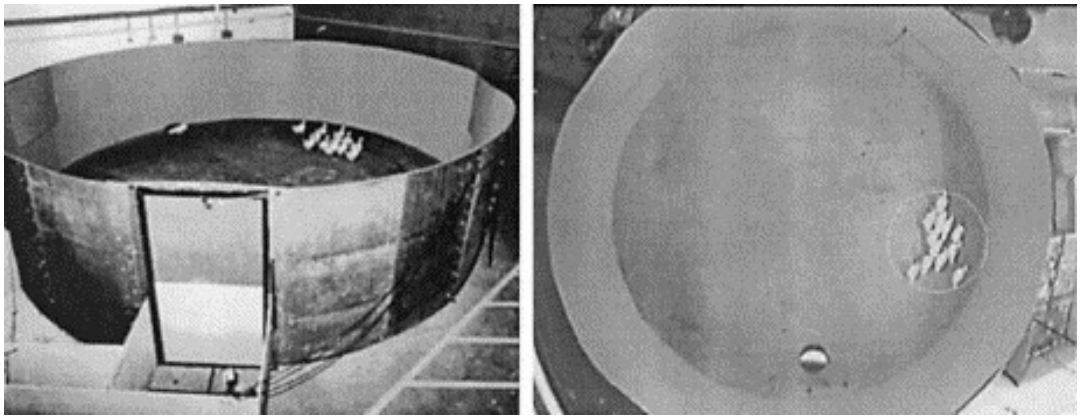


Figure 2.6: Experimental set-up considered in [160]. One robotic herder agent is tasked to herd a flock of geese towards a goal point near the wall of the arena.

the environment through processed video signals. These reference signals were, then, sent to an

external workstation that computed the robot's desired trajectory and subsequently re-transmitted trajectory updates by radio to the herder as illustrated in Figure 2.6. The flock, captured on camera, was considered as one object of varying shape, with the trajectory of its centre $\mathbf{x}_{gcm}(t)$ fed to the control algorithm (2.13) to achieve the goal defined in (2.5).

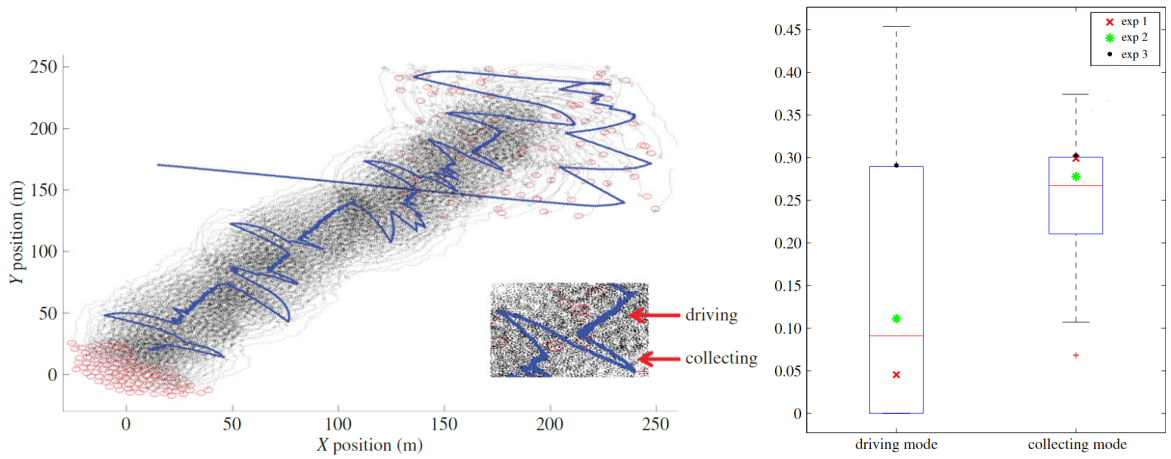
As a first attempt at controlling a multi-agent system by the action of one external agent, Vaughan *et al.* highlighted several limitations in their approach that remained to be addressed by future researchers. For instance, the absence of obstacles in the testing arena (later addressed in [65] for a single herder and in [67, 121] for multiple herders), the reduced size of the to-be-controlled flock (later addressed by [151]) and the difficulty of solving the task when the goal position was far from arena's walls (later addressed in [67, 88]).

Motivated by Vaughan *et al.*'s work ([160]) with real animals, Strömbon *et al.* attempted to model the "driving and collecting" behaviour observed in real sheepdogs [151]. At each time step, the dynamics of a sheepdog herder was designed so as to make it move at a constant speed in a straight line towards a heuristically chosen goal point $\tilde{\mathbf{x}}(t)$ in the herding arena. With R_{in} defining an influence radius surrounding the herder and recalling that N_T is the total number of target agents, when all targets were within a distance $R_{in}N_T^{2/3}$ from the herd centre of mass (i.e., targets were subject to flocking attraction), the herder would move towards $\tilde{\mathbf{x}}(t) = \mathbf{x}_{gcm}(t)$ "driving" the herd towards the goal point. Otherwise the herder would move towards the farthest target from the herd ($\tilde{\mathbf{x}}(t) = \mathbf{x}_k(t) \ :| \ \|\mathbf{d}_{k,gcm}\| = max_i\|\mathbf{d}_{i,gcm}\|, \ \forall i$), "collecting" all stray targets one at a time. To realistically replicate a sheepdog careful approach to target herds or flocks, when the distance between the herder and the targeted point $\tilde{\mathbf{x}}$ becomes smaller the herder's influence radius R_{in} is scaled down to $0.3R_{in}$.

As can be seen from Figure 2.7(a), the resulting model produces a herder that spontaneously switches from a "side to side" quasi-oscillating motion when collecting targets to straight line trajectories when driving the herd. Note that the quasi-oscillating movements are not directly coded into the model, but are an emergent property of the multiagent system.

As in Vaughan's case, targets for the model in [151] were simultaneously attracted towards each other (i.e., exhibited a flocking tendency) and repulsed directly away from the herder. This collective target behaviour led the herder to possibly stall in local minima between two or more sub-flocks during the collecting phase. To avoid such events, if the number of neighbouring target agents to flock is below a certain threshold, the herder targeted the local centre of mass of each sub-flock ($\tilde{\mathbf{x}}(t) = \mathbf{x}_{lcm}(t)$).

Strömbon *et al.* analysed the model for different flock sizes and, then, compared it with data collected in the real world; the recorded actions of a sheepdog and a herd of sheep, with GPS trackers mounted on their collars. The comparison of numerical simulations and experimental results showed that the herding model was able to successfully capture the time spent by the real shepherd dogs in the "driving" or "collecting" condition (Figure 2.7(b)) clearly demonstrating how global



(a) (b)

Figure 2.7: (a) "Side to side" emerging trajectory performed of the herder agent presented in [151] while herding $N_T = 100$ target agents and (b) comparison time spent "collecting" and "driving" by herder model (box plots) and a real sheepdogs (coloured markers) .

emerging behaviour could be replicated using simple rule-based modelling approaches.

It is worth noting that the model presented in [151] has been more recently used in a gamified version of the herding task for investigating herding strategies of human players [164].

Finally, it is important to appreciate that the above approaches to solving the herding problem using virtual attraction/repulsion forces assume that herder agents have global knowledge of the herding environment (and of the agents within). Tsunoda *et al.* questioned whether this global knowledge of the environment was necessary given the assumption of target flocking behaviour, [154]. To test this, an advanced version of (2.13) was implemented in [154], where the herder was allowed to target the farthest agent from the goal instead of the herd's centre of mass. Extensive numerical simulations of this updated model, coupled with targets' behaviour (i.e., flocking, velocity matching, collision avoidance [30, 125]) allowed the herder to succeed and to outperform both the original model (2.13) ([160]) and the herding model proposed by Strömbon *et al.* [151]. Not surprisingly, the updated model was particular effective in conditions where the herder's visibility is limited, which often occurs when the targets agents farthest away from the herder are hidden behind those closest to it (see Figure 2.8).

2.4.2 Path planning solutions

Motivated by a desire to better understand and solve the herding problem within realistic environments, in [65] Lien *et al.* adopted a global rule-based roadmap approach – abstract representations of walkable paths as a directed graph [170] – to model the trajectories of one herder agent and multiple target agents in an environment that contained obstacles. The roadmaps associated with the

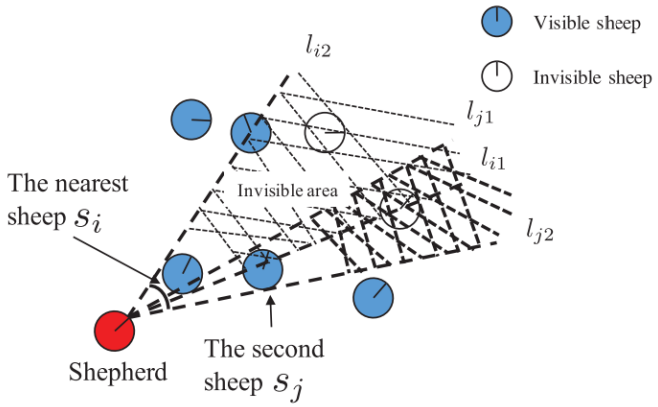


Figure 2.8: Limited field vision considered by Tsunoda *et al.* in [154].

target agents were coupled with repulsive and attractive dynamics to the herder and neighbouring targets, respectively. The nodes of the herders' roadmap changed accordingly to several considerations associated with the "to-be" travelled path and flock position. When a herder "approached" the flock, the herder's path would be planned by making a node appear near the target and disappear when reached by the herder. Once the flock was approached, the "steering" points were computed through pattern formation techniques allowing the herder to perform multiple tasks; for example herding, patrolling or collecting (see Figure 2.2 for a graphical representation of the herding algorithm [65] when extended to multiple herders). The resulting model reportedly enabled the herder to handle large flocks (i.e., up to $N_T = 50$ target agents) and different types of target agents (e.g., exhibiting various flocking tendencies).

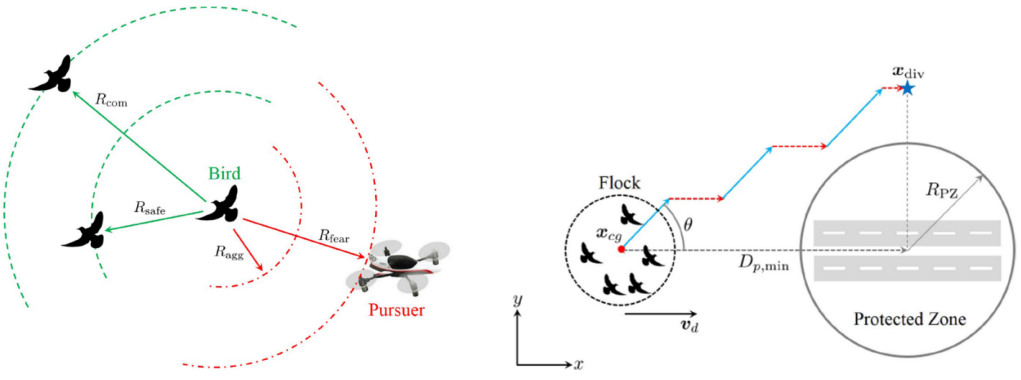
Unfortunately, replicating the apparent effectiveness of the model proposed in [65] proved problematic due to the lack of formal information (i.e., mathematical formulation) of the implemented algorithms therein. Accordingly, Bennet *et al.* attempted a reconstruction of Lien's herder model with the goal at providing a simulator software package for herding algorithms [11]. The developed simulator was then used to compare Lien's ([65]) and Vaughan's ([160]) herder models with the rule based model proposed by Miki *et al.* in [99]. Miki *et al.* modelling approach was to let herder and target agents distinguish two layers of concentric interacting zones around themselves; an inner "handling zone" and a external "watching zone" for the herder, an inner "personal zone" and a external "safety zone" for the targets. Agents reacted according to simple heuristic rules based on the events in one or in both regions being entered by any other agent. As long as no herder agents enters the safety zone of a target, this would only flock with neighbouring targets inside its personal zone. At the same time, a herder would actively herd only targets within their handling zone, and search for the next target to corral among the agents within its watching zone.

Most recently, Paranjape *et al.* [113] has adopted a path planning solution to the problem of

driving an Unmanned Aerial Vehicle (UAV) to divert animals (i.e., birds) from unsafe zones. More specifically, the task goal (2.5) was for a UAV to drive targets towards a point from where the flock could safely resume its flight (see Figure 2.9). Targets were modelled through the composition of Reynolds flocking model [125] and an "evasive response" depending on herder proximity. That is, each target agent was surrounded by sensing areas of different sizes depending on the type of interaction with other agents as illustrated in Figure 2.9(a) (e.g., velocity alignment with flock members, repulsion from the herder). [113] then designed a boundary control strategy called the "n-waypoint" herding algorithm (see Figure 2.9(b)), with a herder's trajectory computed to push targets towards sub-goals – the waypoints – sampled from a set of permissible positions $\mathscr{X}_p$ the herder could occupy to divert the flock without scattering it. The herder would either choose the fastest path FLY to the next waypoint $\mathbf{x}_p \in \mathscr{X}_p$ or, once reached, maintain the same position and ENGAGE with the flock for a predefined dwell time τ_e

$$\dot{\mathbf{y}}(t) = \begin{cases} \text{FLY} & \text{if } \|\mathbf{y}(t) - \mathbf{x}_p(t)\| \geq \epsilon \\ \text{ENGAGE} & \text{otherwise} \end{cases}. \tag{2.14}$$

Assuming targets as nodes of a communication graph and by ignoring herder-related terms in targets dynamics, Paranjape *et al.* exploited local topological considerations to find the set of the permissible positions $\mathscr{X}_p$. This involved balancing the need to avoid colliding with other targets and repulsion from the herder were balanced w.r.t. (i) the time necessary for targets to fly between two waypoints τ_f, (ii) their engagement time with the herder τ_e, and (iii) goal point distance $\mathbf{d}_{i,\star}(t) = \mathbf{x}_i(t) - \mathbf{x}^\star$.



(a) (b)

Figure 2.9: Multi-agent herding system considered in [113]. Various sensing areas (a) influence targets motion (b).

Numerical simulations with different flock sizes and aggregation strength validated the proposed herding algorithm. Innovatively, parameters of the flocking model (e.g., flock velocity, influence radii) and gains of the herding algorithm were estimated on real experiments with two flocks

of birds and a drone. However, these experiments also confirmed that the success of [113] remains strictly linked to herder's ability to avoid flock scattering or sudden diversions in their flight and, moreover, is dependent on a herder's ability to identify and exploit flocking behaviour of target agents.

2.4.3 Non-linear feedback controllers

In Robotics, recent work by Licitra *et al.* employed a backstepping control strategy for a herder to chase and corral one target at a time [88]. Herders were driven by the non-linear controller $\mathbf{u}(t)$ designed using the recursive technique based on a Lyapunov function of the from [71, 78];

$$\dot{\mathbf{y}}(t) = K_{y1}\mathbf{d}_e(t) + K_{y2}\tilde{\mathbf{x}}_i(t) + K_{y3}\|\tilde{\mathbf{x}}_i(t) - \mathbf{y}(t)\| + K_{y4}\|\tilde{\mathbf{x}}_i(t)\|. \tag{2.15}$$

The controller minimises the distance $\mathbf{d}_e(t) = \mathbf{y}^\star(t) - \mathbf{y}(t)$ between the desired and actual herder trajectory, while the the desired trajectory, $\mathbf{y}^\star(t) = K_y\tilde{\mathbf{x}}_i(t) + \mathbf{x}^\star$ minimises the distance between the targeted agent and the centre of the goal region. The herder switches among different targets at a regular interval of time, analytically derived to guarantee herding success (i.e., in collecting the herd within the goal region of interest). This idea was further developed in [86, 87] where other control strategies and further uncertainties in the herd's dynamics were investigated. For instance, [87], proposed an adaptive control strategy, based on integral concurrent learning and varying lower-bounded switching time, ensured global containment of targets whose unknown interactions with the herder were later approximated with Artificial Neural Networks (ANNs) [86]. This model was able to exploit ANNs and dwell time analysis (i.e., the time the herder engage with a target before switching) to show how it is possible, for a single herder, to herd multiple targets.

Let us emphasise that, differently from previous approaches, such herding solutions do not strictly depend on aggregation tendency of targets but only assume that (i) targets are repulsed from the herder when close or (ii) maintain a stationary position otherwise. The feedback control approach is then used to regulates herder trajectory accordingly to both the desired herding goal and targets trajectory.

2.4.4 Evolutionary and genetic algorithms approaches

Herders modelled using attraction forces in [151, 160] have also been recently improved using evolutionary computation [80, 146]. In justifying this approach, Lakshika *et al.* started from the idea that the two type of agents in the herding problem are semi-competitive [80].The authors argued that, different from a classical prey-predator scenario [13], for the herding problem the win of one group of agents does not necessarily implied the loss of the other. Thus, their behaviours could be "co-evolved" through genetic algorithms. Scope of such co-evolution is to find the best combination of the set of spatial rule (e.g., distance between a target and a herder) and the temporal rule (e.g., a triggering event) of herders and targets.

More specifically, these spatial and temporal rules regulate agents' velocity and trigger the next action, respectively. Genetic co-evolution was then leveraged to find the optimal linear combination of these spatial rules, their corresponding weights and the parameters regulating the temporal rules. The targets spatial rules were taken from Vaughan's and Reynolds models [125, 160]. The targets temporal rule corresponded to the target's responsiveness to the behaviour of herders and included the real world assumption that "sometimes sheep stop to eat grass without responding to the dog" [80]. Herder's spatial rules were defined by (2.13) from [160], while the herder's temporal rule took into account the time interval in which the herder would not actively chase targets.

Intuitively, a herder's evolutionary objective is to corral and drive a herd to a goal position in the shortest possible time. However, the herd evolutionary objective function devised by Lakshika *et al.*, was instead designed to enhance co-evolution and realistic interactions (e.g., flocking with other targets, fleeing the herder). Post model evolution, overall model performance was therefore subjectively evaluated by asking human actors to rate the most life-like co-evolved interactions among (i) the ones proposed by the authors, (ii) a set of ad-hoc designed ones and (iii) a replica of Vaughan's original model [160]. The results revealed that the top ranked spatial and temporal rules involved the use, as objective function for target agents, a combination of their tendency to flock and an enhanced desire to escape from the herder if this enters their personal space (i.e., a "flight" zone).

More recently, Singh *et al.* leveraged similar genetic algorithms on Strömbon's model [151] to replace the constant speed of the original model with a variable velocity [146]. In this study, the herder agent was tasked to achieve the herding goal (2.5) while minimizing its energy consumption and task-completion time. In comparison to [151], the evolved herder agent achieved higher success rate; especially for sufficiently smooth variations in herder's velocity.

2.5 Control solutions for pursuit-evasion and encirclement herding tasks

As anticipated, another approach is to frame the task as a *pursuit-evasion game*, as done for example by [38, 69, 142]. Here, the case of one target agent ($N_T = 1$) evading one pursuer, the herder ($N_H = 1$), is solved by computing off-line the optimal solution of a dynamic programming problem. More specifically, the solution of an algorithm based on Dijkstra's shortest path was proposed in [69] to solve the herding problem on a grid while Shedied ([142]) framed the design of herder's trajectory on a plane as a constrained optimization problem, particularizing the solution obtained to the dynamics of non-holonomic vehicles (later considered for multiple herders in [120]). A dynamic programming solution was also applied in [38] at the herding scenario in [87], here both the dynamics of the single herder and the one target were assumed unknown.

A particular set of solutions for multiple herders performing *encirclement herding tasks* have

been proposed to distribute herder agents around the herd to be corralled. Attempts at designing such dynamical encirclement, for aerial robots, included leveraged non linear control techniques like Lyapunov theory [141] or decentralized model predictive control [95]. A herding strategy based on elliptical orbits to entrap a target agent whose position is uncertain was proposed by Montijano *et al.* ([102]) while both Varava *et al.* developed a "herding by caging" solution, based on geometric considerations and motion planning techniques to arrange the herders around the herd [149, 159]. A similar formation was presented in [23], and further developed in [22], to let herders identify clusters of flocking adversarial agents, dynamically encircle and drive them to a safe zone. Recently, Sebastian *et al.* developed analytical and numerical control design procedures to compute suitable herding actions to herd evading agents to a desired position, even when the non-linearities in the evaders' dynamics yield to implicit equations [140].

2.6 Summary

In this Chapter, we provided an overview of the existing literature on multi-agent herding systems. The last two decades have seen researchers from a variety of different fields (e.g., mathematics, engineering, cognitive and computer sciences) employ a variety of different approaches (e.g., attraction forces, path planning, non-linear control) to design herder agents that are able to successfully influence the behaviour of groups of target agents (otherwise not directly controllable), with a more recent, yet growing, emphasis on tasks involving multiple-herders.

To help compare the different herding solutions that have been proposed, we first provided a general formulation of the problem for N_H dynamical herders and $N_T \geq N_H$ targets (Section 2.1). Based on the types of herding tasks (herding behaviours) summarised in Table 2.2, the relevant state of the art solutions were surveyed for the multiple herders case in (Section 2.3 and the single herder case in Section 2.4.

As explicitly stated in [53, 67, 154] and implicitly conveyed in [120, 151, 160] ,this overview highlighted how the success of most of the proposed solutions depends on the existence of a target flocking tendency. This is because a low aggregation strength among targets or the complete absence of flocking behaviour significantly increases the complexity of the herding task; that is, each target needs to be tracked and collected (more-or-less) independently from the others. Thus, to design herder agents able to corral non-flocking targets, identifying effective target selection strategies (i.e., herders decision-making strategies) become an essential part of solving the herding problem.

Additionally, this overview emphasised how, for herds exhibiting flocking behaviour, the most common decision making strategies, for either single and multiple herders, include either (i) corralling the herd centre of mass [80, 84, 151, 160], or (ii) chasing the target furthest from the herd (one target at time) [84, 88, 109, 151, 154], or (iii) for herders to distribute themselves behind the flock, implementing pattern formation techniques [53, 67, 102, 113, 120, 140, 159]. Taking inspiration from the results in [107, 109], in next Chapter we investigate a range of simple, yet effective,

dynamic herding strategies that are able to corral non-flocking target agents. The dynamic herding strategies consists of local feedback control laws for $N_H \geq 2$ herder agents and different dynamic target selection rules driving how the herders make decisions about what targets to corral next.

CHAPTER 3

LOCAL CONTROL RULES AND DYNAMIC TARGET SELECTION STRATEGIES

In this Chapter, we present a simple, yet effective, dynamic herding strategy consisting of local feedback control laws for the herder agents and a set of target selection rules that drive how herders make decentralised decisions on what targets to corral (Sections 3.1-3.3). We numerically analyse how these strategies perform in nominal conditions, and how robust they are to parameter perturbations, uncertainties and unmodeled disturbances in target agent dynamics (Section 3.4). We first (i) assess how different choices of the target selection rules affect the overall effectiveness of the herding control laws and target selection strategies we propose (Section 3.4.1 and (ii) investigate whether the oscillatory containment behaviour observed in some dyadic human herding tasks [109] can indeed emerge spontaneously from the local rules adopted to drive herder agents, rather than having to be explicitly encoded in the dynamics of herders (Section 3.4.3). We then test how herders adopting our local control laws and target selection strategies perform against multiple parameter perturbations (Section 3.4.4) and compare the proposed approach to other key approaches proposed in the literature (Section 3.4.5). Finally, we test the effectiveness of the proposed target selection strategies for solving the herding problem for robotic systems using ROS simulations and real robots in the Robotarium platform [118, 171] (Section 3.5).

The work in this Chapter, in collaboration with Dr. Davide Fiore (University of Naples Federico II, Italy), has been published on Autonomous Robots [9].

3.1 The case study herding problem

We start by considering the case of a small group of *herders* chasing a much larger group of *target agents* whose dynamics, as often happens with natural agents such as fish, birds or bacteria, is stochastic and reflects random Brownian motion. However, contrary to what is usually done in the

rest of the literature [23, 52, 65, 84, 120], we do not consider the presence of any flocking behaviour between target agents, making the problem more complicated to solve as each target needs to be tracked and corralled independently from the others. A herder's action is based on global knowledge of the environment, including the positions of all other herder and target agents. Note that, with respect to other solutions in the literature [22, 65, 120, 149] (surveyed in Chapter 2), our approach will not involve the use of ad-hoc formation control strategies to force the herders to surround the herd, but rather enforces cooperation between herders by dynamically dividing the plane among them by means of simple yet effective (decision-making) rules that can be easily implemented in real robots.

We consider the problem of controlling $N_H \geq 2$ herder agents required to corral a group of $N_T > N_H$ target agents in the plane ($\mathbb{R}^2$) towards a goal region and then contain them within that goal region. Using polar coordinates, we denote (r_j, θ_j) and (ρ_i, ϕ_i) as the polar position of the j-th herder and i-th target agent, respectively, as shown in Figure 3.1. The position of the i-th agent when it is targeted by the j-th herder will be denoted as $\tilde{\mathbf{x}}_{i,j}$ or in polar coordinates as $(\tilde{\rho}_{i,j}, \tilde{\phi}_{i,j})$. Finally, the goal region is a circular *containment region* G, of radius $r^\star$ centred at $\mathbf{x}^\star$. Without loss of generality, we set $\mathbf{x}^\star$ to be the origin of $\mathbb{R}^2$.

Assuming the herders have their own trivial dynamics (2.1) in the plane, we reformulate the *herding problem* as the design of the control action u governing the dynamics of the herders given by

$$m\ddot{\mathbf{y}}_j = u(t, \mathbf{x}_1, \ldots, \mathbf{x}_{N_T}, \mathbf{y}_1, \ldots, \mathbf{y}_{N_H}), \tag{3.1}$$

where m denotes the mass of the herders assumed to be unitary, so that the herders can influence the dynamics of the target agents (whose dynamics will be specified in the next Section) and guarantee that

$$\|\mathbf{x}_i(t) - \mathbf{x}^\star\| \leq r^\star, \qquad \forall i, \forall t \geq t_g, \tag{3.2}$$

where $\|\cdot\|$ denotes the Euclidean norm; that is, all target agents are contained, after some finite *gathering time* t_g, in the desired region G. A herding trial is said to be *successful* in the time interval $[0, T]$ if condition (3.2) holds for some $t_g \in [0, T]$. Note also, that herder and target agents can move freely in $\mathbb{R}^2$, and that an annular *safety region* B of width $\Delta r^\star$ is assumed to surround the goal region G, that the herders maintain between themselves and the goal region where all targets are contained within G.

3.2 Target dynamics

Consistent with [107, 109], we consider the most general scenario of target agents being independent particles moving randomly in the plane. Consequently, targets kinematic model (i.e., first order Ordinary Differential Equations – ODEs) is sufficient to fully describe their motion through the environment, and derive possible herding solutions [84, 88, 120, 148, 151, 160]. Such herding solution, on the contrary, will be chosen in the next Section as second order ODEs to achieve a deeper and more precise control on herder agents movements in the environment.

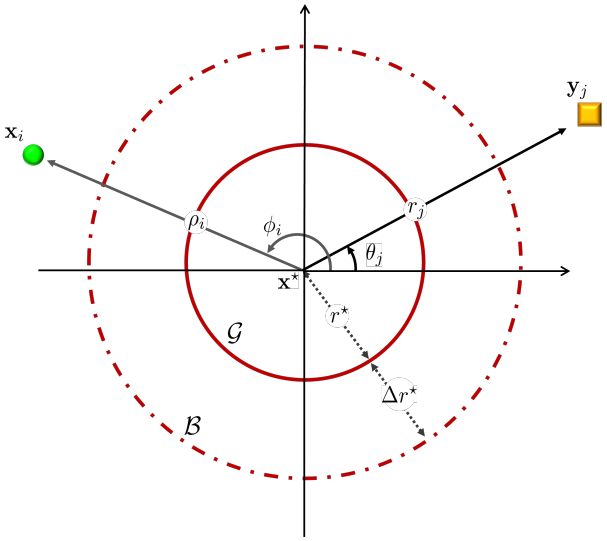


Figure 3.1: Illustration of the spatial arrangement of the herding problem investigated in the current Chapter. The herder agent $\mathbf{y}_j$ (yellow square), with polar coordinates (r_j, θ_j), must corral the target agent $\mathbf{x}_i$ (green ball), with polar coordinates (ρ_i, ϕ_i), into the containment region G (solid red circle) of centre $\mathbf{x}^\star$ and radius $r^\star$. The buffer region B, of width $\Delta r^\star$, is depicted as a dashed red circle. See main text for more details.

In particular, we assume that, when interacting with the herders, target agents are repelled from them and move away in the opposite direction, while in the absence of any external interaction, they randomly diffuse in the plane. Specifically, we assume target agents (2.2) move according to the following stochastic dynamics

$$d\mathbf{x}_i(t) = \mathbf{V}_{r,i}(t)dt + \alpha_b d\mathbf{W}_i(t), \tag{3.3}$$

where $\mathbf{V}_{r,i}$ describes the repulsion exerted by all the herders on the i-th target agent, $\mathbf{W}_i = [W_{i,1}\ , W_{i,2}]^\top$ is a 2-dimensional standard Wiener process and $\alpha_b > 0$ is a constant. We suppose the distance travelled by the target agents depends on how close the herder agents are and model this effect by considering a potential field centred on the j-th herder given by $v_{i,j} = 1/(\|\mathbf{x}_i - \mathbf{y}_j\|)$, exerting on the target agents an action proportional to its gradient [120, 125]. Specifically, the dynamics of the i-th target agent is influenced by all the herders through the reaction term

$$\mathbf{V}_{r,i}(t) = \alpha_r \sum_{j=1}^{N_H} \frac{\partial v_{i,j}}{\partial \mathbf{x}_i} = -\alpha_r \sum_{j=1}^{N_H} \frac{\mathbf{x}_i(t) - \mathbf{y}_j(t)}{\|\mathbf{x}_i(t) - \mathbf{y}_j(t)\|^3}, \tag{3.4}$$

where $\alpha_r > 0$ is a constant. Uncertainties on the repulsive reaction term (3.4) can be seen as being captured by the additional stochastic term in (3.3). Note that the velocity of all target agents is completely determined by (3.3)-(3.4) and we do not assume any upper bound on its maximum value.

3.3 Herder dynamics and control rules

Our solution to the herding problem consists of (i) a local control law to drive the motion of the herder towards a selected target and then push that target inside the goal region, and (ii) a target selection strategy that determines which target a herder decides to corral. When the herd is all corralled, the herders switch back to an idling condition by keeping themselves within the safety region surrounding the goal region.

3.3.1 Local control strategy

For the sake of comparison with the polar coordinate strategy presented in [107, 109], we employed a similar control law. This control law does not result in shortest possible path solutions for herders, but rather ensures circumnavigation of the goal region and the avoidance of target agents already contained within the containment regions (i.e., thereby ensuring the herders do not pass through the containment region and adversely scatter the contained targets). Specifically, the control input of each j-th herder (3.1) is defined as $\mathbf{u}_j = u_{r,j}\,\hat{r}_j + u_{\theta,j}\,\hat{\theta}_j$, where $\hat{\theta}_j = \hat{r}_j^{\perp}$ are unit vectors and $\hat{r}_j = [\cos\theta_j, \sin\theta_j]^\top$, and

$$u_{r,j}(t) = -b_r \dot{r}_j(t) - R(\tilde{\mathbf{x}}_{i,j}, t), \tag{3.5}$$

$$u_{\theta,j}(t) = -b_\theta \dot{\theta}_j(t) - T(\tilde{\mathbf{x}}_{i,j}, t). \tag{3.6}$$

Here the constants b_r, $b_\theta > 0$, and the feedback terms $R(\tilde{\mathbf{x}}_{i,j}, t)$ and $T(\tilde{\mathbf{x}}_{i,j}, t)$ are elastic forces that drive the herder towards target i and then push target i towards the containment region G. These forces are defined as

$$R(\tilde{\mathbf{x}}_{i,j}, t) = \epsilon_r \Big[r_j(t) - \xi_j(t)\,(\tilde{\rho}_{i,j}(t) + \Delta r^\star) - (1 - \xi_j(t))\,(r^\star + \Delta r^\star) \Big], \tag{3.7}$$

$$T(\tilde{\mathbf{x}}_{i,j}, t) = \epsilon_\theta \Big[\theta_j(t) - \xi_j(t)\tilde{\phi}_{i,j}(t) - (1 - \xi_j(t))\psi(t) \Big]. \tag{3.8}$$

with ϵ_r, $\epsilon_\theta > 0$, and where ξ_j regulates the switching policy between collecting and idling behaviours. That is, $\xi_j = 1$, if $\tilde{\rho}_{i,j} \geq r^\star$, and $\xi_j = 0$, if $\tilde{\rho}_{i,j} < r^\star$, so that the herder is attracted to the position of the i-th target $\tilde{\mathbf{x}}_{i,j}$ (plus the radial offset $\Delta r^\star$) when the current target is outside the containment region ($\xi_j = 1$) or close to the boundary of the buffer region at the idling position $(r^\star + \Delta r^\star, \psi)$, in polar coordinates, ($\xi_j = 0$) otherwise . The value of the *idling angle* ψ depends on the specific choice of the target selection strategy employed, which are discussed next. As for the target agents, we do not assume any upper bound on the maximum velocity of the herders.

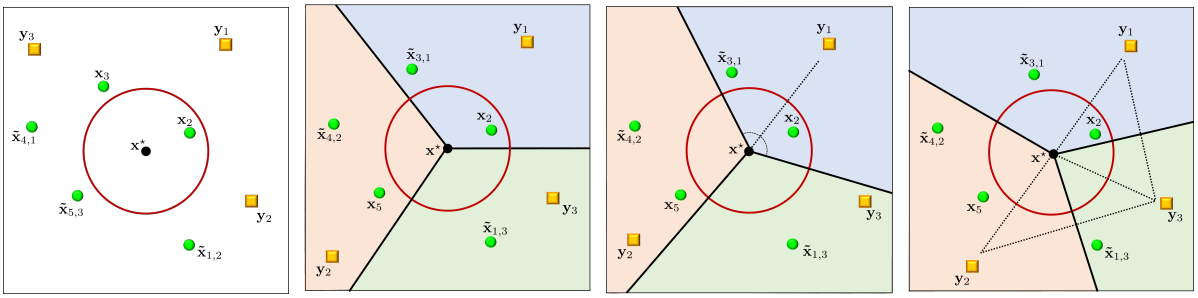
It is important to note that the control laws (3.5)-(3.6) are much simpler than those presented by [107] as they do not contain any higher order nonlinear terms nor parameter adaptation rules aimed at capturing differences in the behavioural intentions of human actors performing the dyadic herding game investigated by [107] (see [107] for further details). In Appendix A, we detailed our simplification of the original model proposed in [107] and expanded in [109], to identify the essential

characteristics required for artificial agents to achieve successful performance, while also allowing for the spontaneous emergence of the different behavioural modes observed in human players (i.e., oscillatory vs. non-oscillatory behaviour). More specifically, using numerical simulations we demonstrate how both non-oscillatory and oscillatory human herding behaviours can emerge from simpler bottom-up dynamical rules (i.e., (3.5)-(3.6)); see also Section 3.4.3).

3.3.2 Target selection strategies

In the case of a single herder corralling multiple agents, the most common strategy in the literature is for the herder to select the target that is either the farthest target agent from the goal region ([84, 88, 109, 151, 154]), or the centre of mass of the herd [80, 84, 151, 160]. When two or more herders are involved, the problem is usually solved using a formation control approach, letting the herders surround the herd and then drive them towards the goal region [53, 67, 102, 113, 120, 140, 159]. Rather than using formation control techniques or solving off-line or on-line optimisation problems as in dynamic target assignment problems (e.g., [18]), here we present a set of simple, yet effective, target selection strategies that exploit the spatial distribution of the herders. That is, the potential solutions investigated here allow herders to cooperatively select targets without requiring any computationally expensive optimisation problem to be solved on- or off-line.

More specifically, we investigated four different herding strategies, which are graphically illustrated in Figure 3.2 for $N_H = 3$ herders and can be defined as follows.



(a) Global search (b) Static partitioning (c) Leader-follower (d) Peer-to-peer

Figure 3.2: Graphical representation of the target selection strategies. Herders are depicted as yellow squares, target agents as green balls. The colours in which the game field is divided correspond to regions assigned to different herders. Herder $\mathbf{y}_j$ is currently chasing target agent $\tilde{\mathbf{x}}_{i,j}$, while target agent $\mathbf{x}_i$ is not chased by any herder.

Global search strategy (no partitioning) Each herder selects the farthest target agent from the containment region which is not currently targeted by any other herder (Figure 3.2(a)). Note that this is the simplest possible strategy and is employed here for the sake of comparison only and was not implemented in real robot experiments detailed in Section 3.5.2.

Static arena partitioning At the beginning of the trial the herding plane is partitioned in N_H circular sectors of width equal to $2\pi/N_H$ rad centred at $\mathbf{x}^\star$. Each herder is then assigned one sector to patrol and selects the target agent within their sector that is farthest from G (Figure 3.2(b)). Note that this is the same herding strategy used in [107] for $N_H = 2$ herders.

Dynamic leader-follower (LF) target selection strategy At the beginning of the trial, herders are labelled from 1 to N_H in anticlockwise order starting from a randomly selected herder which is assigned the *leader* role. The plane is then partitioned dynamically into different regions as follows. The leader starts by selecting the farthest target agent from G whose angular position $\tilde{\phi}_{i,1}$ is such that

$$\tilde{\phi}_{i,1} \in \left(\theta_1 - \frac{1}{2}\frac{2\pi}{N_H}, \theta_1 + \frac{1}{2}\frac{2\pi}{N_H}\right],$$

where θ_1 is the angular position of the leader at time t. Then, all the other *follower* herders ($j = 2,\ldots,N_H$), in ascending order, select their targets as the agent farthest from G such that

$$\tilde{\phi}_{i,j} \in \left(\theta_1 - \frac{1}{2}\frac{2\pi}{N_H} + \zeta_j, \theta_1 + \frac{1}{2}\frac{2\pi}{N_H} + \zeta_j\right],$$

with $\zeta_j = 2\pi(j-1)/N_H$. As the leader chases the selected target and moves in the plane, the partition described above changes dynamically so that a different circular sector with *constant* angular width $2\pi/N_H$ rad is assigned to each follower at any time instant. In Figure 3.2(c) the case is depicted for $N_H = 3$ in which the sector $\left(\theta_1 - \frac{\pi}{3}, \theta_1 + \frac{\pi}{3}\right]$ is assigned to the leader herder while the rest of the plane is assigned equally to the other two herders.

Dynamic peer-to-peer (P2P) target selection strategy At the beginning of the trial herders are labelled from 1 to N_H as in the previous strategy. Denoting as ζ_j^+ the angular difference between the positions of herder j and herder $(j+1) \bmod N_H$ at time t, and as ζ_j^- that between herder j and herder $(j+N_H-1) \bmod N_H$ at time t, then herder j selects the farthest target agent from G whose angular position is such that

$$\tilde{\phi}_{i,j} \in \left(\theta_j - \frac{\zeta_j^-}{2}, \theta_j + \frac{\zeta_j^+}{2}\right].$$

Unlike the previous case, now the width of the circular sector assigned to each herder is also dynamically changing as it depends on the relative angular positions of the herders in the plane.

The *idling angle* ψ in (3.8) is set equal to the angular position $\tilde{\phi}_{i,j}$ of the last contained target for the global search strategy, otherwise it is set equal to the angular position corresponding to the half of the angular sector assigned at each time to the herder. The idling angle ψ represents a "rest" angular position for herders, as such, its choice is motivated by the design goal of reducing unnecessary herder motion when all target agents are gathered in the containment region (i.e., when (3.2) is satisfied).

A crucial difference between the different herding strategies presented above is the nature (local vs global) and amount of information that a herder's selection decision in based on. Specifically,

when the *global search strategy* is used, every herder needs to know the position $\mathbf{x}_i$ of every target agent in the plane, not currently targeted by other herders (e.g., sharing a common sensing infrastructure scanning the whole herding space for targets to be corralled). However, in the case of the *static arena partitioning*, a herder only needs to know its assigned (constant) circular sector and with the position $\mathbf{x}_i$ of the sub-set of targets within that sector.

For the dynamic target selection strategies, even less information is generally required. Indeed, in the *dynamic leader-follower strategy* the herders, knowing N_H, can either self-select the sector assigned to themselves (if they act as leader) or self-determine their respective sector by knowing the position of the leader $\mathbf{y}_1$ (e.g., each herder is equipped with proximity sensors to detect targets in their sector; the leader herder has to communicate with each follower herder, while the latter only need to communicate with the leader). Similarly in the *dynamic peer-to-peer strategy* herders can self-select their sectors by using the angles ζ_j^+ and ζ_j^- (e.g., a communication protocols is needed only between each herders and its two closest neighbours).

Note that in the event of perfect radial alignment of the herder and its target, the herder might push the target away, rather than towards the goal region. Although this condition is very unlikely to persist due to the random motion of the passive agents, this problem can be avoided by extending the herder dynamics in (3.1) by a circumnavigation force $\mathbf{u}_j^\perp(t)$. This force is orthogonal to the vector $\Delta\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{y}_j$, and its amplitude depends on the angle χ_{ij} between $\Delta\mathbf{x}_{ij}$ and $\mathbf{y}_j$, such that it is maximum when the two vectors are parallel ($\chi_{ij} = 0$) and zero when they are anti-parallel ($\chi_{ij} = \pi$). Specifically, it is defined as:

$$\mathbf{u}_j^\perp(t) = \bar{U} \cdot v(t) \cdot \cos^2\left(\frac{\chi_{ij}}{2}\right) \frac{\Delta\mathbf{x}_{ij}^\perp}{\|\Delta\mathbf{x}_{ij}\|}, \tag{3.9}$$

where $\bar{U} > 0$ is the maximum amplitude, and $v \in \{-1, 1\}$, whose value depends on which halves of the assigned sector the herder is currently in thereby guaranteeing that the target agent is always pushed toward the interior of the sector.

3.4 Numerical validation

The herding performance of the proposed control strategies was first evaluated through a set of numerical experiments aimed at (i) assessing their effectiveness in achieving the herding goal; (ii) comparing the use of different target selection strategies; (iii) studying the robustness of each strategy to parameter variations. The implementation and validation of the strategies in a more realistic robotic environment is reported in the next Section where ROS simulations are included.

3.4.1 Performance metrics

We defined the following metrics (see Appendix B.1 for the formal definitions) to evaluate herding performance. Specifically, we computed the (i) gathering time t_g (ii) the average length d_g of the path travelled by the herders until all targets are contained, (iii) the average total length d_{tot} of the

path travelled by herders during a herding trial, (iv) the mean distance D_T between the herd's centre of mass and the centre of the containment region, and (v) the herd agents' spread $S_{\%}$.

Note that *lower* values of t_g correspond to *better* herding performance; herders taking a shorter time to gather all the target agents in the goal region. Also, *lower* values of D_T and $S_{\%}$ correspond to a *tighter* containment of the target agents in the goal region, while *lower* values of d_g and d_{tot} correspond to *more efficient* herd gathering and containment, respectively.

3.4.2 Performance analysis

We carried out 50 numerical trials with $N_T = 7$ target agents and either $N_H = 2$ or $N_H = 3$ herders, starting from random initial conditions. All trials were found to be successful, that is, condition (3.2) was satisfied. (A description of the simulation setup and the simulation parameters employed are reported in Appendix B.2.)

	Global	Static	LF	P2P
		$N_H = 2$		
t_g [a.u.]	8.5 ± 1.5	15.2 ± 9.6	15.3 ± 9.2	13.3 ± 6.2
d_g [a.u.]	139 ± 34	102 ± 42	92 ± 49	143 ± 53
d_{tot} [a.u.]	841 ± 27	493 ± 51	423 ± 29	418 ± 56
D_T [a.u.]	1.3 ± 0.1	1.4 ± 0.6	1.5 ± 0.6	1.3 ± 0.5
$S_{\%}$ [%]	0.1 ± 0.05	0.2 ± 0.1	0.2 ± 0.2	0.2 ± 0.1
		$N_H = 3$		
t_g [a.u.]	5.8 ± 0.9	11.2 ± 7.9	11.2 ± 8.7	8.3 ± 5.6
d_g [a.u.]	88 ± 21	107 ± 98	84 ± 74	58 ± 43
d_{tot} [a.u.]	1242 ± 26	757 ± 67	885 ± 56	950 ± 79
D_T [a.u.]	0.6 ± 0.1	0.9 ± 0.5	0.8 ± 0.3	0.7 ± 0.1
$S_{\%}$ [%]	0.1 ± 0.03	0.2 ± 0.2	0.2 ± 0.2	0.2 ± 0.3

Table 3.1: Average performance and standard deviation over 50 successful trials of different herding strategies for $N_T = 7$ target agents.

The results of our numerical investigation are reported in Table 3.1. Kruskal-Wallis Test [79] were conducted to examine the differences between the different target selection strategies. As expected, when herders search globally for agents to corral, their average total path, d_{tot}, is significantly larger than when dynamic target selection strategies are used ($\chi^2 = 142.5$, $p < 0.001$), indicting that this strategy was the least efficient (in addition to requiring complete information about the agents). With regard to the aggregation of the herd, as measured by D_T and $S_{\%}$, for $N_H = 2$ all of the other (non-global) strategies investigated produced comparable results in terms of both mean and standard deviation, with the exception of $S_{\%}$ for the peer-to-peer strategy ($\chi^2 = 12$, $p < 0.003$). However, the two dynamic strategies exhibited better gathering performance (t_g and d_g) than the static arena

partitioning strategy and when, $N_H = 3$, the peer-to-peer strategies achieving significantly lower t_g ($\chi^2 = 13$, $p < 0.01$) and d_g ($\chi^2 = 20$, $p < 0.0001$).

In summary, the results of the numerical simulations revealed that a higher level of cooperation between herders and a more efficient coverage of the plane, as in the dynamic target selection strategies, yielded overall better herding performance. The implication investigated in Section 3.5.1 and 3.5.2, is that these strategies might also be more effective for robots applications given that such systems are bound to move at limited speed.

3.4.3 Emerging complex behaviours

Recall that the herding task explored here has been inspired by the one recently employed to explore and model the coordination dynamics that emerges between pairs of human herders [106, 109, 129], as discussed in Chapter 2. To complete the task successfully, human herders learned to adopt two modes of behavioural coordination. The first, known as search-and-recovery or S&R behaviour, involving human herders corralling the targets that were approximately farthest from the containment area one at a time. Once all targets were contained, some human teams also learned to switch to an oscillatory containment mode of behaviour (COC), in which the players perform semi-circular movements around the herd.

In Appendix A we analysed how, starting from bottom-up designed herder models, it was possible to observe the emergence of oscillating herding behaviours from herders that simply divide the game field in two fixed halves (what we called the *static arena partitioning* strategy). When we tested herding scenarios in which herders adopted the dynamic selection strategies presented in Section 3.3, oscillatory motion of the herders emerged spontaneously without the need of including extra nonlinear terms in the model as done in [109] in all scenarios (see Figure 3.3 for an example of spectral classification). Specifically, we found oscillatory motion to emerge in all the trials when the *global* or *peer-to-peer* strategies were adopted with both herders moving around the containment region as also observed when pairs of humans were asked to herd agents in the virtual reality setting [109]. The *static* and *leader-follower* strategies were also found to give rise to oscillatory motion of the herders in a fraction of the trials with only one of the two herders oscillating (a type of behaviour we classify as non-cooperative).

3.4.4 Robustness analysis

Next, we analysed the robustness of the proposed herding strategies to variations in herd size and the magnitude of the repulsive reaction to the herders exhibited by the target agents (Figure 3.4). Specifically, we varied N_T between 3 and 60 and the repulsion parameter α_r in (3.4) between 0.05 and 2.5, while keeping $N_H = 2$; we found that all strategies succeed in herding up to 60 agents across a large region of parameter values (see the blue areas in Figure 3.4()). As expected, the global strategy, where herders patrol the entire plane, was again found to be the least efficient in terms of total

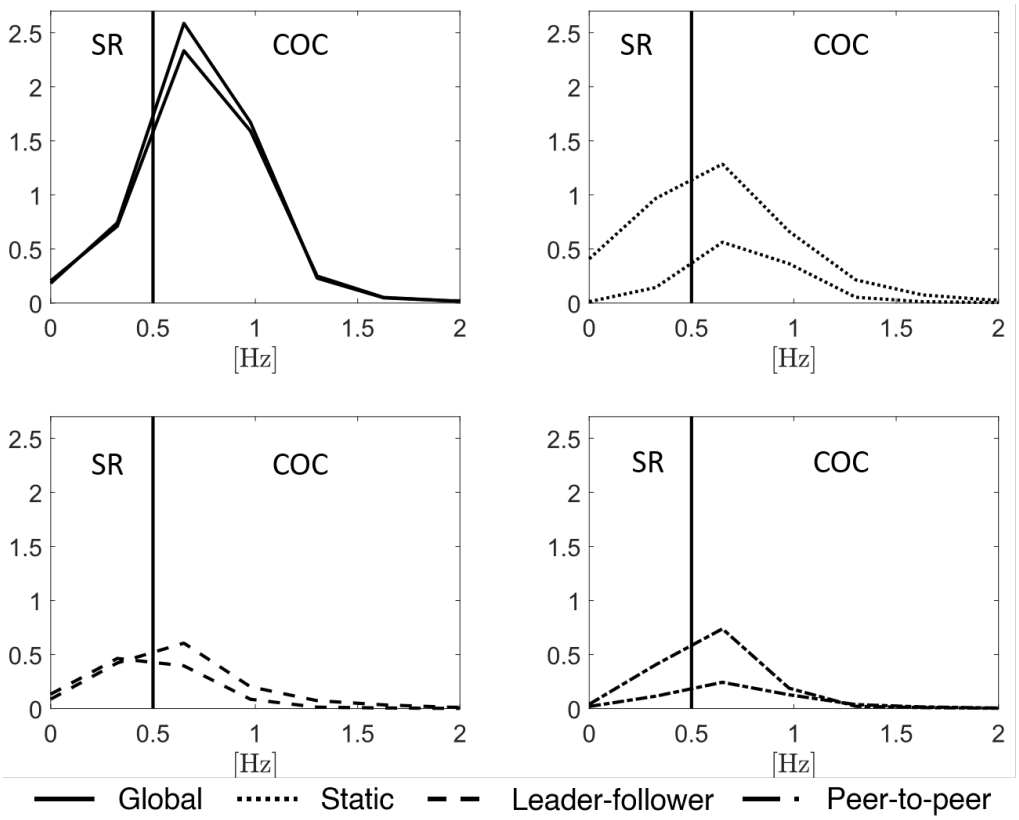
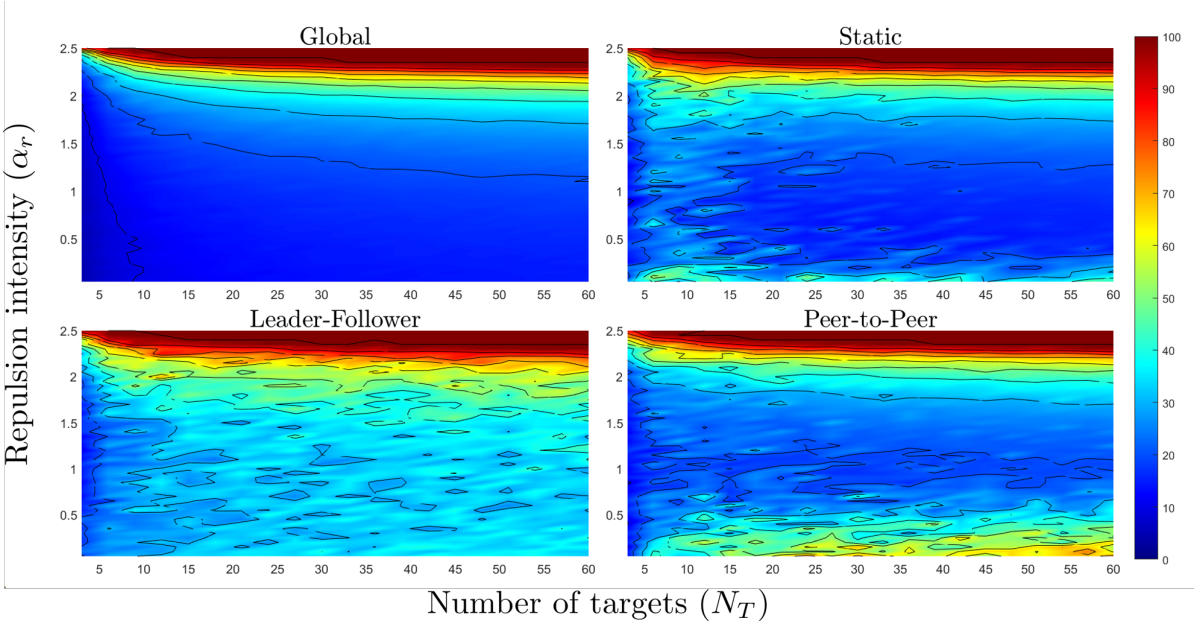
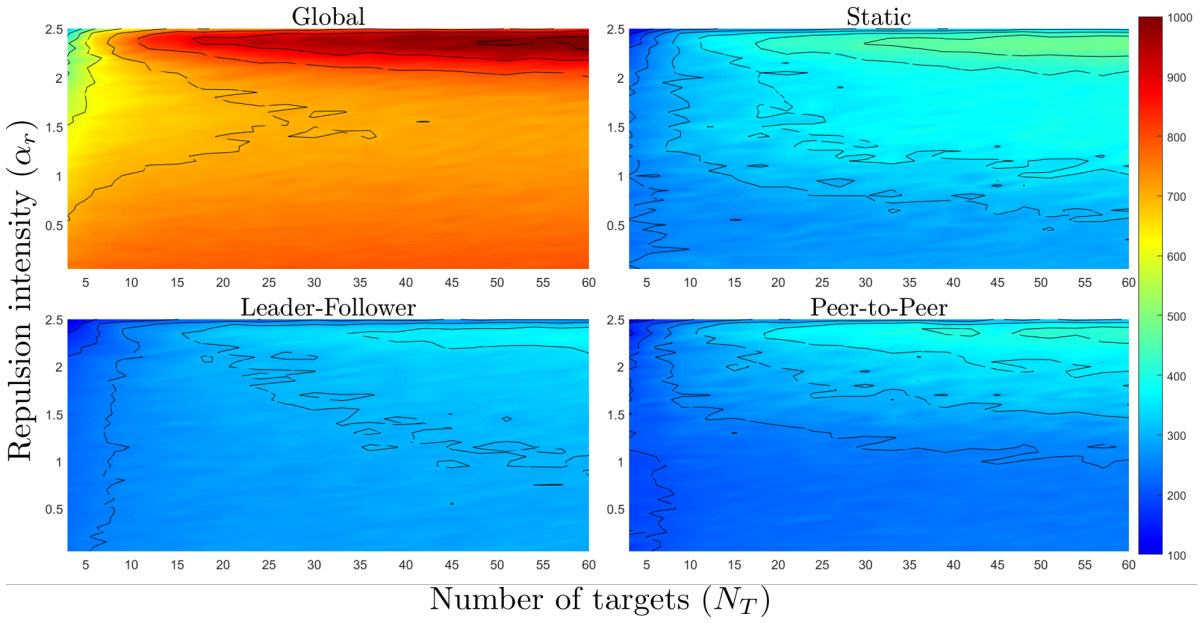


Figure 3.3: Comparison of power spectra exhibited by $N_H = 2$ herders adopting different target selection strategies in a successful trial when corralling $N_T = 7$ target agents. Peak values of power spectrum are used to classify the behaviour in search and recovery (SR) or coupled oscillatory containment (COC) (see Appendix B.1 for their definition). Herders not dividing (solid line), statically dividing (dotted line) and cooperatively dividing (dash-dotted line) the game field have a peak frequency on the right side of the threshold $\omega_c = 0.5\,\text{Hz}$ and their coupled behaviours are classified as COC. On the other hand, herders adopting the Leader-follower strategy (dashed line) have peak frequencies on both sides as the leader herder mostly engage in a non-oscillatory behaviour and the follower herder in a oscillatory one.



(a) Gathering time t_g. Lower values correspond to faster herding.



(b) Total distance travelled d_{tot}. Lower values correspond to more efficient herding.

Figure 3.4: Robustness analysis of the proposed herding strategies for two herders ($N_H = 2$) to variation of herd size N_T and repulsive reaction coefficient α_r. N_T was varied between 3 and 60 agents, with increments equal to 3, while α_r between 0.05 and 2.5, with increments equal to 0.05. For each pair (N_T, α_r) the corresponding metric was averaged over 15 simulation trials starting with random initial positions. The coloured plots were obtained by interpolation of the computed values.

distance travelled by the herders (Figure 3.4(a)), with the dynamic peer-to-peer strategy offering the best compromise in containment performance (Figure 3.4()) and efficiency (Figure 3.4(a)).

To further validate these findings we carried out 50 simulation trials for the herding scenario in which $N_H = 3$ herders are required to herd $N_T = 60$ target agents. In this more challenging scenario not all the trials were completed successfully, due to at least one target agent escaping the containment region. Thus, we averaged the resulting performance over the successful trials only (Table 3.2). Herders adopting the global and peer-to-peer strategies successfully herded all agents in over 50% of the trials. The global search strategy also resulted lower average gathering time ($t_g = 12.96$) and lower herd spread ($S_{\%} = 0.48$). However, the path travelled to achieve the goal (d_{tot}) was significantly larger compared to the local (static and dynamic) selection strategies ($\chi^2 = 54$, $p < 0.0001$).

	Global	Static	LF	P2P
Successful trials	49	11	7	26
t_{g} [a.u.]	12.9 ± 1.8	18.2 ± 4.2	16.1 ± 5.1	15.9 ± 6.6
d_{g} [a.u.]	211 ± 23	191 ± 43	138 ± 31	148 ± 66
d_{tot} [a.u.]	1226 ± 25	708 ± 28	742 ± 74	734 ± 97
D_T [a.u.]	6.5 ± 0.7	10.8 ± 4.1	8.2 ± 3.3	8.9 ± 6.8
$S_{\%}$ [%]	0.5 ± 0.05	0.9 ± 1.5	3.4 ± 5.8	1.1 ± 1.7

Table 3.2: Average performance over successful trials of different herding strategies for $N_T = 60$ target agents and $N_H = 3$ herders.

3.4.5 Comparison with other approaches

We compared our proposed herding strategy with two other solutions presented in the literature, by considering as a benchmark scenario the case of $N_H = 3$ herders chasing $N_T = 7$ targets.

The herder model (2.15) presented in [88] was adapted to this scenario, while the model (2.6) proposed in [120] was applied as is. Specifically, the model in [88] was originally designed to drive a *single* herder to corral multiple targets by switching between targets according to a dwell-time condition (i.e., at fixed time intervals the herder agent switch between target to corral). To apply this model to our benchmark scenario, we assume that each of the herders corrals only one target at a time without coordinating with the other herders.

In all simulations the dynamics of the targets was set as in (3.3). Note that in [88] target dynamics are assumed to be deterministic but uncertain, while in [120] targets have noisy dynamics but exhibit an aggregating behaviour, or flocking, that is not considered here. Performance averaged over 25 trials is shown in Table 3.3, with both dynamic selection strategies (LF and P2P) proposed here out performing the strategies proposed in [88, 120]. The relative poorer performance of these strategies when compared to the strategies proposed here can be explained by noticing that the model in [88] lacks coordination among multiple herders, while in the control solution in [120] heavily relies on flocking among target agents which is not required in our case of interest. Furthermore, the

chosen models were designed under the assumption of deterministic ([88, 120]) and flocking ([120] targets. As a consequence, both models' poor performance is affected by the stochastic, rather than deterministic, scattering dynamics of the target agents considered here, further confirming the effectiveness of our approach when deployed in this type of scenario.

	Model [88]	Model [120]	LF	P2P
t_g [a.u.]	100 ± 0	100 ± 0	11.8 ± 9	8.6 ± 6.3
d_g [a.u.]	297 ± 90	7.9 ±0.6	85.4 ± 67	61 ± 55
d_{tot} [a.u.]	297 ± 90	7.9 ±0.6	889 ± 52.6	942 ± 74
D_T [a.u.]	3 ± 1.4	8 ± 6	0.8 ± 0.3	0.7 ± 0.1
$S_{\%}$ [%]	3.7 ± 1.1	35.7 ± 6.1	1.7 ± 0.5	0.2 ± 0.3

Table 3.3: Average performance over 25 trials of $N_H = 3$ herders driven by herding models available in literature tasked to corral $N_T = 7$ target agents.

3.5 Experimental (robotic) validation

The numerical validation, presented in Section 3.4, proved the efficacy of our strategies under the ideal assumption of unbounded herding environment, unbounded agents velocities and unbounded communication possibilities among herders. In real world applications, from one perspective, target motion can be mostly unknown (e.g., in case of biological systems) or designed to be faster and/or smarter than herders (e.g., in case of artificial systems). From another perspective, artificial herder's design needs to address physical constraints like hardware limitations (e.g., upper-bounded velocities and accelerations), or not ideal herding conditions (e.g., uneven planar environment slowing their down, impossibility to preserve enough energy to complete the task, faulty communication among herders). To validate the strategies proposed in more realistic robotic settings, we complemented the numerical simulation presented in Section 3.4 with simulations in ROS[1] and experiments on real robots conducted via the Robotarium platform [118, 171]. In the ROS simulations, we addressed the possibility of success of our strategies, and their underlying communication protocols, in a realistic robotic environment. Then, the Robotarium experiments allowed us to test if the constrained environment and the bounded robot dynamics affected herding task completion.

3.5.1 ROS simulations

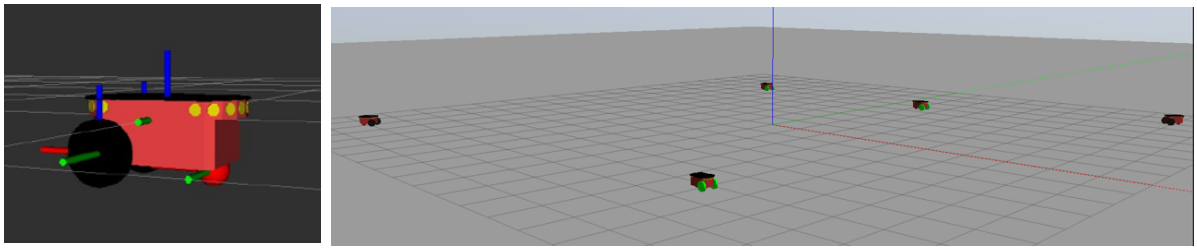
ROS[2] is an advanced software framework for robot software development that provides tools to support the user during all development cycles, from low-level control and communication to deploy-

[1] Code available at `https://github.com/diBernardoGroup/HerdingProblem`

[2] `https://www.ros.org`

ment on real robotic systems. We used the Gazebo software package[3] to test the proposed control architecture on accurate 3D models of commercial robots to simulate their dynamics and physical interaction with the virtual environment.

We considered a scenario where $N_T = 3$ target agents needed to be herded by $N_H = 2$ robotic herders. All agents were chosen to be implemented as Pioneer 3-DX[4], a commercially available two-wheel, two-motor differential drive robot whose detailed model is available in Gazebo (Figure 3.5). The desired trajectories for the robots are generated by using equations (3.3) and (3.5)-(3.6) for the



(a) Robot agent

(b) Simulated environment

Figure 3.5: Overview of Gazebo-ROS application, with 3D model of the Pioneer 3-DX robot (a) and a landscape view of the simulated environment (b).

target and herder robots, respectively, which were used as reference signals for the on-board inner control loop to generate the required tangential and angular velocities (see Appendix B.3 for further details).

Examples of ROS simulations are reported in Figure 3.6, with all of the target selection strategies tested (static arena partitioning, leader-follower, peer-to-peer) resulting in successful robot herding performance. Figure 3.6 also shows that the angular position of the herders remained within the bounds defining the sector of the plane assigned to them for patrolling. The only exception can be seen from an inspection of Figure 3.6(d) illustrating the leader-follower strategy. Here, the follower herder temporarily exceeded the bounds when the lead herder changed its angular position while corralling its target. This is essentially due to the subordinate role of the follower herder with respect to the leader.

3.5.2 Robotarium experiments

Robotarium is a remotely accessible swarm robotics research platform equipped with GRITSBot robots which allows rapid deployment and testing of custom control algorithms [119, 171]. To comply with the limited space of the arena (3.2 m × 2 m) and safety protocols to avoid collisions between robots (robots' diameter is 11 cm) implemented in the platform, we considered a scenario

[3] `http://wiki.ros.org/gazebo_ros_pkgs`

[4] Pioneer 3 - Operations Manual, available at `https://www.inf.ufrgs.br/~prestes/Courses/Robotics/manual_pioneer.pdf` (2020/08/11)

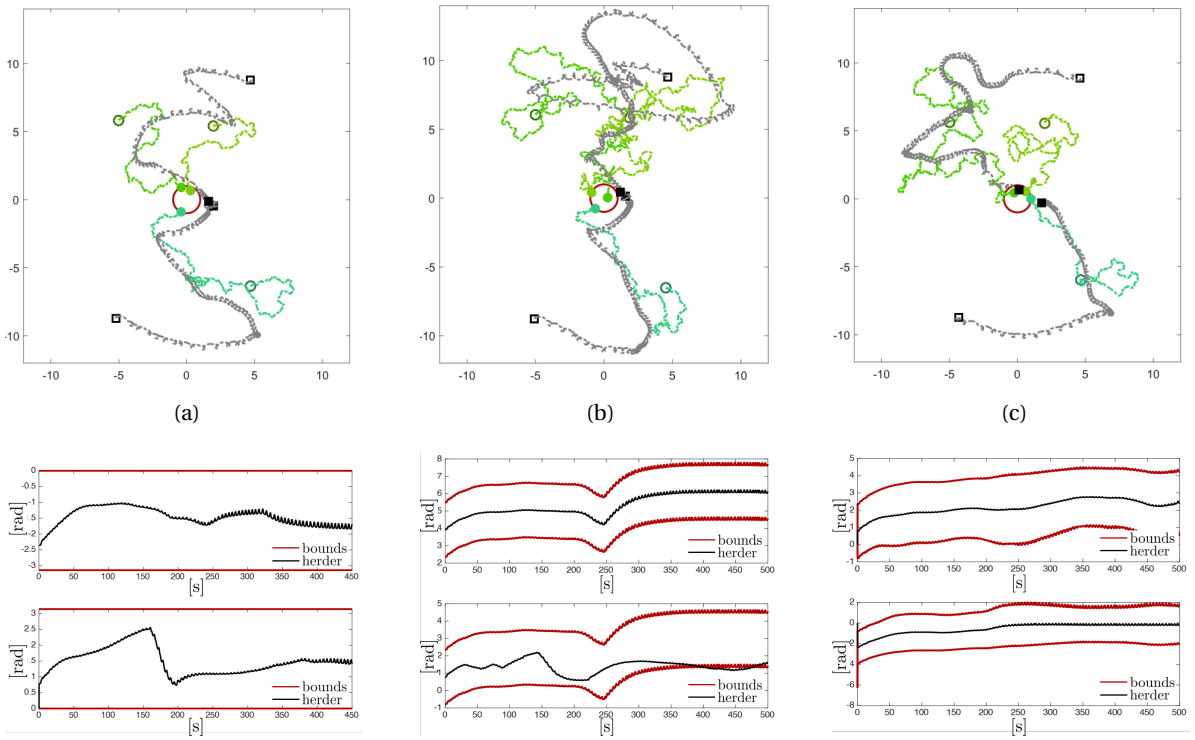


Figure 3.6: ROS simulations. Top panels show the trajectories of target agents (green lines) and herders (grey lines) adopting (a) static arena partitioning, (b) leader-follower and (c) peer-to-peer herding strategies simulated in the Gazebo environment. The containment region G is depicted as a red circle. Black square marks denote the initial and the final (solid coloured) position of the herders. Green circle marks show the initial and the final (solid coloured) position of the target agents. Bottom panels show that all herders are able to collect the herd in less than 500 s by following the angular bounds (red lines) prescribed by the (d) static arena partitioning, (e) leader-follower and (f) peer-to-peer herding strategies.

with $N_T = 4$ target robots and $N_H = 2$ herder robots; a herding scenario that was also considered in [8, 129] to study and model the selection strategies adopted by pairs of human-driven herder agents.

Herder parameters were selected as described in Appendix B.2, while the coefficient of diffusion and repulsion in the dynamics of passive agents (3.3) were scaled to $(\alpha_b, \alpha_r) = (0.001, 0.4)$ to comply with the physical constraints on the hardware of the GRITBots; having a max tangential speed of 20 cm/s and a max rotational speed of about 3.6 rad/s.

The results of the experimental test[5] are reported in Figure 3.7. Both dynamic strategies were found to be effective in containing the target robots with a gathering time t_g less than 70 s, and with the peer-to-peer strategy guaranteeing slightly faster convergence (Figure 3.7(b)) compared to the leader-follower strategy (Figure 3.7(a)) over all the 5 trials performed.

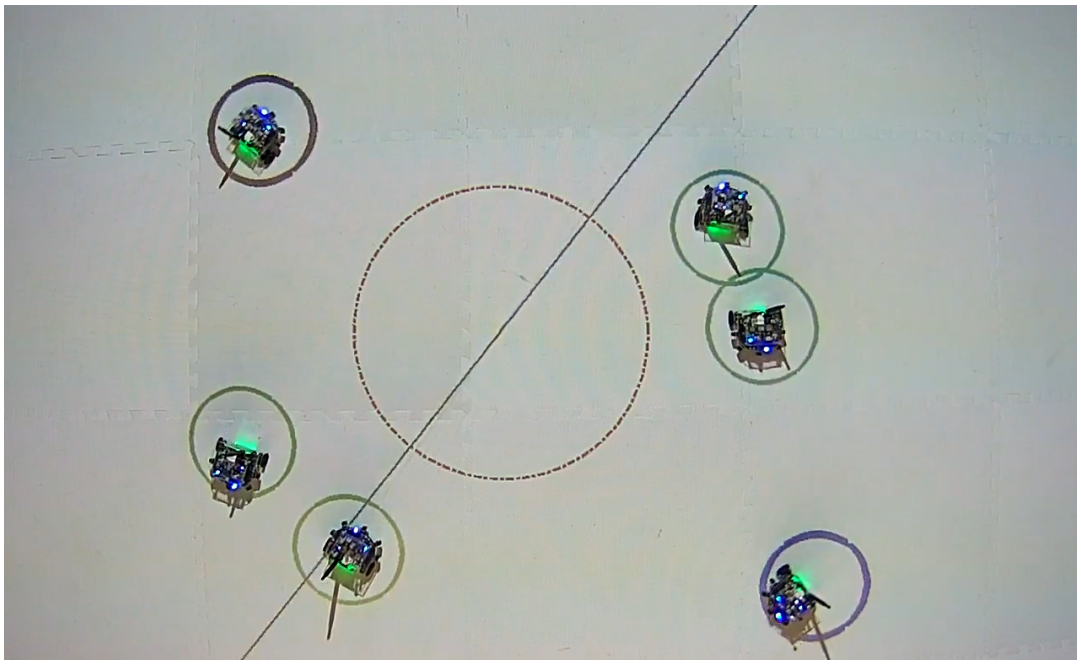
[5] Movies of two illustrative trials are available at `https://github.com/diBernardoGroup/HerdingProblem`

Note that the proposed dynamic control strategies were also found to have good performance when implemented in real robots with constraints imposed on their maximum velocities, assumed unbounded during the design phase described in Section 3.3.

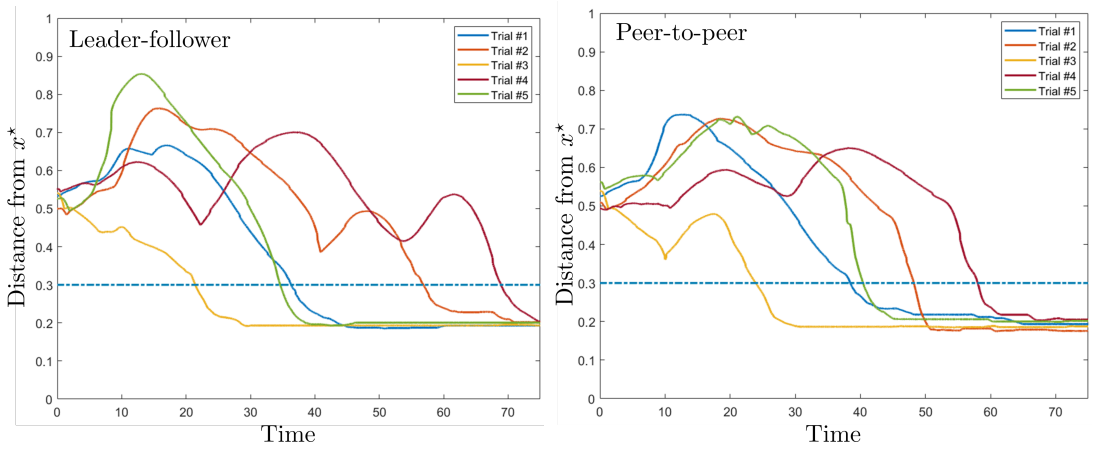
3.6 Summary

In this Chapter we investigated a simple control strategy for solving non-flocking herding problems where the number of target agents is greater than the number of herder agents. Our approach was based on the combination of a set of (i) local dynamical rules driving the herders movements to the position of selected targets and (ii) a target selection strategy that involved partitioning the game area (plane) among the herders, with a herder selecting a target to corral in the sector assigned to them either statically or dynamically (Section 3.3). More specifically, we proposed two dynamic target selection strategies (leader-follower and peer-to-peer) and compared these two strategies against two benchmark and previously used target selection strategies, namely global selection and static game area partitioning. Recall, that for the leader-follower strategy the herding space was partitioned dynamically between the different herders depending on the position of the "lead" herder. For the peer-to-peer strategy, the herding space was dynamically split into equally sized partitions as a function of each herder's local position. For the static partitioning strategy, the game plane was divided at the beginning of a herding trial and remained fixed (static) over time. For the global strategy, there was no area partitioning, with herders simply targeting the herder furthest from the containment region not currently being herded by another herder. Our numerical results show the effectiveness of the control strategy (movement control laws plus target selection strategy) and the ability of the simple control process proposed to cope with an increasing number of target agents and variations of the repulsive force they feel when the herders approach them (Section 3.4). Interestingly, our numerical evidence also suggests that the oscillatory motions of herders observed in experiments with human players [109] may emerge from the local rules of interaction between herders and target agents and do not need to be explicitly encoded in the mathematical model describing their dynamics. Finally, we tested our control strategy via simulations in ROS and experiments on real robots, showing that the purposed local control law and the various dynamic target selection strategies are effective (and viable) in real robotic scenarios (Section 3.5). To the best of our knowledge, our approach is the only one available in the literature to drive multiple herders to collect and contain a group of multiple agents that do not possess a tendency to flock and whose dynamics is stochastic.

In what follows, we will investigate whether it is possible to model the target selection strategies adopted by humans (i.e., distributed human decision making) in joint and team multi-agent tasks (Chapters 4 and 6, respectively) and then employ these "human" target selection models to control AA herders.



(a)



(b) (c)

Figure 3.7: Robotarium experiments. (a) Overview of the Robotarium arena, with GRITSBot robots for a herding scenario with $N_H = 2$ herder robots (circled in black and blue) and $N_T = 4$ target robots (circled in green), and the evolution in time of the distance of the farthest passive robot from the containment region G when the (b) leader-follower and the (c) peer-to-peer strategies are employed for each trial. The mobile GRITBots representing passive agents were initialised with random initial position chosen as $\mathbf{x}_i(0) = 2\,r^\star e^{i\phi_i(0)}$, with $\phi_i(0)$ drawn with uniform distribution in the interval $(-\pi, \pi]$. The radius $r^\star$ of the containment region has been chosen equal to 0.3 m (i.e., equal to a third of the length of the arena).

CHAPTER 4

HUMAN-INSPIRED STRATEGIES TO SOLVE COMPLEX JOINT TASKS

In this Chapter, we used Supervised Machine Learning (Section 4.1) to build a (function approximation) model of how human players, with different expertise levels, select targets to corral during a simulated multi-agent herding task. More specifically, based on the assumptions that (i) the target selection strategies of human herders were reflected in the herders' observed actions and (ii) that human herders use dynamical information on the state of targets and other herder agents (e.g., positions, velocities) around them to make target selection decisions, we trained an LSTM-layered artificial neural networks to predict these action decisions using short sequences of the herding system's state preceding the observed target selection (Section 4.2). Here, we first detail how the resultant LSTM_{NN} models could accurately predict the target selection behaviours of expert and novice herders (Section 4.3). We then demonstrate the utility of the proposed modelling approach for creating artificial herder agents (AA-herders) capable of 'human-like' decision making behaviour by employing the LSTM_{NN} models of novice and expert strategies to control the target selection decisions of the AA-herders (Section 4.4-4.5). Finally, the performance of the AA-herders employing human-inspired target selection strategies was compared to that of herders who selected targets via heuristic target selection strategies and dynamic selection strategies presented in Chapter 3.3 (Section 4.6).

Part of the work in this Chapter has been presented at the 6th Conference on Analysis and Control of Chaotic System (CHAOS 2021) [8]. Additionally, with results in Chapter 5, it has been prepared for publication.

4.1 Overview of supervised machine learning and artificial neural networks

The development and application of Machine Learning (ML) has rapidly increased over the last decade. For example, ML applications are integral to all modern forms of image and speech recognition [5, 37, 57, 145] financial modelling [89], and online product, movie and social interest recommendations [16, 115, 155]. In fact, today, ML is employed in everything from scientific analysis [50, 77], to disease diagnosis, recover and therapy planning [12, 143], to employee identification and selection [1, 46], to credit risk assessment [6, 175], to the control of autonomous cars and robotic systems[33, 114, 156].

In many of these contexts, ML models are trained via Supervised Machine Learning (SML), which is a subcategory of machine learning, whereby computational models or algorithms learn to correctly classify input data, or predict future outcomes states from input data, by leveraging coded, real-word training samples [47, 82]. Training samples include representative task data (e.g., images, sounds, motion data, GPS coordinates, stock prices) that have been labelled with the correct data class or outcome state. These training samples are then used to build an approximate model (function or policy) of how the input data (e.g., the pixels from an image) maps to the correct output label (e.g., cat or dog) [20, 56]. Following training, the efficacy of the resultant model is tested against data not supplied during training, with effective models able to generalise the learned input-output mappings and, ultimately, correctly classify or predict unlabelled data.

4.1.1 Artificial neural and Long Short-Term Memory networks

SML models can be realised in numerous ways. For instance, using decision trees [124, 132], random forests models [17], support vector machines [29, 31] and, of particular importance here, Artificial Neural Networks (ANNs).

In general, ANNs are a composition of elementary units, nodes, that are grouped in interconnected layers, where the connections between nodes can have different weights. A typical ANN includes an input and an output layer, with 1 or more “hidden layers” in between (i.e., deeper ANNs have more hidden layers). Training an ANN to model input-output mappings via SML requires finding the combination of connection weights that maps input data to the correct output class (label) or state prediction. This is achieved by iteratively evaluating the error between the correct and predicted output of the ANN and adjusting the network weights between nodes and layers in order to minimise this error. The error is measured via a loss function, which typically involves calculating the mean squared error or the negative average of the log of the probabilities between the predicted and the correct output states. Network weights are adjusted via a process known as back-propagation, which over training iterations adjusts network weights until the loss function is minimised (i.e., converges/descends to a minimal value; refer to [10] for further details on ANNs).

There are various types of ANNs. For instance, a fully-connected ANN is an ANN where all of the

nodes on one layer are connected to all of the nodes on the subsequent layer. Of relevance in this and following Chapters, is an ANN known as *Long Short-Term Memory* network (hereafter referred to as LSTM$_{NN}$). An LSTM$_{NN}$ is a form of recurrent neural network that in addition to feed-forward connections among node layers also includes feedback connections. The significance of these feedback connections is that they allow the ANN to process and retain information about sequences of consecutive inputs [60, 61]. Accordingly, LSTM$_{NN}$s are commonly used in time-series prediction tasks [3, 21, 45, 111, 167], where the processing and retention of both past and present input states are required to correctly predict future states. LSTM$_{NN}$s are applicable here, as human decision making during skilful action is based on the assessment of dynamical (time varying) task information [7, 176] and, thus, the prediction of future state decisions requires processing sequences of task relevant state input data.

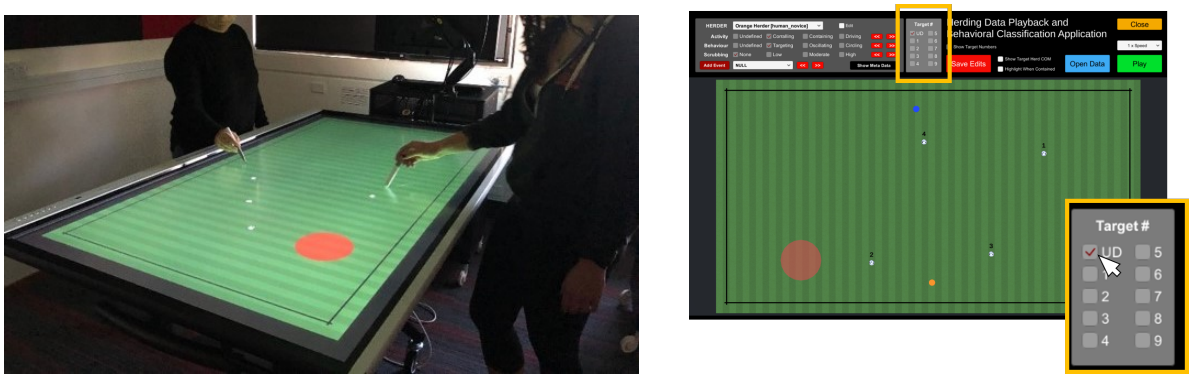
4.2 Human target selection decisions

Effective behavioural coordination is essential to the performance of many everyday social or multi-agent tasks. As discussed in previous Chapters, the multi-agent herding scenario offers a paradigmatic example of a complex multi-agent coordination task [109, 110], in which one or more "herder" agents are required to coordinate their actions to corral a second group of "target" agents into a desired containment area.

As detailed in Chapter 2, different approaches have been proposed to design the control policies of herders [67, 87, 113, 120, 151, 160] with coordination both explicitly embedded in the herder dynamics [67, 84, 120] and implicitly conveyed in how each herder selects which target to corral at each time step. Indeed, key to the success of any herding model is the target selection strategy that regulates which target a herder is attracted towards and corrals at any given time ([9], Chapter 3). Despite the central role that such action decisions play in the success of herding, the target selection rules that have been employed when modelling human performance or developing artificial herders are typically derived and implemented heuristically [67, 84, 151] under the general assumption that target agents tend to aggregate together. Indeed, almost no empirical research has been directed to determining the validity of these heuristic selection strategies, nor what specific task information (e.g., relative location or velocity of targets, relative position of co-herds) herders use to make target selection decisions.

In human-machine interaction, where the complex multi-agent system is composed of both human and artificial actors, understanding how humans select their targets is also important for developing artificial agents capable of successful and robust human interaction [32, 64]. Indeed, like human-human interaction, effective human-machine interaction rests on the ability of artificial agents (AAs) to not only predict the future actions or action decisions of human actors [19, 43, 122], but, in many instances, also enact human-understandable patterns of reciprocal behavioural action [15, 19, 62, 81, 109, 166]. Moreover, AI-driven agents shaped on successful human behaviour

are also able to better train human actors for future human-human task performance [129]. One way of developing artificial agent herders capable of robust human-AA interaction is to couple human inspired local control laws (3.5)-(3.6) with human inspired target selection strategies. Our aim here was to investigate this by first modelling the target selection decisions of human herders who completed a tabletop version of the multi-agent herding task as illustrated in Figure 4.1 and then implement and test these models of AA herders.



(a) (b)

Figure 4.1: The herding task and experimental setup (a); and a screenshot of the playback application developed in Unity3D (b).

Successful novice and expert human performance data from the joint-action herding experiments conducted by [129] was employed to model the target selection strategies of human herders. For this version of the herding task, novice-novice and expert-expert human herders were required to corral and contain target agents randomly disperse around a game field into a predefined containment area. Specifically, the herding task (game), developed with Unity-3D game engine (Unity Technologies LTD, CA), required pairs ($N_H = 2$) of human participants (players) to control virtual herding agents to corral and contain $N_T = 4$ randomly moving target agents within a designated containment area positioned at the centre of a game field. The task was performed on large 70" touch screen monitors (see Figure 4.1()), with the human participants using touch pens to control the location of motion of the herder agents. The targets' movement dynamics were defined by Brownian motion when not being influenced by a herder agent, and when influenced by a herder agent moving in a direction away from the herder agent (similar to (3.3) detailed in Chapter 3). During task performance, the position and velocity of all herders and targets agents (as well as other general game states) was recorded at 50 Hz[1]. Full details of the experimental set-up and data collection process can be found in [129], where the authors also developed "human-like" artificial agent trainers using the dynamical herding model and heuristic target selection strategy proposed in [107, 109] and detailed in Chapter 3 Section 3.3 and Appendix A.

[1]Data available at github.com/FabLtt/ExplainedDecisions.

Given that there were 4 targets and that players could also choose to corral no-target, predicting the target selection decisions of novice and expert human herders corresponded to a multi-label (i.e., 5 label) prediction problem with the output prediction class ID = 0 corresponding to no target and ID = 1, 2, 3, or 4 corresponding to the four targets, respectively.

4.3 Predicting human target selection decisions

To exploit the time-varying nature of our data, we built an Artificial Neural Network (ANN) by alternating Long-Short Term Memory (LSTM) and Dropout layers [60, 61]. We used Bayesian Optimisation to tune the learning rate of the Adam optimiser and the hyperparameters of the ANN. To avoid over-fitting, we stopped the training phase when the Logarithmic loss on the validation set stopped improving. Appendix B.5 has additional details on the learning algorithm.

To train an LSTM_{NN} using SML, we extracted successful state time-series data from novice-novice and expert-expert human pairs who completed the tabletop herding task detailed above. State data was extracted from the time of task onset to when the herders had corralled all the four target agents inside the containment area for the first time. We disregarded herder behaviour during the remaining of the trials as we were only interested in trying to predict target selection decisions prior the onset of the oscillatory containment behaviour the human herders often exhibit once targets are contained within the containment area. This is because human players treat the target herd as a single entity for the latter mode of behavioural containment, with individual target selection decisions no longer occurring [106, 109, 129].

Novice data was extracted from 40 successful trials performed by 10 different novice pairs (4 successful trials from each pair). Note that a human herder was considered to be a "novice" if they were unfamiliar with the herding task prior to the data collection session. Novices repeated the task until they had completed the 4 successful trials included in the novice data-set, with an average of 8 unsuccessful trials per pair.

Expert data was extracted from 48 successful trials performed by 3 pairs of human players with extensive experience ($\gg$10 hours and $\gg$100 successful trials) performing the simulated multi-agent herding task (i.e., several authors from [129]).

At each time step the target a herder was corralling was classified manually by a paid research assistant (naive to the studies purpose) via an interactive data playback Unity3D[2] application that played back the original recorded data-set (see Figure 4.1(a)). Data playback speed could be decreased to 1/8 speed, as well as stepped frame by frame, with each target visually labelled with a fixed number (1 to 4) allowing the research assistant to more easily detect the currently corralled target in case of uncertainty (e.g., a herder being close to two targets at the same time). Additionally, the resulting classification was compared with the results of an automatic coding software. The software leveraged the experimenter's knowledge of target agents' kinematics; in fact, each target was

[2] https://unity.com/, version 2018LTS

designed to move at constant speed in a random direction if not corralled, and to increase its velocity if targeted by a herder. The comparison showed that the naive researcher's and the automatic codes matched at 81.5% but the naive researcher was considered to be more in tune with human decision-making processes. As a result of such manual classification, the target agent a given human herder was corralling was labelled with an integer number $\tilde{i} \in [0, \hat{N}_T + 1]$, with $\tilde{i} = 0$ meaning "no target agent corralled" and $\tilde{i} \neq 0$ being the class ID of the target agent being corralled.

From the resulting labelled time-series data, training samples of expert and novice herders were constructed of length t_i to t_f, where $t_f - t_i = T_{seq}$ (corresponding to 25 time steps or 0.5 seconds of system evolution), with 48 input features; i.e., the *relative distances* between herders and between the herders and each of the 4 targets, each herders' and targets' *distance from the centre of the containment area*, and the *velocities* and *accelerations* of each herder and target. See Appendix B.4 for more details on the feature extraction process and LSTM$_{NN}$ training data[3].

4.3.1 Prediction analysis

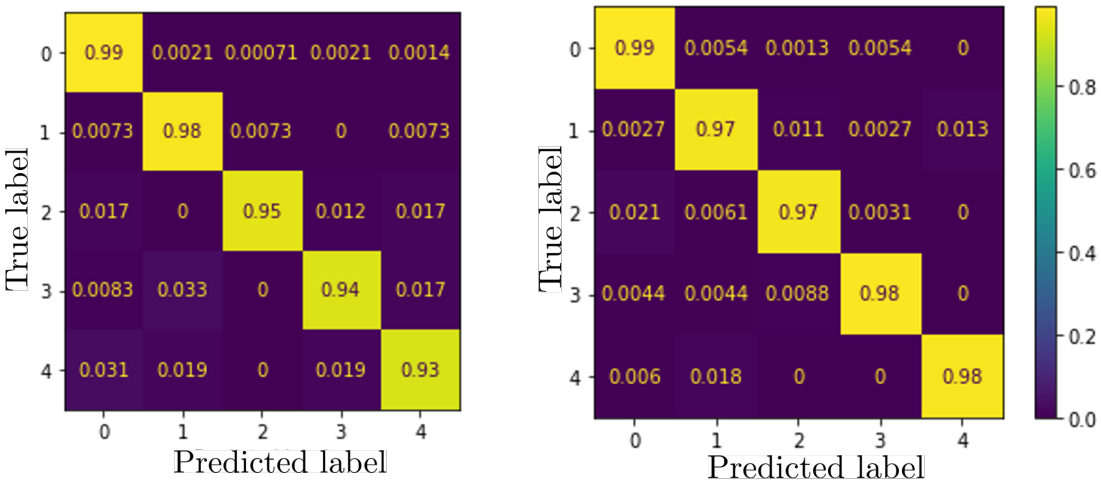
Separate LSTM$_{NN}$ were then trained[4] to predict the next target ID (output label) a novice or an expert herder would corral at $t_{f+\tau_{hor}}$ given a feature input sequence T_{seq}, with a prediction horizon $\tau_{hor} = 1$ time step, which corresponded to predicting the target the herder would corral at the next time-step (equivalent to 20 ms in the future).

The confusion matrices of each LSTM$_{NN}$ tested on a test set of data samples (i.e., samples not employed during training) are reported in Figure 4.2. Each label represents the ID of the target agent the herder would corral at $t_{f+\tau_{hor}}$, with True/Predicted label $\neq 0$ for targets 1 to 4 and True/Predicted label = 0 when no target agent would be corralled (note that the terms 'predicted label' and 'prediction output' will be employed interchangeably). Importantly, the values on the diagonal indicate that each target ID could be correctly predicted between 93% to 99% of the time. Indeed, independent from player expertise, each LSTM$_{NN}$ predicted which target a herder would corral with an average accuracy above 95% (see Table 4.1 for more prediction metrics averaged over multiple test sets and defined in Appendix B.6).

	Accuracy	Precision	Recall	F1 score
Novice	97.90	97.75	97.72	97.73
Expert	98.10	95.89	95.89	95.92

Table 4.1: Average performance [%] of the multi-label predictor trained on sequences of length T_{seq} = 0.5 s on a test set of N_{test}= 2000 samples.

[3]The processed datasets are made available in the public repository https://osf.io/wgk8e/
[4]Code and trained ANNs available at github.com/FabLtt/ExplainedDecisions.



(a) Expert (b) Novice

Figure 4.2: Confusion matrices of the multi-label predictor trained on time-series of length T_{seq} = 0.5 s and tested on N_{test}= 2000 samples for expert (a) and novice (b) data. ID = 0 correspond to "no target", with ID = 1, 2, 3 and 4 corresponding to the actual targets. Values on the diagonal indicates the portion of samples for each prediction output correctly predicted.

4.4 Employing human-inspired strategies for artificial herders

To demonstrate the utility of the proposed modelling approach for creating AAs capable of of human-like target selection decisions, we used the novice and expert LSTM models to control the behaviour of AA-herders. As in Chapter 3, we considered the problem of N_H (artificial) herder agents tasked to corral N_T autonomous target agents. Each AA herder targeted one agent at a time, with the target agent to be corralled determined by either the novice or the expert LSTM$_{NN}$ models. That is, the LSTM$_{NN}$ models were employed as *human-inspired target selection decision* strategies for the artificial herders (AA-herders) completing the same tabletop version of the herding task (Figure 4.1). In what follows, we will refer to the LSTM$_{NN}$ models trained on novice and expert pairs as *novice-inspired* and *expert-inspired* selection strategies, respectively.

The N_T target agents were modelled as stochastic autonomous agents (3.3) subject to a repulsive force away from the herders (3.4); the movements of the AA-herders were controlled by the dynamical model, (3.5)-(3.6), that has been previously demonstrated to capture the movement dynamics of human herders, as detailed in Chapter 3.

For each artificial herder, the LSTM$_{NN}$ models received real-time N_{sv} state features sequences for the current AA-herder, other AA-herders, and the subset of $\hat{N}_T = 4$ stochastic agents closest to the current herder. At the beginning of a trial AA-herders were labelled from 1 to N_H in anticlockwise order starting from a randomly selected herder. The LSTM$_{NN}$ models outputted the ID of the targets (ID = 1 to 4) or no-target (ID = 0), among the $\hat{N}_T \leq N_T$ selected, to be targeted by the AA-herder. If the selection strategy output was "no target" (ID = 0), then the herder would move towards and corral

the target, among the $\hat{N}_T$ selected, farthest from itself.

4.5 Numerical validation

To test the generalizability of the human-inspired target selection models (and modelling approach detailed here), we considered $N_H = 2$ and $N_H = 3$ artificial herders that were required to corral $N_T = 4$ and $N_T = 7$ target agents and performed 50 trials with different initial conditions for each combination of LSTM$_{NN}$ policy (novice-inspired, expert-inspired), AA-herder number ($N_H = 2$ and $N_H = 3$) and number of targets ($N_T = 4$ and $N_T = 7$). Note that AA-herders in the same herding pair or triad were driven by the same target selection policy (i.e., there were no mixed novice-expert herder groups). All AA-herders started from a distance of $6r^\star$ from the centre of the goal region, with a relative angular displacement of $(2\pi)/N_H$. Target agents initial positions were selected randomly in a circular sector around the goal region with mean radius of $3r^\star$. Appendix B.2 has more details on the numerical integration[5] and simulation parameters.

To evaluate the performance of novice- and expert-inspired AA-herders, we employed similar measures to those previously employed to evaluate novice- and expert-human herding performance [109, 129] and AA-herding performance ([9], Appendix B.1). More specifically, we measured the mean distance travelled by the herders d_{g}, the spread rate of the herd $S_{g,\%}$ and the containment rate $I_{g,\%}$ during the gathering time t_{g}, the time from the beginning of a herding trial to the first time instant at which the targets in the herd were all contained. Note that after t_{g}, all herders were able to maintain the herd (on average) inside the goal region until the end of the trial.

As can be seen from an inspection of Table 4.2, both the novice- and expert-inspired AA-herders were able to successfully herd the target agents, irrespective of the number of targets and the number of AA-herders. Interestingly, the overall performance of the novice- and expert-inspired AA-herders was relatively similar with regards to each of the performance measures assessed, with significant differences only observed for containment rate, $I_{g,\%}$, (Kruskal Wallis test [79]; $\chi^2 = 9.45$, $p < 0.003$) when $N_H = 3$ and $N_T = 7$ and the distance travelled by the AA-herders, d_{g} ($\chi^2 = 5.56$, $p < 0.02$) when $N_H = 2$ and $N_T = 7$, with expert-inspired AA-agents exhibiting slightly faster target containment rates, while travelling over more ground, respectively, compared to novices.

Recall, that the novice-human data employed to train the novice-inspired model was extracted from successful herding trials. Thus, while the novice human herders were not 'experts', they had reached a level of behavioural performance consequent with "successful herding". Indeed, the difference in the actual performance of the expert and novice human herders was also rather minimal as can be observed in Table 4.3. Hence, finding little difference in the performance outcome measures of the novice-inspired and expert-inspired herders was partially expected. However, an additional robustness analysis[6] in which the herding performance of novice- and expert-inspired herders were tested on herd-size N_T from 6 to 66 targets revealed that the expert-inspired herders

[5] Code available at github.com/FabLtt/DeepHerding
[6] Code available at github.com/FabLtt/DeepHerding

	$N_T = 4$		$N_T = 7$	
	Novice-insp	Expert-insp	Novice-insp	Expert-insp
		$N_H = 2$		
t_g [a.u.]	13.72 ± 4.4	13.06 ± 4.03	18.42 ± 4	18.50 ± 3.2
d_g [a.u.]	151.05 ± 64.1	143.95 ± 60.7	197.70 ± 51.1	216.10 ± 50.5
D_g [a.u.]	2.2 ± 1.5	2.45 ± 1.66	2.8 ± 1.6	3.35 ± 2.12
$S_{g,\%}$ [%]	1.32 ± 0.6	1.33 ± 0.6	2.39 ± 0.5	2.34 ± 0.6
$I_{g,\%}$ [%]	22.81 ± 8.8	21.06 ± 6.6	22.35 ± 9.2	21.18 ± 6.6
		$N_H = 3$		
t_g [a.u.]	10.17 ± 2.7	10.31 ± 3.3	14.86 ± 6.3	13.37 ± 2.7
d_g [a.u.]	114.81 ± 36.2	124.00 ± 47.6	166.44 ± 65.4	164.53 ± 37.8
D_g [a.u.]	2.12 ± 1.45	2.46 ± 1.45	3.1 ± 1.6	3.23 ± 2.15
$S_{g,\%}$ [%]	1.156 ± 0.5	1.16 ± 0.5	2.27 ± 0.5	2.36 ± 0.4
$I_{g,\%}$ [%]	25 ± 8.2	24.63 ± 7.8	25.98 ± 9.2	20.89 ± 4.9

Table 4.2: Average performance over 50 trials of different human-inspired target selection strategies for $N_T = 4$ and $N_T = 7$ target agents during gathering time t_g.

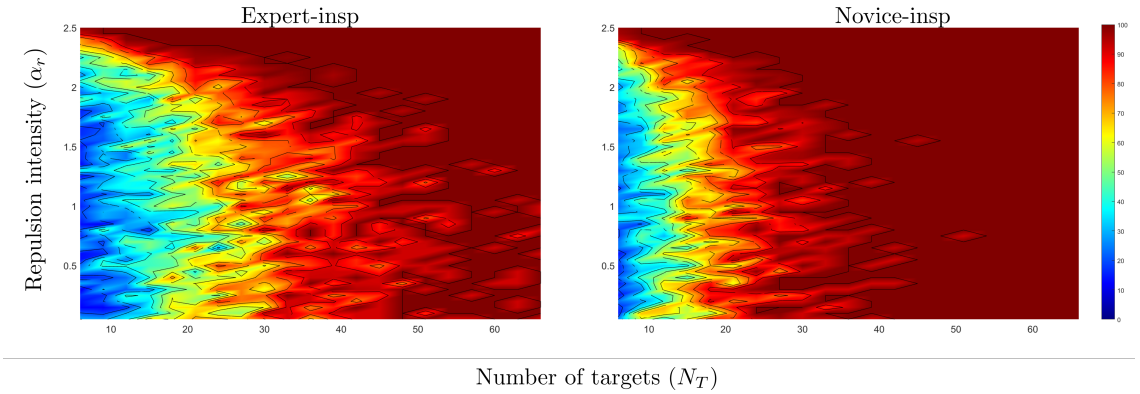
did, in general, preform better than novice-inspired herders for herd-sizes >15 targets in terms of gathering time t_g (blue area in Figure 4.3(a)) and containment rate $I_{tot,\%}$ (red area in Figure 4.3(b))

	Novice pairs	Expert pairs
t_g [a.u.]	26.35 ± 8.67	10.65 ± 3.43
d_g [a.u.]	7.19 ± 4.08	4.6 ± 1.6
D_g [a.u.]	0.99 ± 0.2	0.35 ± 0.22
$S_{g,\%}$ [%]	3.61 ± 1.99	2.7 ± 1.48
$I_{g,\%}$ [%]	18.25 ± 7.18	16.81 ± 6.76

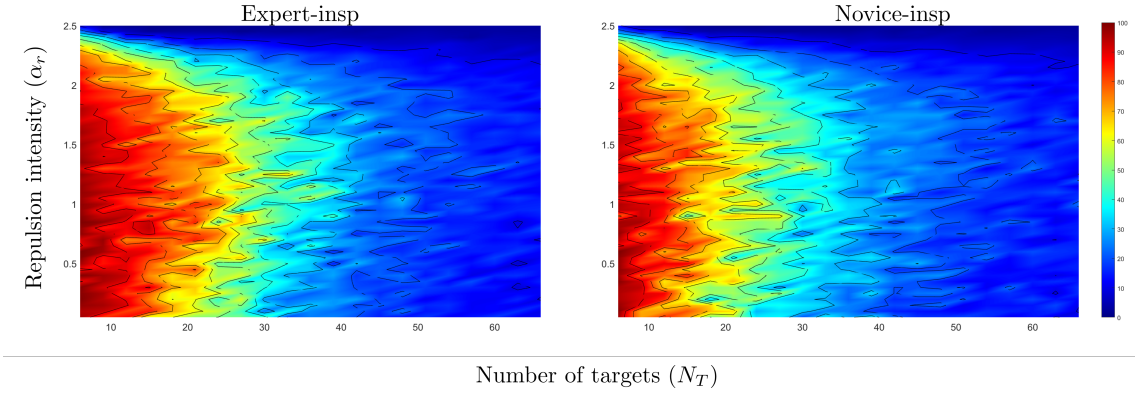
Table 4.3: Average performance over successful trials of novice and expert pairs of actor playing the herding game with $N_T = 4$ passive agents during gathering time t_g. According to Kruskal Wallis statistical test, significant difference between Novice and Expert pairs is only observed for gathering time t_g ($\chi^2 = 24.67$, $p < 0.0001$), distance travelled during gathering phase d_g ($\chi^2 = 5.76$, $p < 0.02$) and herd distance from the containment region during gathering phase D_g ($\chi^2 = 24.33$, $p < 0.0001$) with pairs of experts exhibiting faster target gathering time compared to novices.

4.6 Comparison among target selection strategies

We compared the human-inspired selection strategies against the target selection strategies investigated in Chapter 3 (i.e., the global, static, leader-follower, peer-to-peer) in the benchmark scenario of $N_H = 3$ herders and $N_T = 7$ target agents (see Table 3.3. Recall that the global search strategy and the static arena partitioning are the two rule based strategies that have typically been employed in



(a) Gathering time t_g. Lower values correspond to faster herding.



(b) Total containment rate $I_{\text{tot},\%}$. Higher values correspond to more efficient herding.

Figure 4.3: Robustness analysis of the proposed herding strategies for two herders ($N_H = 2$) to variation of herd size N_T and repulsive reaction coefficient α_r in (3.4). N_T was varied between 6 and 66 agents, with increments equal to 3, while α_r between 0.05 and 2.5, with increments equal to 0.05. For each pair (N_T, α_r) the corresponding metric was averaged over 10 simulation trials starting with random initial positions. The coloured plots were obtained by interpolation of the computed values.

the herding literature [67, 107, 151]. In these approaches, either herders distributed among themselves the N_H agents farthest from the centre of the goal region (*global search strategy*) or they evenly and statically split the space in N_H static sectors and each selected the farthest agent in its own sector (*static arena partitioning*). The other two strategies were the novel strategies proposed in Chapter 3. For the leader-follower strategy the herding space was partitioned dynamically between the different herders depending on the position of the "lead" herder. For the peer-to-peer strategy, the herding space was dynamical split into equally sized partitions as a function of each herder's local position.

From Table 4.4, the simulation results revealed that AA-herders could successfully achieve the task (3.2) within the first 20% of a trial, independent of the target selection strategy employed. Consistent with the results reported in Chapter 3.4, Table 4.4 the global target selection strategy

outperformed the other strategies. However, it is important to appreciate that the global target selection strategy is computationally heavy, as it depends on full knowledge of the environment and accurate estimates of agents distance from the goal. This is in contrast to the significantly lower computational cost of the human inspired strategies developed here and the other three strategies investigated in Chapter 3, which only rely on partial or local information. Note also that although AA herders enacting the static strategy exhibited significantly higher containment rates ($I_{g,\%} = 34.4 \pm 16.6$) compared to the other strategies, this was achieved at the cost of a significantly higher gathering time ($t_g = 32.02 \pm 25.8$) and greater distance travelled ($d_g = 216 \pm 189$).

While the global and static strategies provided effective benchmarks for assessing the overall performance of the human inspired strategies, of more interest was the comparison of the human inspired models to the two dynamic strategies investigated in Chapter 3 (i.e, leader-follower and peer-to-peer), which we assumed were more "human like" than the global and static strategies[129]. Consistent with the latter assumption, AAs controlled via the human-inspired and dynamic selection strategies exhibited similar levels of task performance, with the exception that AAs dynamically dividing the game space as peer-to-peer which took significantly longer than the human-inspired (and the leader-follower) strategies to corral the targets (t_g, $\chi^2 \approx 8$, $p < 0.02$), but also travelling significantly shorter distances (d_g, $\chi^2 \approx 28.7$, $p < 0.0001$) compared to the human-inspired (and the leader-follower) strategies. This is not to say, however, the human inspired strategies were more similar to the leader-follower strategy, as the human inspired strategies did exhibit (on average) much faster gathering times, t_g, than the leader-follower strategy.

	Global	Static	LF	P2P	Novice-insp	Expert-insp
t_g [a.u.]	10.5 ± 1.2	32.02 ± 25.8	31.3 ± 18.1	17.2 ± 8.3	14.86 ± 6.3	13.37 ± 2.7
d_g [a.u.]	102 ± 21	216 ± 189	169 ± 91	101 ± 44	166.44 ± 65.4	164.53 ± 37.8
D_g [a.u.]	2.5 ± 1.7	3.3 ± 1.6	3.4 ± 1.8	2.5 ± 1.6	3.1 ± 1.6	3.23 ± 2.15
$S_{g,\%}$ [%]	2.6 ± 0.6	2.2 ± 0.5	3.1 ± 1.2	2.8 ± 0.9	2.27 ± 0.5	2.36 ± 0.4
$I_{g,\%}$ [%]	11.12 ± 3.2	34.4 ± 16.6	27.6 ± 12.6	20.5 ± 6	25.98 ± 9.2	20.89 ± 4.9

Table 4.4: Average performance over 50 trials of benchmarking (Global, Static), dynamical (LF, P2P) and human inspired (Novice-, Expert-insp) target selection strategies for $N_H = 3$ and $N_T = 7$ target agents during gathering time t_g (see Appendix B.1 for the formal definitions).

4.7 Summary

In this Chapter we demonstrated how short state information (feature) sequences (i.e., 500 ms of herding system evolution) could be used to train LSTM_{NN} models to accurately predict (above 95% on average) the target selection decisions of human actors, independently of expertise level (Section 4.2-4.3). Additionally, we provided evidence that the resultant LSTM_{NN} models of human target selection decisions could be employed to control the decision dynamics of AAs. We demon-

strated this for both the expert and novice models (Section 4.4) and, moreover, demonstrated how the models could be used to control the decision making behaviour of AAs within multi-agent task contexts not included in the training data (Section 4.5). The generalised herding contexts included both increases in the number of herding agents and the number of targets to be corralled, as well as across a wide range of target repulsion magnitudes (see Figure 4.3). In particular, the numerical experiments reported here highlighted that herders adopting human-inspired strategies can be successful (i) with or without herd flocking behaviour and (ii) do not need global knowledge of the environment, nor knowledge of the exact herd size (i.e., have low computational cost).

Similar results, both in terms of prediction and herding performance, were also observed for LSTM_{NN} trained to predict the target selection decisions much further in the future (i.e., longer prediction horizon). The results for the short prediction horizon models were presented here, while the results for the long prediction horizon models (i.e., 0.5, 1 and 2 seconds in the future) are reported in Appendix C.3.

Finally, it is important to appreciate that the dynamic target selection strategies presented in Chapter 3 and employed for comparison purposes here reflect white-box policies that exploit global information to dynamically partition the searching space. In contrast, the the human-inspired strategies models developed and investigated here correspond to black-box policies (i.e., the mapping to between input and output states is unknown). Thus, while the result suggest that the human inspired target selection models were more similar to the two dynamical targets selection strategies investigated in Chapter 3 (i.e, leader-follower and peer-to-peer), it is very difficult to draw any definitive conclusions about this potential relationship, without understanding the information (features) the human herders employed to make their target selection decisions. Accordingly, to better understand the decision making behaviour of human herders, in the next Chapter we employed explainable AI (Artificial Intelligence) tools to identify the state information that underlies the target selection decisions of human herders, including what differentiates expert from non-expert performance.

CHAPTER 5

UNDERSTANDING HUMAN DECISION IN FAST PACED JOINT TASK

Chapter 4 provided evidence that it is possible to train artificial neural networks ($LSTM_{NN}$s) to model the target selection decisions of human herders completing the tabletop, multi-agent herding task developed by [106, 129]. Recall that this task was chosen because it (i) provided a paradigmatic example of a multi-agent task where skilful action or expertise is required for task success and (ii) rests on the ability of herders to rapidly and continuously coordinate their target selection decisions throughout task performance.

In this Chapter, we first detail how the target selection decisions of human herders can be predicted at timescales, or prediction horizons, that extend beyond the timescale at which target section actions are enacted (Section 5.1). That is, we demonstrate how $LSTM_{NN}$ can be trained to predict the target selection decisions of human herders well before those target selection actions are enacted. Following a more advanced analysis of the $LSTM_{NN}$ models trained on novice and expert data (Section 5.3), we then demonstrate how explainable AI tools can be employed to identify differences in the decision-making processes of expert and novice herders. More specifically, we demonstrate how the $LSTM_{NN}$ models that were able to accurately predict the target selection decisions of expert or novice herders differently weighted task information (feature inputs) (Section 5.4).

The work in this Chapter, with results in Chapter 4, has been prepared for publication.

5.1 Predicting fast paced distributed human decision

As noted in the previous Chapter, a timely and important challenge – given the increasing potential for interactive robotic and artificial intelligence (AI) technologies to significantly enhance and expand individual, and team, performance capabilities – is modelling the decision making processes of human actors for the design of interactive artificial agents (AAs). While effective prediction is of-

ten of primary concern in the development of such models [76, 131], understanding what differentiates human performance, including what state information human actors employ when making action decisions is equally important [28, 63, 100, 133]. Here, we investigated whether we could leverage recent advances in explainable AI methods to understand the decision-making activity of expert and non-expert (novice) human actors performing the fast-paced, tabletop, multi-agent herding task [9, 107].

In Chapter 4, the target selection decision of novice and expert pairs of human herders were modelled using Supervised Machine Learning (SML) and LSTM$_{NN}$. Given a brief sequence of system evolution (T_{seq} = 500 ms), the resulting models could effectively predict the ID (1 to 4, or 0 for no target) of the target the herder would corral 20 ms in the future. Here, we model the same target selection predictions, but this time for sequence lengths of T_{seq} = 1 s and with respect to two different prediction horizons: a shorter (immediate) prediction horizon, $\tau_{hor} = 1$, which corresponded to predicting the target the herder would corral 40 ms in the future; and a much longer (delayed) prediction horizon of $\tau_{hor} = 16$, which corresponded to predicting the target the herder would corral 640 ms in the future. As in the previous Chapter, data of novice and expert herders completing the herding task presented in [129] was employed for model training with pairs of human herders sharing the same level of expertise. More specifically, we (i) extracted T_{seq}-long time series of relevant state variable (i.e., input features), (ii) chose the corresponding output label (i.e., ID of the agent targeted) τ_{hor} time steps later and (iii) trained an LSTM-layered ANN for each combination of expertise and τ_{hor} as described in Chapter 4 and Appendices B.4-B.6[1].

5.1.1 Immediate and delayed decision

Choosing T_{seq} = 1 s, the shorter prediction horizon of $\tau_{hor} = 1$ (or 40 ms in the future), essentially reflected decisions already made and/or currently being enacted by a human herder. In contrast, the longer prediction horizon of $\tau_{hor} = 16$ (or 640 ms in the future), involved predicting the target selection decision of a human herder well before the decision was enacted and/or a player's behavioural intention was typically observable. This was validated by calculating the average time it took a human herder to move from one target to the next when switching targets. The latter was determined by calculating the human herders inter-target switching time. These inter-target switching movement times were calculated as the time from when a human herder moved outside the region of repulsive influence of the current target being corralled and entered the region of repulsive influence of the next target to be corralled, with an average inter-target movement time of 556 ms for novice herders and 470 ms for expert herders (see Figure 5.1 for the distributions of inter-target movement times). Note we also investigated $\tau_{hor} = 8$ and 32, see Appendix C.2.

[1] Code and trained ANNs available at github.com/FabLtt/ExplainedDecisions

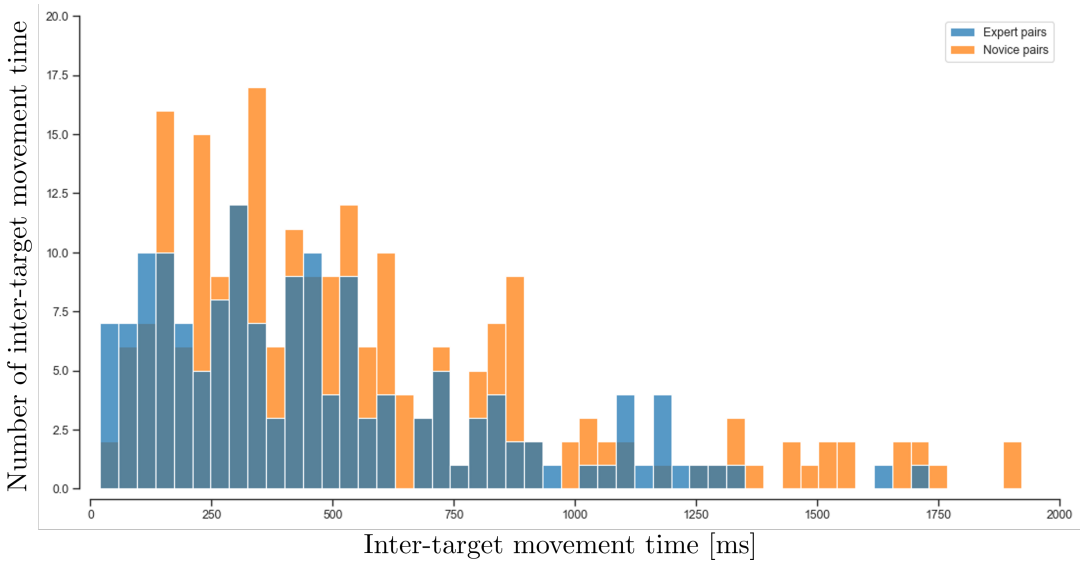


Figure 5.1: Distribution of inter-target movement times [ms] of expert (blue) and novice (orange) herders. The average inter-target movement time was 556 ms (65% of the total inter-target movement times < 600 ms) for novices and 470 ms (72.5% of the total inter-target movement times< 600 ms) for experts.

5.1.2 Prediction analysis

Prediction performance reported in Table 5.1, and defined in Appendix B.6, reveals how SML trained LSTM$_{NN}$ models could predict which target a herder would corral at an average accuracy exceeding 95%, independent of prediction horizon and level of expertise (see Figure 5.2 for the confusion matrices of each LSTM$_{NN}$ when tested on a benchmark set of data samples). Supplementary results for τ_{hor} = 8 and 32 are also reported Appendix C.2.

	Accuracy	Precision	Recall	F1 score
Shorter timescale (40 ms)				
Novice	96.7 ± 0.4	96.6 ± 0.4	96.5 ± 0.4	96.6 ± 0.4
Expert	96.8 ± 0.2	93.7 ± 0.7	92.4 ± 0.7	93.1 ± 0.6
Longer timescale (640 ms)				
Novice	96.2 ± 0.5	96.2 ± 0.6	96.1 ± 0.6	96.1 ± 0.6
Expert	95.6 ± 0.3	90.7 ± 0.6	89.4 ± 0.9	90 ± 0.7

Table 5.1: Average performance [%] of the multi-label predictor trained on time-series of length T_{seq} = 1 s and tested on 10 sets of N_{test}= 2000 samples.

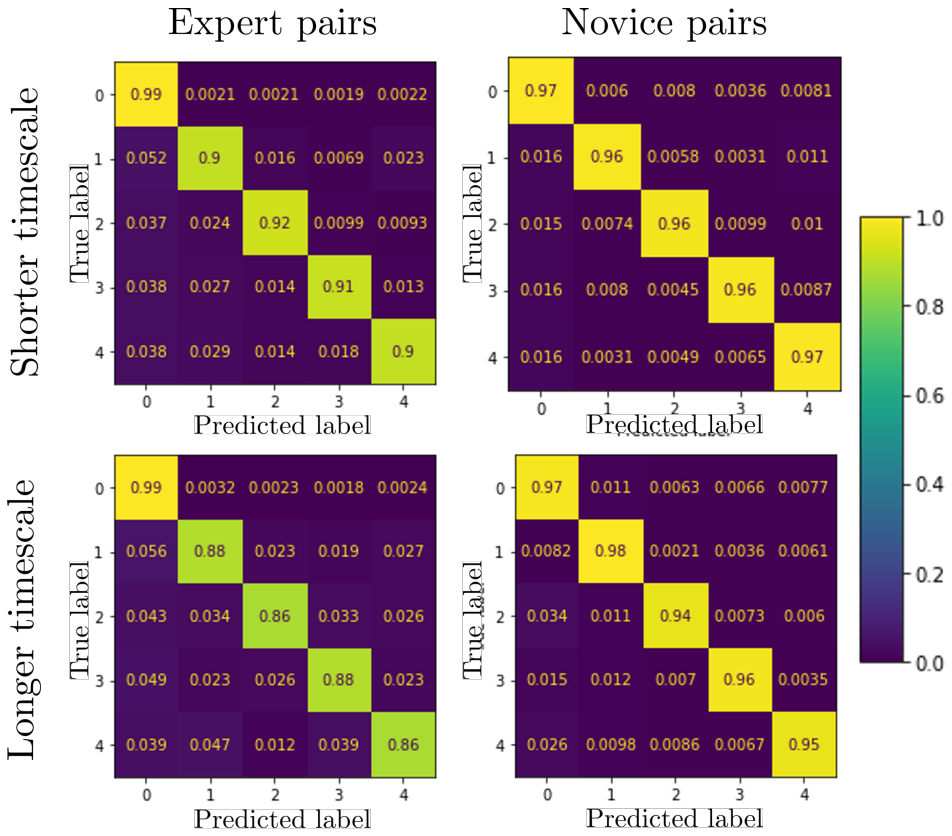


Figure 5.2: Confusion matrices of the multi-label predictor trained on time-series of length $T_{seq} = 1$ s and tested on N_{test}= 2000 samples for different combination of expertise (columns) and prediction horizon (rows).

5.2 Sample type prediction analysis

Recall that the data samples used to make a target selection prediction are vector time-series of the herding system's state evolution for $t \in [t_i, t_f]$, where $t_f - t_i = T_{seq}$, and the prediction outputs are chosen as the ID of the target that will be corralled at $t_{f+\tau_{hor}}$ with $\tau_{hor} \neq 0$. Accordingly, it is important to appreciate that there are two sub-categories of sample type (see Figure 5.3). First, in the time interval T_{seq}, a human herder could either continuously corral the same target agent or *transition* between different targets. Here we classified these two types of T_{seq} samples as *non-transitioning* and *transitioning* behavioural sequences, respectively. The second subcategory of sample, corresponded to whether a herder *switched* targets between t_f and $t_{f+T_{hor}}$. That is, at T_{hor}, a herder could be corralling the same target that was being corralled at the end of T_{seq} or switch to a different target, with these two possibilities classified as *non-switching* and *switching* behaviour, respectively.

Interestingly, the majority of samples, at $\tau_{hor} = 1$, in the expert and novice data-set corresponded to "non-transitioning and non-switching" behaviour (79.28% and 63.87%, respectively). The implication is that experts transitioned between targets less often than novices and were more persistent

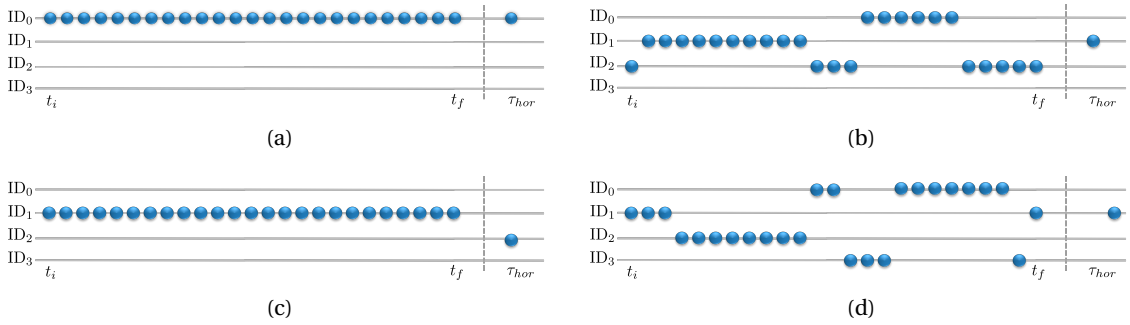


Figure 5.3: Examples of samples in different categories: (a) “non transitioning and non switching” sample, (b) “transitioning and switching” sample, (c) “non transitioning and switching” sample and (d) “transitioning and non switching” sample.

	Non transitioning				Transitioning				Mixed	
	Non switching		Switching		Non switching		Switching			
	Accuracy	% samples	Accuracy	% samples	Accuracy	% samples	Accuracy	% samples	Accuracy	% samples
Shorter timescale (40 ms)										
Novice	97.92%	63.87%	63.18%	2.41%	97.56%	33.14%	69.25%	0.58%	96.75%	100%
Expert	98.86%	79.28%	55%	1.16%	92.16%	18.88%	57.34%	0.68%	96.82%	100%
Longer timescale (640 ms)										
Novice	97.35%	48.61%	92.48%	17.69%	97.69%	26.59%	92.22%	7.12%	96.21%	100%
Expert	98.86%	72.67%	71.45%	7.63%	89.14%	12.62%	74.24%	7.08%	93.48%	100%

Table 5.2: Average performance [%] of the multi-label predictor, per type of samples, trained on time-series of length T_{seq} = 1 s and tested on multiple sets of N_{test}= 2000 samples each.

in corralling a given target (or no target) compared to novices. Conversely, novice herders switched between targets more frequently and corralled targets for shorter periods of time compared to experts.

Given the latter finding, it was important to determine whether the difference in the amount of one type of sample versus other type of samples in the expert and novice data sets skewed model accuracy. To test this, the LSTM$_{NN}$ models predicting expert and novice target selection decisions at τ_{hor} = 1 and 16 were tested against novel test sets composed of either (i) mixed samples, (ii) only “transitioning” samples, (iii) only “non-transitioning”, (iv) only “ transitioning and switching” samples, and (v) only “non transitioning and switching” samples. Further results for different values of input sequence length T_{seq}, prediction horizon τ_{hor} and type of samples are reported in Appendix D.1.

Results, reported in Table 5.2, demonstrated that model accuracy was dependent on sample type. Specifically, the novice and expert LSTM$_{NN}$’s models exhibited high accuracy (between 89% and 99%) for “non-switching” samples for both prediction horizons and for transitioning and non-transitioning data sequences. However, the accuracy for “switching” samples was significantly lower, particularly for expert models (between 55% and 75%) and when τ_{hor} = 1 (between 55% and 70%).

	Non transitioning				Transitioning				Mixed	
	Non switching		Switching		Non switching		Switching			
	Accuracy	% samples	Accuracy	% samples	Accuracy	% samples	Accuracy	% samples	Accuracy	% samples
Longer timescale (640 ms)										
Novice	95.75%	25%	93.25%	25%	96.23%	25%	95.32%	25%	97.42%	100%
Expert	95.45%	25%	94.38%	25%	97.71%	25%	94.69%	25%	94.72%	100%

Table 5.3: Average performance [%] of the multi-label predictor, per balanced type of samples, trained on time-series of length T_{seq} = 1 s and tested on N_{test}= 2000 samples.

The reason for this reduction in performance was due to a lack of "switching" samples within the training data. For the shorter prediction horizon $\tau_{hor} = 1$, "switching" samples represented less than 2% of samples. As can be discerned from an inspection of Table 5.2, this resulted in an accuracy of only 63% to 70% for switching samples for the $\tau_{hor} = 1$ models. This was not an issue when modelling the target selection decisions of novices at $\tau_{hor} = 16$, as the number of switch samples was above 12%. However, for experts, the number of switch samples was less than 8% when $\tau_{hor} = 16$ and, thus, model accuracy was also much lower, only between 70% and 75% for switching samples.

Due to the significant difference between the type of samples in novice and, especially, expert datasets, we trained LSTM$_{NN}$ models for $\tau_{hor} = 16$ with samples evenly representing each type. As observed in Table 5.3, models trained on an even distribution of behaviours more robustly predict transitioning and switching behaviour, with an average accuracy above 94% for all sample types. Importantly, this indicates that when $\tau_{hor} >> 0$ and there are enough target switching events within the training data, LSTM$_{NN}$ models can be trained to effectively predict when a human herder will remain herding the same target and switch to a new target.

5.3 Cross expertise prediction analysis

The latter results, (Table 5.3) provided clear evidence that the LSTM$_{NN}$ models could accurately predict what target agents expert and novice herders would choose to corral at both short and long prediction horizons. Nevertheless, the difference in the distribution of samples between experts and novices suggested that the specific target selection decisions made by the human herders were dependent on the level of experience. This potential dependence was tested by comparing the performance of the expert trained LSTM$_{NN}$ models attempting to predict novice target selection decisions and vice versa.

As expected, when an LSTM$_{NN}$ trained on one expertise (i.e., a novice trained LSTM$_{NN}$) was used to predict test samples extracted from the opposite expertise (i.e., expert data-set), performance decreased significantly, with the model prediction operating at near chance levels (see Figure 5.4). More specifically, for the shorter prediction horizon, the LSTM$_{NN}$ models trained on expert pairs predicted novice samples with an average accuracy of only 59.6%. Similarly, the model trained on novice pairs only predicted expert samples with an average accuracy of 61.5%. Performance was

even worse for the longer prediction horizon, with average accuracy dropping to 45.18% and 51.58%, for expert-to-novice and novice-to-expert predictions, respectively.

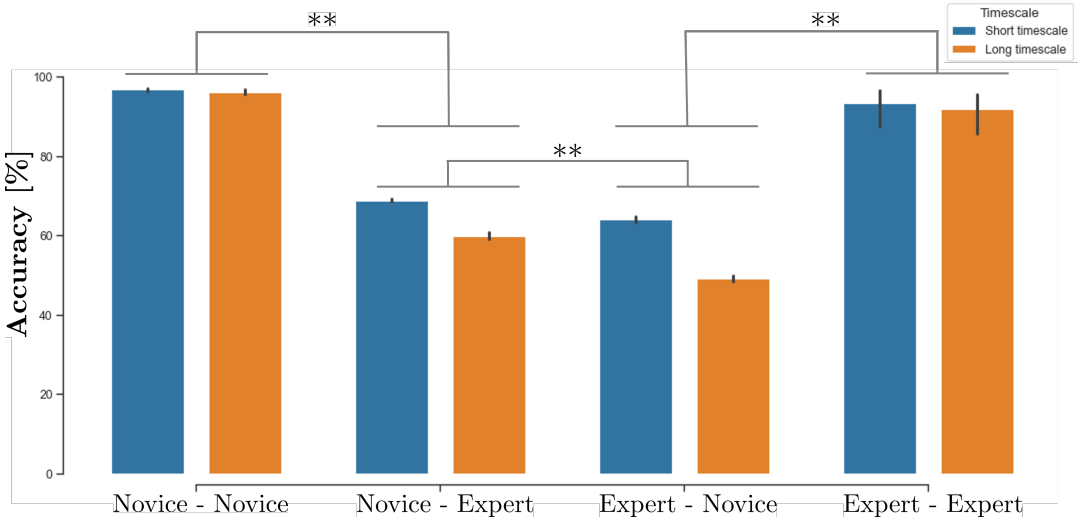


Figure 5.4: Average accuracy [%] values of the multi-label predictor trained on time-series of length T_{seq} = 1 s and tested on 10 test sets of N_{test}= 2000 samples for different combination of expertise between the Train and Test data sets. $**$ indicates a significant paired samples t-test difference of $p < 0.01$. Note there were no significant differences between the accuracy of novice and expert models for both prediction horizon's when tested on the same expertise level (i.e., novice-novice and expert-expert (all $p > 0.1$).

5.4 Understanding differences in human target selection decisions

The significant difference in the performance of LSTM$_{NN}$ models trained and tested on the same level of expertise (Table 5.1) compared to different levels of expertise (Figure 5.4) indicated that the novice and expert LSTM$_{NN}$ models weighted input state variables differently when making target selection predictions. The related implication is that novice and expert herders employed state information differently when making target selection decisions. Hence, we explored whether these differences could be determined by employing explainable AI techniques to identifying how the LSTM$_{NN}$ models weighted state input features.

5.4.1 Overview of explainable AI techniques and SHapley Additive exPlanation

As anticipated in Chapter 4, artificial neural networks (ANNs) have underpined many of the advances in machine learning and artificial intelligence (AI) in recent years [6, 37, 57, 155]. Despite the practical utility and effectiveness of ANNs, the large number of connection weights within ANNs, particularly Deep-ANNs, makes it difficult to directly access and understand how information, as input, propagates though an ANN to provide a given output. For this reason, ANNs are often re-

ferred to as "black-box" models. However, a desire to better understand and interpret the validity and reliability of ANNs and other black-box models, as well as the growing demand for more transparent, ethical and trustworthy AI and AA systems [163], has resulted in a renewed interest in the application and development of explainable AI techniques [93, 101, 112, 116, 147].

In short, explainable AI techniques make the internal processes of a black-box model understandable to human investigators by deriving linear explanation functions of the effects that input features have on the output states. Popular methods include LIME [126], DeepLIFT [144], and, more recently, SHapley Additive exPlanation (SHAP) [91, 92], which we employed here.

With respect to the analysis of the input-output mappings of an ANN, the SHAP algorithm pairs each input feature with a SHAP value. The *higher* the SHAP value, the *greater* the influence of the feature on the output. Importantly, SHAP values are locally accurate. Thus, to derive a measure of global feature importance one can average the importance of a feature over the test set used to assess generalised model accuracy. The result is an average SHAP value for a given input to output mapping, with this value indicting the overall significance of a given input feature for a given output prediction.

In what follows we demonstrate how SHAP can be employed to identify the task information underling the target selection decisions of human herders, including what differentiates expert from non-expert decision making performance. That is, we employed SHAP to identify what input information (features) provided to an LSTM_{NN}, that could effectively predict expert and novice target selection decisions, most influenced the ANNs prediction output (i.e. the decision made).

5.4.2 Feature rankings association analysis

We computed SHAP values[2] for each sample in the test set (predicted with the accuracy scores reported in Table 5.1) and then rank-ordered each input feature in terms of its importance in making target selection predictions as captured by its average SHAP value (see Appendix B.7 for more details on the SHAP algorithm being employed).

Before assessing what (if any) specific input features were weighted differently, we first computed the ordinal association of SHAP value rankings between the different LSTM_{NN} models using the Kendall's τ rank correlation coefficient [98]. Kendall's τ is a non-parametric statistical test of rank order association, with $\tau = 0$ corresponding to the absence of an association[3], $\tau = 1$ corresponding to perfect association (matched rankings), and $\tau = -1$ corresponding to opposite ranking orders (negative association). We computed Kendall's τ between the SHAP rankings of the full input feature set and the top 10, 5 and 3 features, ranked by SHAP, for (i) the novice and expert LSTM_{NN} models for each prediction horizon, and (ii) the long and short prediction horizons for expert and novice LSTM_{NN} models.

[2]Computed SHAP values are made available in the public repository https://osf.io/wgk8e/. These can be analysed with the code available at github.com/FabLtt/ExplainedDecisions

[3]Although $\tau = 0$ is the null-hypothesis (and one cannot draw conclusions from non-significant results), it does provide a robust and intuitive assessment of rank order independence

As can be seen from Figure 5.5, the Kendall's τ analysis revealed that there was very little association between the novice and expert SHAP rankings for both short and long prediction horizons, with average $\tau < 0.4$, ($p > 0.26$) for all feature sets, further suggesting that novice and expert herders employed different state information when making targets selection decisions. With regard to the association between the short and long prediction horizons for novice and expert models, there was also very little association between SHAP rankings for the full, top 10 and top 5 ranked feature sets for each level of expertise (average $\tau < 0.4$, $p > 0.3$). It is important to note that there was slightly greater association (yet still not significant) between the top 3 ranked features set (average $\tau < 0.5$, $p > 0.3$) for each level of expertise. As detailed below, this was due to the relative herder-to-target distance features often ranked within the top 3 features across target ID, implying that this information played a key role in target selection decisions irrespective of the prediction horizon (see Appendix D.2 for a detailed summary of Kendall's τ values for each target prediction class and model comparison).

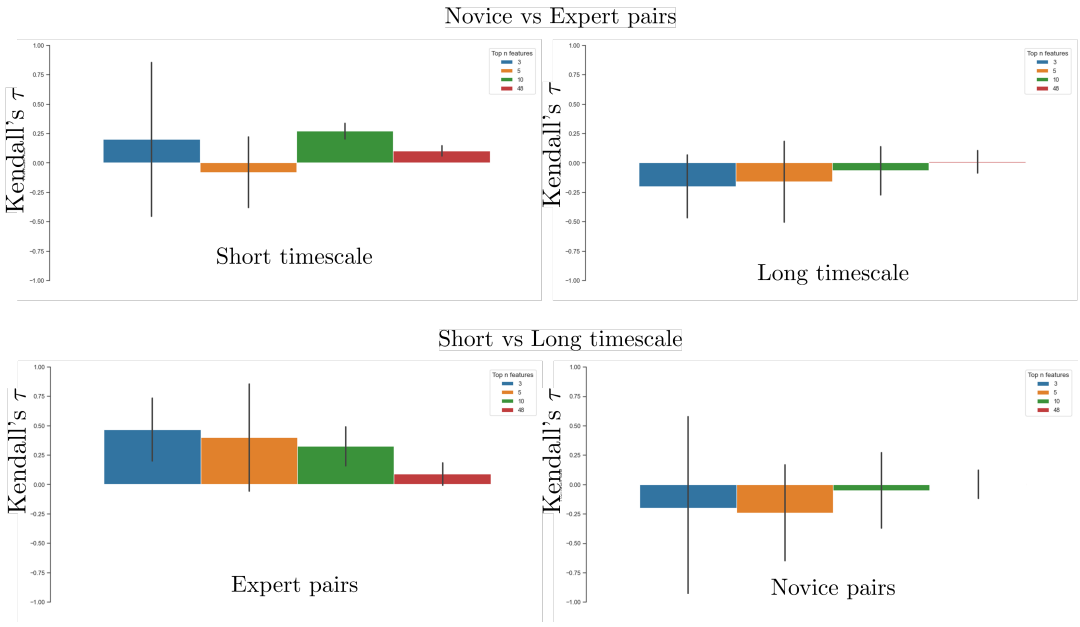


Figure 5.5: Kendall's tau values for subgroups of top ranked input features averaged over labels. The top panel shows the Kendall's τ comparison between novice and expert SHAP rankings for short (top left) and long (top right) prediction horizon. The bottom panel shows the Kendall's τ comparison between the short and long prediction horizons for expert (bottom left) and expert (bottom right) models.

5.4.3 SHAP analysis

To highlight what specific input features played a major role in the target selection predictions for novice and expert herders, Figure 5.6(a) illustrates the top ten most prevalent features for non-zero prediction outputs (i.e., ID = 1 to 4), with the different input features (e.g., distance from goal, ve-

locity, etc.) defined with regard to the current herder in question ("1st-herder"), the partner herder ("2nd-herder"), the "will-be-corralled target" at τ_{hor} (i.e., correct prediction output) and the "other" targets that were not predicted to be corralled at τ_{hor}.

Consistently with the Kendall's τ analysis for the top 3 feature sets, the relative distance between herders and targets played a key role in the target selection predictions; these input features nearly always had the highest SHAP values for each target ID, largely independent from expertise and prediction horizon (filled circle marks in Figure 5.6(a)).

Interestingly, the relative distance between the two herders was only ranked within the top 10 features for expert target selection predictions (empty diamond marks in Figure 5.6(a)). This implies that expert herders were more attuned to their co-herders location within the herding environment than novice herders and modulated their target selection decisions accordingly. However, both novice and expert pairs were attuned to the partner herder's distance from the targets , suggesting that novices were, at least, aware of partner's whereabouts – either to seek partnership and workload sharing, or to not be hindered. Expert target selection predictions were also influenced by the herder's current distance from the containment area, whereas novice target selection predictions were not (square marks in Figure 5.6(a)). Expert predictions at both short and long time horizons were also more reliant on target direction of motion information compared to novice predictions, whereas novices target selection predictions were more reliant on target and herder acceleration compared to experts, suggesting a deeper focus on their own actions than partner's.

Finally, similar SHAP results were observed with regard to no-target (ID = 0) predictions (Figure 5.6(b)), with novice target selection predictions more reliant on acceleration (and velocity) information compared to expert predictions. Refer Appendix D.3 for a detailed summary of SHAP feature values for each model, prediction horizon and target ID.

5.5 Summary

Results in this Chapter revealed that independently of expertise level, both immediate and delayed target selection decision of human herders performing a fast paced joint task could be accurately modelled by training LSTM-layered ANNs (Section 5.1).

A key finding was that the trained LSTM_{NN} models were expertise specific, in that, when the expertise level of the training and test data mismatched, prediction performance dropped to near chance level (Section 5.3). Consistent with action decisions during skilful action being a function of an actor's level of situational awareness [24, 96] and trained attunement to the information that best specifies what action possibilities will ensure task completion [68, 158, 176], this was a result of the expert and novice LSTM_{NN} models weighting input features differently.

The differences between expert and novice pairs of herder was further explored using the explainable AI technique SHAP, with the average SHAP feature rankings revealing that although expert and novice pairs both employed similar relative target distance information, expert decisions

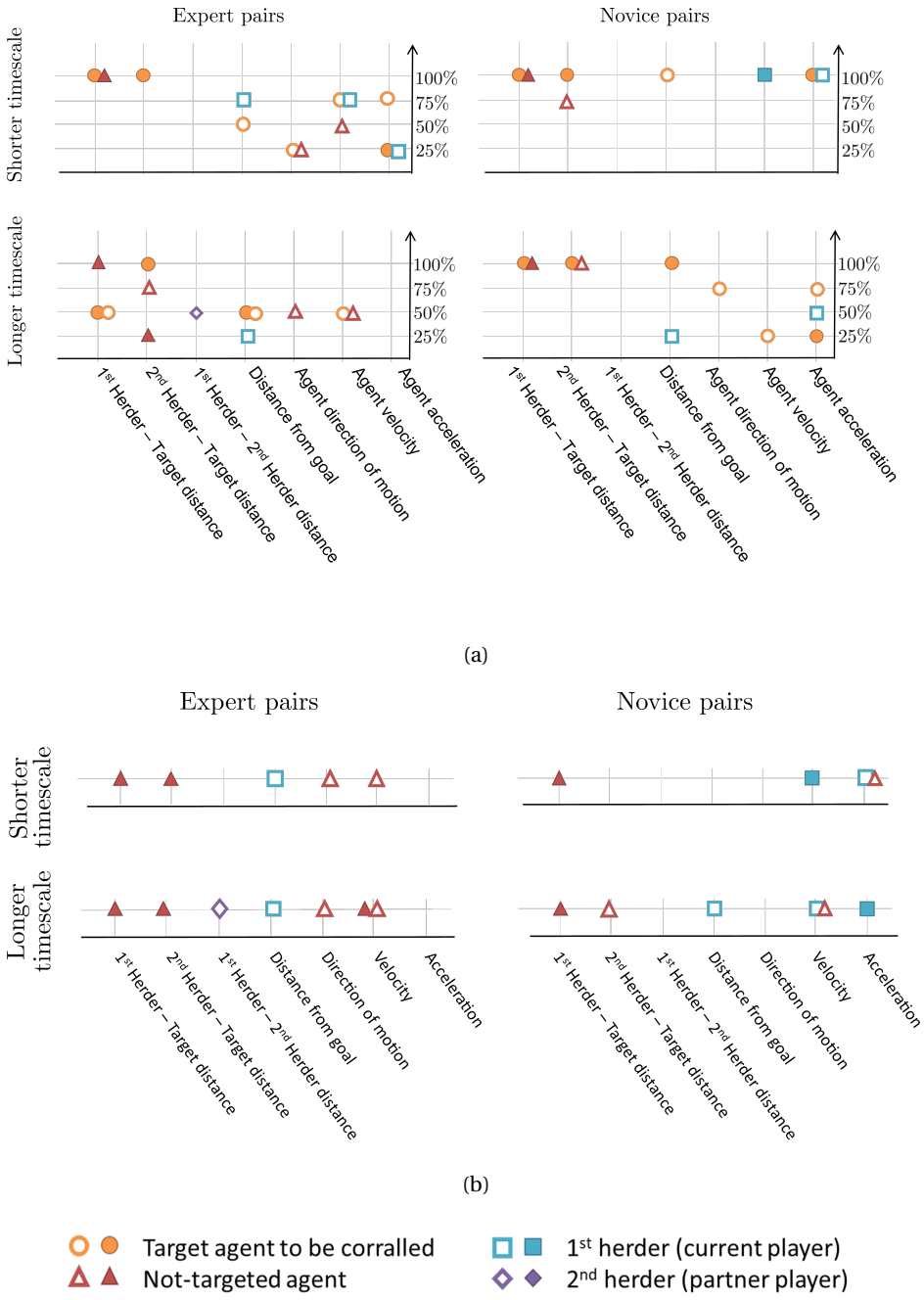


Figure 5.6: SHAP results for target prediction (a) ID = 1 to 4 and (b) ID = 0 (no target), as a function of prediction horizon and expertise. Features type is listed on the x-axis. In (a) the y-axis represents the proportion of targets $N_T = 4$ a state variable was found on average to be ranked within the top-ten features, such that "100%" represents a feature that was a top-ten feature with respect to all targets. In (b) only features ranked with an average ranking in the top-ten features ID = 0 are displayed. Filled marks indicate state variables ranked on average in the top five features (i.e., from 1st to 5th place) while empty marks indicate state variables ranked on average between 6th to 10th place.

were more influenced by direction of motion information compared to novices. In contrast, novice decisions were more influence by target velocity and accelerations rather than by partner's actions. This is consistent with players being novices not only to the herding task itself but to the partner player (i.e., to the partner's level of expertise or decision making process). Together with finding that experts transitioned between targets less often than novices (see Section 5.1), this suggests that experts were more attuned to information that better specified the prospective state of the herding system, including what targets afforded corralling by their co-herder, and only switched targets when such target transitions were critical to future task realisation. This possibility is, of course, consistent with expertise being commensurate with knowing how to perform tasks more efficiently and prospectively compared to pairs of novices [40, 104, 157, 176]

In the next Chapter we will explore the degree to which the current SML modelling and explainable AI approach can be adapted to herding tasks that involve teams of herders (i.e, 3 human herders) and much slower decision timescales (e.g., actions decisions taken over tens of seconds). Of particular interest will be whether the proposed approach could be be employed to predict and understand human decision-making events across a variable prediction horizon (i.e., non-fixed time scale predictions).

CHAPTER 6

MODELLING FUTURE ACTION DECISIONS AT VARIABLE PREDICTION HORIZONS

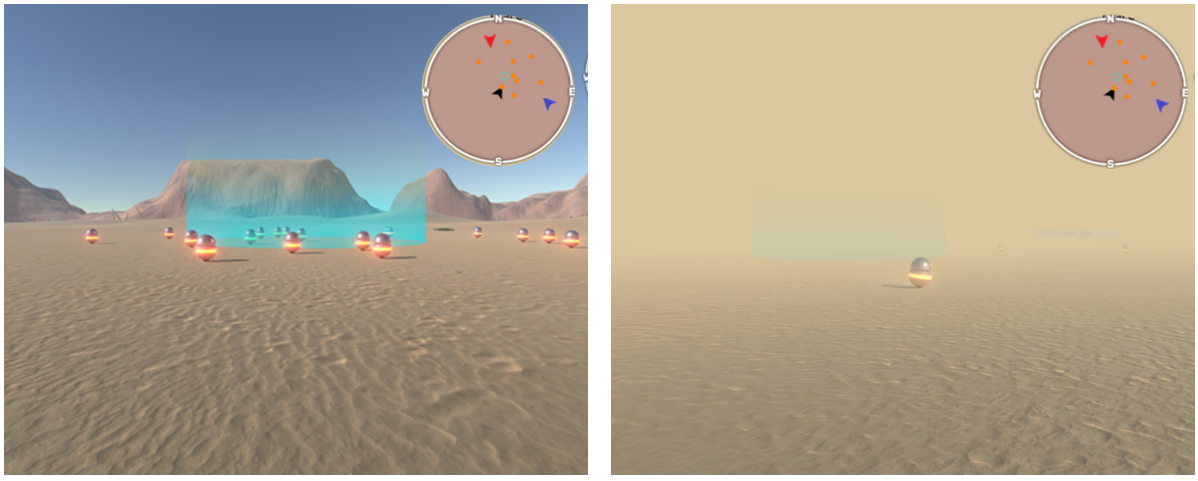
Chapters 4-5 demonstrated how contemporary SML and explainable AI methods can be employed to both model and understand the decision making behaviour of human actors within a multi-agent task setting. In this Chapter, we apply the same modelling approach to capture the decision-making behaviour of human actors playing a 3-player online herding game called "Desert Herding" (Section 6.1). Of particular interest, was whether the modelling approach could be employed to predict and understand the target selection strategies of human herders at both fixed (Section 6.2) and variable (Section 6.3) prediction horizons. As in Chapter 5, we then leveraged SHAP to identify the key informational variables (features) underlying the players' target selection decisions (Section 6.5).

6.1 Multi-player desert herding game

As noted previously in this Thesis, the action decisions made by human actors during multi-agent behaviours are typically highly tuned, context specific, responses to the unfolding dynamics of a task [24, 68, 157] and for fast-paced perceptual-motor tasks, like the tabletop herding task investigated in Chapters 4-5, often occur <500 ms prior to action onset [169]. Thus, the results reported in Chapter 5 not only demonstrate how the modelling approach employed here can effectively predict and understand human decision making during multi-agent behaviour, but can potentially do so at timescales well in advance of a human actor's conscious (or unconscious) intent. To explore this possibility further we examined whether we could predict the target selection decisions of human herders that had completed the multiplayer online herding game called "Desert Herding" within the study first presented by Prants *et al.* in [123].

For this herding game three human herders (players) were required to work together to corral herds of targets randomly distributed around a larger 500 m x 500 m desert landscape into a containment area positioned at the centre of the desert game area. Players downloaded the multiplayer "Desert Herding" game, developed with Unity-3D game engine (Unity Technologies LST, CA), on their home computer and completed the task remotely. The game server, run on an EC2 instance of Amazon Web Services, recorded all game state data at 90-Hz.

As illustrated in Figure 6.1, each player had access to a first-person perspective of the game field and an Heads-Up Display (HUD) located at the top-right corner of their screen. Each player controlled a humanoid avatar using standard keyboard and mouse PC/MAC game controls (i.e., the "W-A-S-D" keys to control forward, left, backward, right directions respectively, and the mouse for rotation). The avatar for each player was a different colour (i.e, red, blue, white) so that players could identify each other. The targets were small autonomous robots that were repelled away from player avatars if the player avatars got within a 10 m radius of the robot target. When repelled away from a player avatar, an autonomous robot target would enter a "running" mode and move in the opposite direction of the player's approach vector. When not being corralled (chased) by a player avatar, the robot targets exhibited a small degree of Brownian motion around their current resting position. Target status was visible to players via a visual light positioned around the centre of a target's chassis. Targets emitted an orange light when at rest, a red light when being corralled (repelled), and a blue light when successfully corralled into the containment area. As in the tabletop herding task, the targets were not programmed to flock together; see [123] for more details on the "Desert Herding" and the data collection process.



(a) (b)

Figure 6.1: Examples of the multiplayer desert herding environment, first introduced in [123], under full (a) and partial (b) visibility conditions. The HUD display is shown on the top-right corner of the participant display.

The number of targets (herd size) could range from 1 to 100 targets, however, here we consider

data from teams required to corral 9 targets. In addition to herd size, player visibility could be manipulated via the inclusion of environmental fog. Here we model the target selection decisions of players completing the task under either low visibility conditions (i.e., a 15 m visibility horizon; see Figure 6.1(b)) or full (high) visibility conditions (i.e., >150m visibility horizon). As noted above and illustrated in Figure 6.1, players also had access to a HUD positioned at the top-right corner of their screen. For the data being modelled here, the HUD displayed the real time position of all three players and all target agents, overlaid on a game space map. The location of the containment area (blue coloured circle in Figure 6.1) was also displayed on the HUD and, for the game data modelled here, was always fixed at the centre of the herding space. Finally, players could verbally communicate with each other; see [123] for a complete description of the task and possible task manipulations.

6.1.1 Experimental data

We extracted and modelled the target selection decisions of players who participated in the "Desert Herding" study conducted by Prants *et al.* [123]. For this study, ten three-person teams completed four separate 1-hour game sessions. Each session included 16 trails, in which herder size (9 vs 18 targets), visibility (low visibility [fog] vs high visibility [no-fog]) and the presence (or not) of the HUD was manipulated. Here we chose to model the target selection decisions for both the high and low visibility conditions, when the number of targets, $N_T = 9$ (initial modelling revealed similar results for trials containing $N_T = 18$ targets) and when the HUD was present. Note that for the data explored here, the HUD would always show the position of all $N_H = 3$ herder and $N_T = 9$ target agents w.r.t. cardinal points as in Figure 6.1.

A trial was considered successful if all $N_T = 9$ targets were corralled and remained in the containment region for 5 consecutive seconds. Otherwise, the trial would terminate if participants spent longer than 5 minutes performing the task. We chose to focus on Session 4, as by this final session all teams had reached a high level of performance and were able to succeed on nearly all trials. A total of 20 successful trials were obtained resulting in a 20 trial x 3 player data set.

To employ the supervised machine learning (SML) algorithm presented in Chapters 4-5, we first needed to label each player's times series data with the ID of the target (or no target) corralled at each time step (i.e, the correct prediction output). Differently from the coding method employed in Chapter 4, where the ground truth about the ID of the agent targeted at each time step was manually classified by a paid research assistant, here the ID of the target that a player was currently corralling was automatically determined from the recorded data. At each time step, the target a player was corralling was identified as the target agent (already numbered 1 to 9 in the data recordings) which (i) had a distance from the herder smaller than 10 m in game space, (ii) was in "running" mode and (iii) was the closest to the player. Using this automatic method, the prediction output was set to ID = 0 if a player was not currently engaged with (corralling) a target or ID = $1, 2, \ldots, N_T = 9$, otherwise.

6.2 Predicting decision making at fixed prediction horizons

To investigate the decision making processes of team members applying the same fixed timescale (prediction horizon) approach presented in Chapters 4-5 to the newly acquired experimental data, we again selected as input to a Long-Short-Term-Memory neural network (LSTM$_{NN}$), N_{train} time series of N_{sv} state variables as described in Appendix B.4, of length $t_f - t_i = T_{seq}$, with the output label corresponding to the ID of the target agent being corralled at $t_{f+\tau_{hor}}$, with constant $\tau_{hor} > 0$. Further details on the learning algorithm are discussed in Appendix B.5.

Because the herding environment considered here is much larger and more varied (e.g., presence of rises and troughs) than the tabletop herding environment explored in Chapters 4-5, the time interval between player target selection decisions was much larger. Here we chose a sampling time $dt = 0.2$ s, a sequence length $T_{seq} = 5$ s and prediction horizons of $\tau_{hor} = 25$ or 65 steps, which corresponded to making target selecting predictions at 5 s and 13 s, respectively. The choice of $T_{seq} = 5$ s was motivated by computing the players inter-target movement times (i.e, target switching time), as introduced in Chapter 5.1.1. Again, these were computed as the time it took a player to move outside the repulsive radius of a target and into the repulsive radius of the following one. As can be seen from an inspection of Figure 6.2, the average inter-target movement time was typically greater than 5 s. Note that we also validated the modelling approach for $T_{seq} = 2$ s of system evolution (see Appendix E) with results comparable to the $T_{seq} = 5$ s case discussed in this Chapter.

We trained separate LSTM$_{NN}$ models for each combination of timescale and visibility condition (i.e., 4 models in total) from $N_{train} = 10000$ training samples, with the model accuracy tested against 100 sets of $N_{test} = 2000$ test samples each. As reported in Table 6.1, each LSTM$_{NN}$ could successfully predict which target would be corralling at τ_{hor} with an average accuracy exceeding 97% for all conditions and prediction horizons (τ_{hor}). The F1 score and model precision and recall were also above 97% for both visibility conditions and prediction horizons. Kruskal-Wallis tests revealed that the performance of longer (13 s) prediction horizon and Partial visibility models were significantly between than the shorter and full visibility models, respectively (average $\chi^2 = 9.3$, $p < 0.04$). However, the magnitude of the differences was rather trivial (< 0.15% on average).

6.3 Predicting decision making at variable prediction horizons

Figure 6.2 shows that the decision-making timescales were distributed between 0.07 s and 11 s; this suggested that it was not enough to be able to just predict "which target would be corralled at a particular time in the future", but that in order to fully understand the decision making behaviour of players we also needed to model targeting switching events; that is, the variable timescale decision events that corresponded to when participants decide to change the target agent they engaged; i.e., what we will refer to as *target switching decisions*.

Recall from the analysis reported in Chapter 5, that for a fixed τ_{hor} in the future, a herder could either keep corralling the same target in the sequence preceding the decision or it could switch to

	Accuracy	Precision	Recall	F1 score
Shorter timescale				
Full visibility	97.76 ± 0.27	97.71 ± 0.29	97.89 ± 0.27	97.78 ± 0.28
Partial visibility	97.9 ± 0.31	97.92 ± 0.31	97.89 ± 0.3	97.9 ± 0.31
Longer timescale				
Full visibility	97.9 ± 0.31	97.94 ± 0.31	97.98 ± 0.31	97.95 ± 0.31
Partial visibility	98.01 ± 0.32	98.02 ± 0.33	98.01 ± 0.32	98.01 ± 0.32

Table 6.1: Average performance [%] of the multi-label predictor trained on time-series of length T_{seq} = 5 s and tested on 100 sets of N_{test}= 2000 samples for shorter (T_{hor} = 5 s) and longer (T_{hor} = 13 s) prediction horizons. Although comparable, prediction performance was statistically different across visibility and timescale (Kruskal-Wallis test, average $\chi^2 = 9.3$, $p < 0.04$)

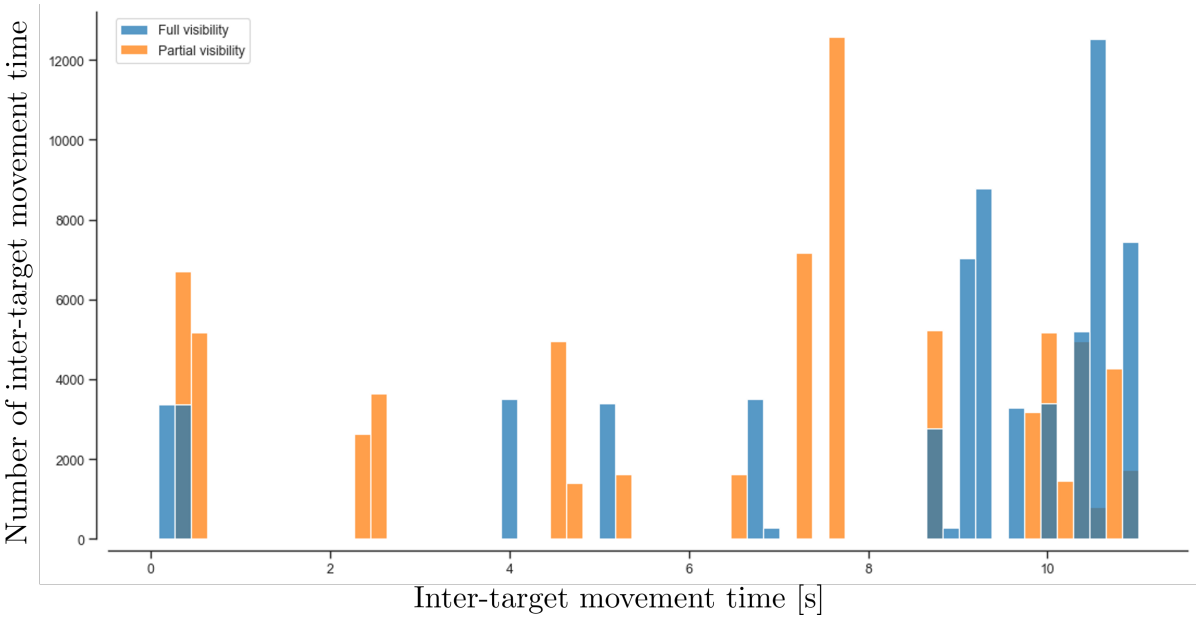


Figure 6.2: Distribution of inter-target movement times [s] averaged across player for the full (blue) and low (orange) visibility conditions (i.e, no-fog vs. fog) for the 20 successful trials employed for model training and testing, with an average inter-target movement time of 8.12 s (74.4% of the total inter-target movement times > 7 s) for full visibility trials and 6.57 s (62.6% of the total inter-target movement times > 7 s) for low visibility trials.

a different one. Furthermore, that the latter behaviour was observed more often for novice herders, suggesting that expert herders enacted a more decisive decision process than novices, switching less often (see Section 5.3). Assuming the players were sufficiently "expert" by Session 4 and that switching events were of more importance for understanding the target selection decisions of players in the desert herding game, we trained LSTM$_{NN}$s to predict the ID of the next engaged target, independently from how far in the future this target switching event was observed.

More specifically, for each sequence of $t_f - t_i = 25$ time steps of the system evolution, the output label was the ID of the next target corralled that was different from the current target engaged with at t_f, with one limiting restriction, $\tau_{hor,max}$, which was the maximum step in the future explored. Thus, if in $\tau_{hor,max}$ consecutive time steps in the future a player engaged with the same target, then the ID label was selected as that same ID (i.e, the ID of the target agent engaged at $t_{f+\tau_{hor,max}}$).

For T_{seq} = 5 s, we considered a minimum and maximum prediction horizon $t_{\tau_{hor,min}} = dt$ and $t_{\tau_{hor,max}}$ = 20 s (or 100 times steps in the future), resulting in 97.4% switching samples for the full-visibility condition and 97.3% for the partial visibility condition. Note that switching samples represent the cases that either a participant started engaging a target (ID = 0 to ID $\neq$ 0), or a participant stopped engaging a target (ID $\neq$ 0 to ID = 0) or that it switched between engaged targets.

Prediction performance, averaged over 100 sets of N_{test} = 2000 samples and reported in Table 6.2, confirmed that both the partial and full visibility LSTM$_{NN}$s models were able to predict the next corralled target with an accuracy above 91%. Although small, a Kruskal Wallis statistical test [79] on data reported in Table 6.2 indicated that the difference in the accuracy of the full and partially visibility models was statistically significant ($\chi^2 = 80.75$, $p < 0.0001$), with the full visibility model slight outperforming the partial visibility model. Although small, the performance difference between the full and partial visibility models further highlight the importance of understanding how input features (i.e., state variables) are weighted by each model. For example, this difference could reflect the difference in information available to players between the two conditions, with fog introducing an environmental perturbation and, consequently, uncertainty in target switching decision. Before leveraging the explainable AI tool SHAP to gain insights on how input information was weighted by the LSTM$_{NN}$s models, in the next Section we first explored whether the variable prediction horizon models (i.e., the models predicting target switching decisions) could also accurately predict target selection decision at fixed prediction horizons (and vise versa).

	Accuracy	Precision	Recall	F1 score
Full visibility	92.15 ± 0.54	92.39 ± 0.54	93.45 ± 0.51	92.87 ± 0.51
Partial visibility	91.25 ± 0.61	91.35 ± 0.63	92.14 ± 0.59	91.66 ± 0.6

Table 6.2: Average performance [%] of the multi-label predictor trained on time-series of length T_{seq} = 5 s, and variable prediction horizon, when tested on 100 sets of N_{test}= 2000 samples.

6.4 Robustness of using variable prediction horizons

In addition to testing the overall accuracy of the variable prediction horizon models, it was also important to investigate if there was a functional relationship between prediction horizon length and accuracy. To investigate this we examined if the fixed prediction horizon models reported in Section 6.2 (i.e., models where $\tau_{hor} = 25$ or 65 steps) and the variable prediction horizon models in the previous Section 6.3 could accurately predict the IDs of the to-be-corralled targets across a range of different fixed prediction horizons τ_{hor} in the future. That is, we aggregated all test samples by prediction horizon and tested the different fixed and variable prediction horizon LSTM$_{NN}$s models on these "aggregated" test sets. As can be seen from an inspection of Figure 6.3, the models trained to predict the decision at the fixed prediction horizons of $\tau_{hor} = 25$ or 65 steps (dotted and dashed lines) only retained peak performance when tested on samples selected at the same timescale (prediction horizon) of the trained model. In contrast, the models trained to predict the next corralled target, at a variable prediction horizon, were able to correctly predict with an accuracy above 90% the target that would be targeted next, independently from the timescale and experimental condition (solid lines) when $t_{f+\tau_{hor}} > 2$ s. That is, the accuracy of the variable prediction models was invariant over the timescale $t_{\tau_{hor,min}} = dt$ and $t_{\tau_{hor,max}} = 20$ s.

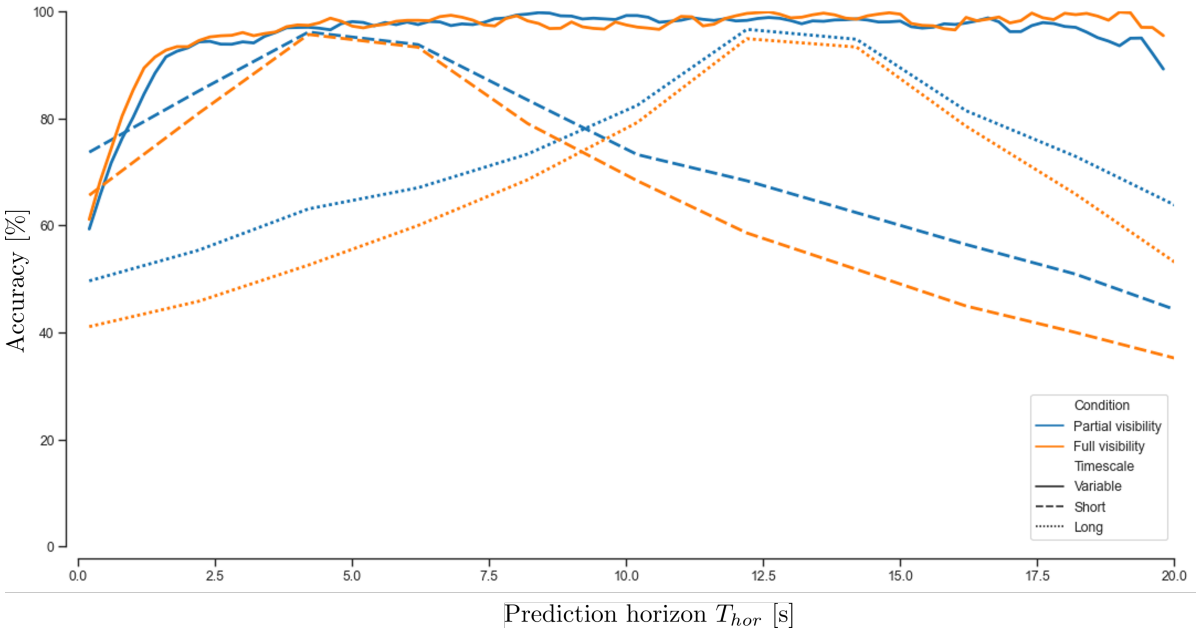


Figure 6.3: Prediction accuracy of ANNs modelling team members decision making during last session of experiments by [123] when faced with full and partial visibility, tested on samples aggregated by prediction horizon τ_{hor}.

6.5 Explanation of target switching decisions

As in Chapter 5, we employed SHapley Additive exPlanation (SHAP,[92]) to better understand how the different input features to the variable prediction horizon LSTM_{NN} models influenced the prediction outputs. Recall that given a trained "black-box" model, the SHAP algorithm pairs each input feature with a SHAP value representing the weight that a feature has on the model output; the higher the SHAP value, the greater the influence that feature has on the model output. Global feature importance can therefore be assessed by calculating the averaged SHAP value for a given feature over the test set.

Here, for each LSTM_{NN} trained to predict target switching decisions at a variable prediction horizon (Table 6.2), we computed[1] the SHAP values for $N_{test} = 6000$ samples and rank-ordered each feature in terms of its average importance (i.e., average SHAP value). Given these SHAP rankings, we first employed Kendall's τ rank order test [98] to examine whether there were differences in the features rankings for the partial and full visibility conditions. Recall that $\tau = 1\,(-1)$ corresponds to perfect (negative) association and $\tau = 0$, the null-hypothesis, corresponds to the absence of association.

As illustrated in Figure 6.4, the Kendall's τ analysis revealed that the rankings of the entire input feature set and the top 75, 50 and 20 features presented little association between the input-out mappings for the partial and full visibility models (average $\tau \leq 0.04$, $p \leq 0.61$). There was slightly greater association between the top 10, 5 and 3 feature sets, however, these associations were also not significant (average $\tau \geq 0.12$, $p \geq 0.46$). Interestingly, the closest similarity observed between the top 3 features (average $\tau = 0.4$) was due to the relative herder-to-target distance features always ranked as top features across target ID. This is consistent with the results reported in Section 5.4.2, Chapter 5, in which these same features played a primary role in the target selections decisions of expert and novice herders completing the tabletop herding task.

Figure 6.5 provides visual illustrations of feature importance for the partial and full visibility models, with the top twenty most prevalent features for non-zero prediction outputs (i.e., ID = 1 to 9). The importance of relative herder-to-target distance can also be easily discerned from an inspection of this Figure, with how far (or close) players were from target agents always within the top twenty most important features for both the partial and full visibility models. That is, players tending to select targets that were closer to themselves than to other players. Interestingly, independent of visibility condition, a player's own velocity and target acceleration were also key to target predictions. More specifically, a player/herder's (filled blue square mark) velocity was key to target selection decisions for 100% of all target ID predictions, with the to-be-corralled targets acceleration (filled orange circle marks) important for over 90% of prediction outcomes. The importance of these features was likely due to (i) players selecting new targets just after they had corralled a target into the containment area, with the to-be-corralled target typically stationary (i.e., had zero or close

[1] More details on the SHAP algorithm being employed can be found in Appendix B.7

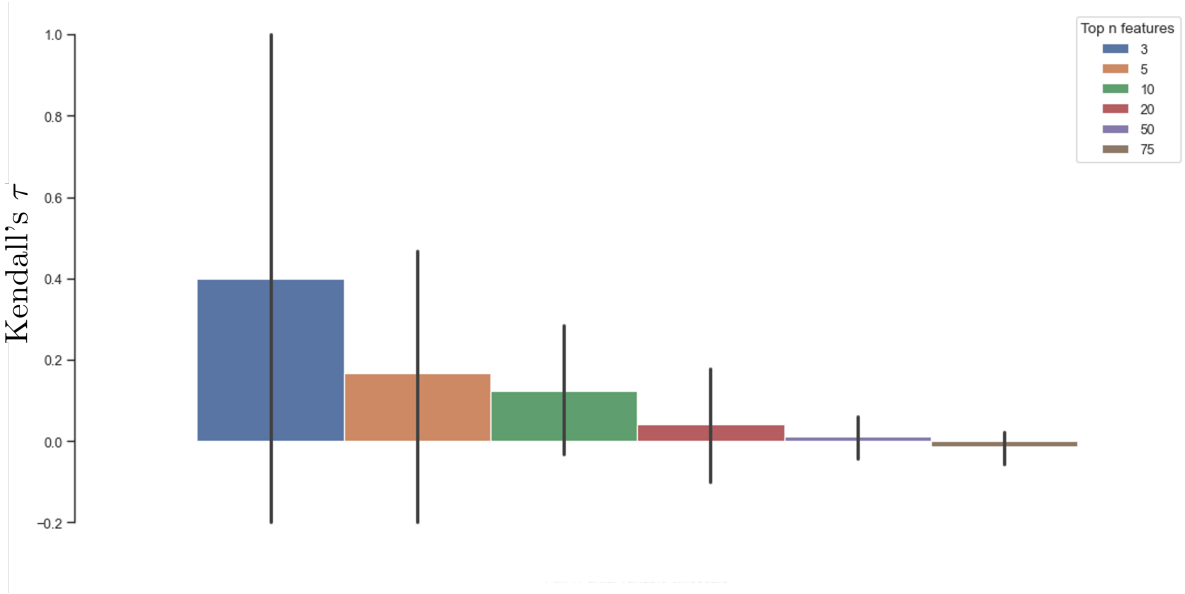


Figure 6.4: Kendall's tau values for subgroups of top ranked input features, averaged over labels. A summary of Kendall's τ values for each target prediction class is reported in Appendix E.

to zero velocity) and (ii) players choosing targets that were not increasing or decrease speed, as the later only occurred if a target was already been corralled by another player.

Another interesting finding that can be discerned from an inspection of Figure 6.5, was that the relative distance between players (i.e., co-herders) had little impact on target switching decisions. In fact, inter-herder distance almost never appeared in top twenty features. This may have been due to the large game area and that players could freely move anywhere in the game area (in contrast, for example, to the tabletop herding task, where players were restricted to one side of the table). In other words, how close or far a player was from another player was of little importance. This is in contrast to the significant importance of the relative distances between herders and targets, again, indicating that how close a target was to a player (or other players) was key, not the relative distance between players/herders.

With regard to what differentiated the target prediction of the partial and full visibility models, the distance of the targets from the containment goal area was weighted as more important (on average) for partial visibility model predictions compared to the full visibility model predictions. One possible reason why goal distance information was weighted as more important for the low visibility model, was that in the presence of environmental impairment caused by the fog, players could often not see any target within their first person field of view and, hence, would have relied more heavily on the HUD to select targets, naturally choosing targets that were furthest from the containment area. In contrast, players in the full visibility condition may simply have chosen targets that were in their current (first person) field of view and, thus, may not have always selected the target furthest from the containment area (just the target furthest from their current location in

their current field of view).

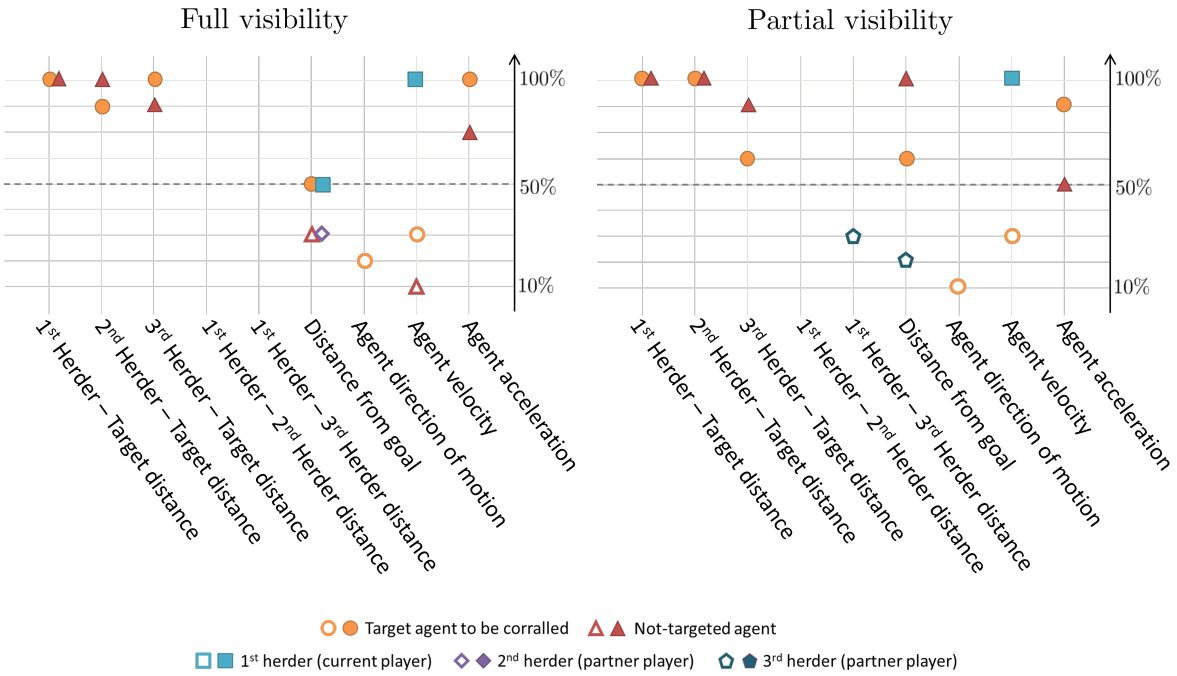


Figure 6.5: Explanation of the Multi-label predictor of "agent to be targeted" prediction outputs, for full and partial visibility conditions. Feature type is listed on the x-axis, with the y-axis representing the portion of $N_T = 9$ targets for which a state variable was found on average to be a top twenty global feature, such that "100%" represents a feature that was a top-twenty feature with respect to all targets. Filled marks indicate state variables used to predict a portion of targets equal or above 50% while empty marks indicate state variables used to predict smaller portions.

6.6 Summary

In this Chapter, we extended our SML and explainable AI pipeline to model and understand the target selection decisions of human actors playing a multiplayer (3-person) online Desert Herding game. Due to the large game area of the Desert Herding game, we first extended the modelling approach reported in Chapters 4-5 by demonstrating how SML-trained LSTM_{NN} models could successfully predict the target selection decisions of human players (herders) at longer fixed prediction horizons of 5 and 13 seconds (Section 6.2). Motivated by an analysis of inter-target switching movement times, which revealed how the interval between a player's target switch decisions ranged from 0.07 to 11 seconds, we then demonstrated how the same modelled approach could also be extended to predict target switching decisions at variable prediction horizons. That is, LSTM_{NN} models could be trained to predict the next target a player would corral independent of how far in the future this target switching decision occurred (Section 6.3). Moreover, LSTM_{NN} models trained to predict target switching decisions at a variable prediction horizon were also found to produce more accurate test predictions at fixed prediction horizons compared to fixed interval prediction models, even when tested on the prediction horizons aggregated by a characteristic fixed timescale (Section 6.4). Finally, we employed SHAP to explore what task information players might have employed when making their target selection decisions. This analysis revealed that the relative distance between targets and herders, and target and herder motion played a key role in players' target selection decisions independent of visibility, but that players in the partial-visibility condition may have more heavily relied on the containment region distance information contained within the HUD compared to when completing the task in the full visibility condition (Section 6.5).

It is worth noting that, with regards to the experiments employed in this Chapter, Prants *et al.* not only explored team behaviour when access to task-relevant information varied (e.g., in presence or absence of a HUD), but allowed team members to communicate via the use of a teleconferencing system (Zoom; San Jose, California) [123]. This was not the case in the experiments conducted by Rigoli *et al.* [129] employed in Chapter 4-5. As a consequence, the SHAP results discussed in this Chapter rest on the assumption that explicit information on agents position (either herder or targets) or players' intention could have been verbally shared. Nevertheless, we here considered only task relevant state variables as a, preliminary, inspection of experiments' audio recordings highlighted that (i) the majority of communication did not consist of words in the traditional sense (e.g., grunts, nods, agreement sounds) and that (ii) their meaning could not be trivially fed to ANN. In the next Chapter, we will discuss the importance of expanding the proposed modelling and explainable AI approach with non-trivial information (e.g., verbal, haptic, visual information) along with possible applications and open paths.

CHAPTER 7

CONCLUSIONS

In this Thesis, the emergence of coordinated behaviour during skilful multi-agent activity was studied using a combination of mathematical, supervised machine learning (SML) and explainable AI tools. More specifically, the work presented here addressed the following research questions.

(i) Can local control rules and dynamic decision strategies replicate the global coordinated behaviour observed in humans?

(ii) Can effective human decision making in complex multi-agent tasks be modelled using SML and Long-Short-Term-Memory Artificial Neural Networks?

(iii) Can artificial agents equipped with human-like decision-making processes complete multi-agent tasks effectively?

(iv) Can the explainable AI technique SHAP be employed to uncover the task-specific information that underlies effective human decision making during multi-agent activity?

7.1 Summary of the main results

After introducing the Thesis motivation and open research questions in Chapter 1, a detailed literature review of the model solutions for the multi-agent herding task was given in Chapter 2. We adopted the herding problem as a representative case of multi-agent activity as it is defined by the interaction of two sets of autonomous agents; herders and targets. The first (herders) directly controllable and the second (targets) indirectly influenced by the first. Popular state of the art solutions for the design of artificial herder agents (AA herders) were discussed, with an emphasis on the fact

that little attempt has previously been made to design AA herders able to corral non-flocking targets. Moreover, that the majority of the previously proposed solutions adopt simple heuristic rules when distributing the herding workload (i.e, target selection decisions) among multiple herders.

The most general scenario of a small set of AA herders faced with a larger set of non-flocking targets was tackled in Chapter 3. More specifically, we investigated herding strategies composed of local control rules, regulating herder agents attraction to a point behind each targeted agent, and different target selection strategies (i.e., different herder decision-making policies). Numerical Matlab experiments showed that the proposed control law, as well as the various target selection polices, could not only complete a general herding task successfully, but were (i) not dependent on target flocking behaviour and (ii) robust to variations in herd size and the magnitude of repulsive action exerted by herders on targets. These findings were also validated in ROS simulations and using real robots, further demonstrating how the proposed control law and dynamic selection strategies could be adopted in real-world robotic applications with low computational requirements.

Human decision making during fast-paced joint-action behaviour was modelled in Chapter 4. Here, we demonstrated how Long-Short-Term-Memory artificial neural networks (LSTM-ANNs) could be trained using SML to accurately predict the target selection decisions of human actors playing a two-person tabletop herding game. To validate whether these models could be used to control the decision behaviour of artificial agents (AA), we embedded these models into the control architecture of AA herders and tested their efficacy. As expected, numerical simulations revealed that the AA herders, driven by local control rules and human-inspired decision making LSTM models, could successfully complete complex multi-agent herding tasks, even when faced with operating conditions not considered during the training phase.

Motivated by the high prediction accuracy of the LSTM models of human decision making in Chapter 4, in Chapter 5 we demonstrated how the proposed SML modelling approach could be used to predict human decision making behaviour before an human actor's behavioural intent. More specifically, we demonstrated how SML-trained LSTM models could predict the target selection decisions of human actors preforming a 2-person herding task at both short (immediate) and long (future) prediction horizons. The results, reported in Chapter 5, also demonstrated how the proposed modelling approach could effectively differentiate expert from non-expert (novice) decision making performance; that is, that the developed LSTM models of human target selection decisions were expertise specific. Given the latter result, we then exploited the explainable AI tool SHAP [92] to uncover what state information novice and expert human actors employed when making target selection decisions. The results revealed key differences in the task information employed by experts and novice human herders, providing initial evidence that explainable-AI techniques can be employed to better identify and understand the information that supports human decision making.

In Chapter 6, we further validated the proposed SML and explainable AI approach by modelling and explicating the target switching decisions of human actors playing a 3-person online herding game. Of particular importance, we demonstrated how LSTM models could be trained to predict the

decision making behaviour of human actors at variable timescales and that such models could more accurately predict decision events at fixed time-scales than models trained at specific timescales.

In summary, the findings presented in this Thesis highlight the utility of dynamical and SML methods for modelling the coordinated action and decision making behaviour of humans and AA completing multi-agent tasks. In the next Section, we discuss possible applications and future research paths with regard to modelling of human-behaviour, including how the results observed for the multi-agent herding problem could be expanded to more realistic scenarios. We then discuss the potential importance of explainable AI tools for understanding human decision making.

7.2 Possible applications and future work

The ability of AAs to reciprocally respond to and predict human behaviour is key to effective human-machine interaction [19, 43, 122]. This is true whether the interactions are pre-orchestrated, such as in factory assembly lines, or during tasks that involve dynamic improvisation, as in team sports [59, 103]. Here we demonstrated how this can potentially be achieved using simply human inspired control laws and SML trained ANN decision making models. Of course, future research is required to further validate this possibility.

One possible avenue of the future research would be to investigate whether similar control laws and decision models could be employed to develop AAs and robots capable of performing social tasks that entail more explicit coordination or competition, such as when rescuing and evacuating a group of individuals from dangers, or during mob containment or crowd control events [94, 105, 153]. Although we demonstrated that our local control laws and selection strategies were robust to large variations in operating conditions (Chapters 3-4), future research should also explore if the human-inspired models are equally robust when interacting with other human co-actors and not simply other artificial agents. In other words, more directly determine the robustness of the AAs techniques detailed here during real-time human-machine interaction. Human-AA studies could also explore whether the human-decision making models developed here could be employed for the predictive or anticipatory correction of human action decisions during task training and real-time engagement [129]. Furthermore, as in [81, 109], Turing Test methodologies could be employed to examine whether the proposed models and modelling approach results in AA behaviours that are indistinguishable from real human behaviour.

With regard to the SML-ANN decision making models explored here, future research could also validate if the generalised robustness of human-inspired SML trained decision making models is a reflection of the generalised flexibility of human decision making and perceptual-motor performance [54, 152, 168]. Previous research has demonstrated how human herders can easily adapt to the addition of new targets [106, 129] and herding environments. Therefore, it seems safe to assume that human herders would have also adapted to the addition of a new, third herder (as explored in Chapters 3 and 4), with this latent flexibility being transferred to, and entailed within, resultant

ANN models. Thus, the present results not only highlight how employing human-inspired decision-making models to control the decision dynamics of AAs has the potential to create more 'human-like' AAs, but that employing such models could also result in artificial and robotic agents that are capable of the similar levels of task generalisation and flexibility.

Although effective human-machine interaction depends on machines (either virtual avatars or physical robots) being able to predict human responses or action decisions, the production of reciprocal actions and movements is also crucial. That is, the ability to perform human-like joint- or multi-agent action is not only dependent on replicating how human actors decide where and when move (i.e., human decision policies) but, also, on how human actors actually move (i.e., the dynamics of human motion). In Chapter 3 and Appendix A, we proposed a local dynamical control law, simplified from [107, 109], that was able to capture the coordinated motion that human herders produce during the table top herding task. Interestingly, we discovered that although not explicitly defined in the model, the same oscillatory containment behaviour observed in human pairs still emerged, suggesting that while the more complicated model in [109] may have better captured the intentional dynamics of human actors (i.e., the intentional engagement in coupled oscillatory containment; see [117]), the simplified version explored here is sufficient to replicate the patterns of behavioural movement human herders exhibit during the tabletop herding task.

Perhaps more importantly, however, we also demonstrated how the control strategy investigated and employed here could solve the herding problem *without* assuming that (i) targets flock and (ii) herders have global knowledge of the state of the herding system. Indeed, to the best of our knowledge, our approach is the only one available in the literature that is able to drive multiple herders to collect and contain a herd of non-flocking target agents (whose intrinsic dynamics are stochastic) using only partial or local state information. One implication is that our approach could be generalised from the design of herder robots or AAs required to corral targets whose dynamics are known [53, 151, 160], to the design of herder robots and AAs that could successfully corral targets with unknown or not fully modelled dynamics (e.g., crowds to be evacuated or oil spills to be cleaned).

We acknowledge that although some of the target selection strategies investigated in the current Thesis did not always rely on global knowledge of the herding systems, the inputs to the SML-ANN decision models did (by-and-large) entail full state information. Accordingly, another interesting avenue of future research is to explore whether the SML-ANN approach proposed here can also model human decision making when full access to the state of the behavioural or task system is inaccessible. This could be addressed, for example, by attempting to model the target selection decisions of human players in the desert herding game (Chapter 6), but only using the local state input information provided in the first-person field of view; i.e. in the experimental condition of team members not equipped with Heads-Up Displays.

Finally, to our knowledge, no previous research has employed an explainable AI technique to try to understand and explain the decision making behaviour of human actors during skilful action, let alone identify the differences between expert and novice actors (Chapter 5) or experimental con-

ditions (Chapter 6) within the context of coordinated joint-action. To date, research on explainable AI has predominately focused on the ability of such techniques to make AI models more understandable to human users [48, 165] and to argument or enhance the decision making capabilities of human users [4]. And, while these work have often drawn connection to cognitive and psychologies models and theories of human decision making, the utility of explainable AI for specifically understanding the how, why and when of human decision making has been of little interest (for an exception, see [100]).

It is important to appreciate that we openly acknowledge that employing explainable AI and ANN models to understand human decision making assumes (i) that the input features employed for model training contains the informational variables employed by human actors and (ii) that there is a relationship between the dynamics of network weights on input features and the actual dynamical information employed by human actors when making those task decisions. Although this Thesis provide initial support for the possibility that explainable AI techniques can provide a powerful tool for explaining the decision making process of human actors, including what information best supports optimal or near optimal task performance, these assumptions need to be further validated in future work. Furthermore, while we focused on informational variables relevant to a visual-motor coordination task in this Thesis, the modelling and explainable-AI approach proposed here could be employed across a wide array of other task and informational settings (i.e., visual, auditory, haptic, linguistic, etc). Indeed, the implications for both basic scientific research and the applied development of training and decision making assessment tools seems unbounded, and, thus, the validity of employing explainable AI methods to identify and explain the structural organisation and context sensitivity of human behaviour and decision-making should become a focal area of future research.

APPENDIX A

EMERGENCE OF OSCILLATORY MOTIONS IN HERDING MODELS

In this appendix, we will first detail the human-inspired model proposed by Nalepka *et al.* in [107, 109], with a particular focus on how the model components influence the effectiveness and patterning of behavioural dynamics that emerge (Section A.1). We then derive a simpler herding model (Section A.2) by removing the terms that explicitly model the oscillatory behaviour observed in human behaviour. The comparison of the original model and the simplified version (Section A.3) motivates the choice of the latter to be the local control rules presented in Chapter 3, and later adopted in Chapter 4 to test human-inspired target selection strategies.

A.1 Human-inspired herders

As reported in Chapter 2, the herding problem has been adopted as paradigmatic example to study the emergence of coordinated behaviours in human joint tasks [106, 109, 129]. This research on dyadic human herding has observed that successful pairs of players discovered two herding strategies: search and recovery (S&R) and coupled oscillatory containment (COC). Initially all human pairs adopt an S&R strategy, with players appearing to corral the farthest target agent from the containment region within a dynamic region of responsibility. A smaller subset of successful pairs also discover an oscillatory mode of target containment (termed coupled oscillatory containment or COC) where both players create a oscillatory "wall" around the herd (refer to Figure 3.1 for an illustration of the spatial arrangement of herders and target agents).

The movement dynamics exhibited by human herders was modelled using environmental coupled task dynamic model [134, 135]. In short, using a polar coordinate system centred at the middle of the containment area ($\mathbf{x}^{\star}(t)$), the movement dynamics of a human herder j was captured by modelling changes in the herder's radial $r_j(t)$ and angular $\theta_j(t)$ distance via the dynamical equa-

tions

$$r_j(t) = -b_r\,\dot{r}_j(t) - R(\tilde{\mathbf{x}}_{i,j}, t), \tag{A.1}$$

$$\theta_j(t) = -b_{\theta,j}(t)\dot{\theta}_j(t) - T(\tilde{\mathbf{x}}_{i,j}, t) + O_{\mathrm{R,VP}}(t) + (1-\xi_j)\,O_{\mathrm{HKB}}(t), \tag{A.2}$$

Here the time-varying terms $R(t)$ and $T(t)$ are nonlinear mass-spring equations that attract the herders radial position $\mathbf{y}_j(t) = r_j\,e^{\mathrm{i}\theta_j}$ to the location of the target to be corralled $\tilde{\mathbf{x}}_{i,j}(t) = \tilde{\rho}_{i,j}\,e^{\mathrm{i}\tilde{\phi}_{i,j}}$ and defined as

$$R(\tilde{\mathbf{x}}_{i,j}, t) = \epsilon_r\left[r_j - \xi_j\,(\tilde{\rho}_{i,j} + \Delta r^\star) - (1-\xi_j)\,(r^\star + \Delta r^\star)\right], \tag{A.3}$$

$$T(\tilde{\mathbf{x}}_{i,j}, t) = \epsilon_\theta\left[\theta_j - \xi_j\tilde{\phi}_{i,j}\right]. \tag{A.4}$$

Here ξ_j is defined as

$$\xi_j(t) = \begin{cases} 1, & \tilde{\rho}_{i,j}(t) \geq r^\star \\ 0, & \tilde{\rho}_{i,j}(t) < r^\star \end{cases} \tag{A.5}$$

so that the herder dynamics is attracted to the position of the i-th chased target (plus a radial bias $\Delta r^\star$) when the current target is outside the containment region ($\xi_j = 1$) or to the boundary of the buffer region otherwise ($\xi_j = 0$).

The intrinsic oscillatory dynamics observed in each human herder was captured by the nonlinear term [70], composed by Rayleigh-like and Van der Pol-like oscillators

$$O_{\mathrm{R,VP}}(t) = \beta\,\dot{\theta}_j^3 + \gamma\,\theta_j^2\,\dot{\theta}_j, \tag{A.6}$$

with $\beta > 0$ and $\gamma > 0$.

Once the target agents are within the containment region, the parametric Hopf-bifurcation [55] process $b_{\theta,j}$

$$\dot{b}_{\theta,j} + \delta\left[b_{\theta,j} - \eta\big(\tilde{\rho}_{i,j} - (r^\star + \Delta r^\star)\big)\right] = 0, \tag{A.7}$$

where δ and η are positive constants, causes the model to produce oscillatory movements around the gathered herd. More precisely, the value of $b_{\theta,j}(t)$ governs the switch between S&R and COC strategies in the herder dynamics. If the radial distance $\tilde{\rho}_{i,j}$ of the targeted agent is greater than $(r^\star + \Delta r^\star)$ (that is, it is outside the buffer region), then $b_{\theta,j}(t)$ will quickly settle to a positive value to which it corresponds a damped dynamics of θ_j. On the other hand, if the target is inside the buffer region, then $b_{\theta,j}(t)$ becomes negative and so the dynamics of θ_j will be oscillatory (see [106, 109, 117] for more details).

Inspired by the work of Haken-Kelso-Bunz (HKB) [49], the function HKB(t) is employed to capture the coupled (in-phase and anti-phase) nature of the oscillatory containment behaviour that emerges between human herders. This function takes the form

$$O_{\mathrm{HKB}}(t) = (\dot{\theta}_j - \dot{\theta}_{j'}) \cdot (A - B(\theta_j - \theta_{j'})^2), \tag{A.8}$$

where $\dot{\theta}_{j'}$ and $\theta_{j'}$ denote the angular velocity and position of the partner herder.

It is worth noting that simulated herders driven by (A.1)-(A.2) presented in the original research papers [106, 109] were assigned a specific portion of the herding plane to be responsible for during game play and only monitored and corralled targets within this region of interest. This division strategy was defined as *static arena partitioning* in Section 3.3, where we also presented other possible dynamic division strategies of the herding environment.

The model described by (A.1)-(A.2), and first proposed by Nalepka *et al.*, was well motivated by previous research on the dynamic of human behaviour but adopted a theoretically driven, *top-down* approach to qualitatively model the experimental observations in [106, 107, 109]. In particular, the model effectively replicates the coupled oscillatory behaviours exhibited by human players by *explicitly* modelled non-linear oscillatory behaviour into the herder's time-varying parameter $b_{\theta,j}$ and the switch between S&R and COC behaviour via the switching term ξ_j.

A crucial open question was to understand whether the emergence of oscillatory behaviour could stem from simpler local rules of motion and interaction, and could therefore be captured by employing a more *bottom-up* modelling approach.

A.2 A simplified herder model

To identify the essential ingredients in the herder model required to replicate observed emerging behaviours, we gradually simplified the human-inspired model (A.1)-(A.2) and compared the emerging dynamics of these models (note that we used the same "static arena partitioning" strategy, presented in Section 3.3, in all simulations). More specifically, we compared the emerging properties and performance of four different dynamical models;

1. the *qualitative human-inspired* top-down model (A.1)-(A.2),

2. the *non-cooperative* herder model, defined as

$$r_j(t) = -b_r \dot{r}_j - R(t) \tag{A.9}$$
$$\theta_j(t) = -b_\theta \, \dot{\theta}_j - T(t) - O_{\mathrm{R/VP}}(t) \tag{A.10}$$

   obtained by removing the coupling term (A.8) to investigate if direct coordination among herders influences herding abilities,

3. the *non-oscillatory* herder model, defined as

$$r_j(t) = -b_r \dot{r}_j - R(t) \tag{A.11}$$
$$\theta_j(t) = -b_\theta \dot{\theta}_j - T(t) + (1-\xi_j)\, O_{\mathrm{HKB}}(t) \tag{A.12}$$

   obtained by removing the nonlinear oscillatory terms (A.6) to investigate how intrinsic oscillating properties affect herding performances,

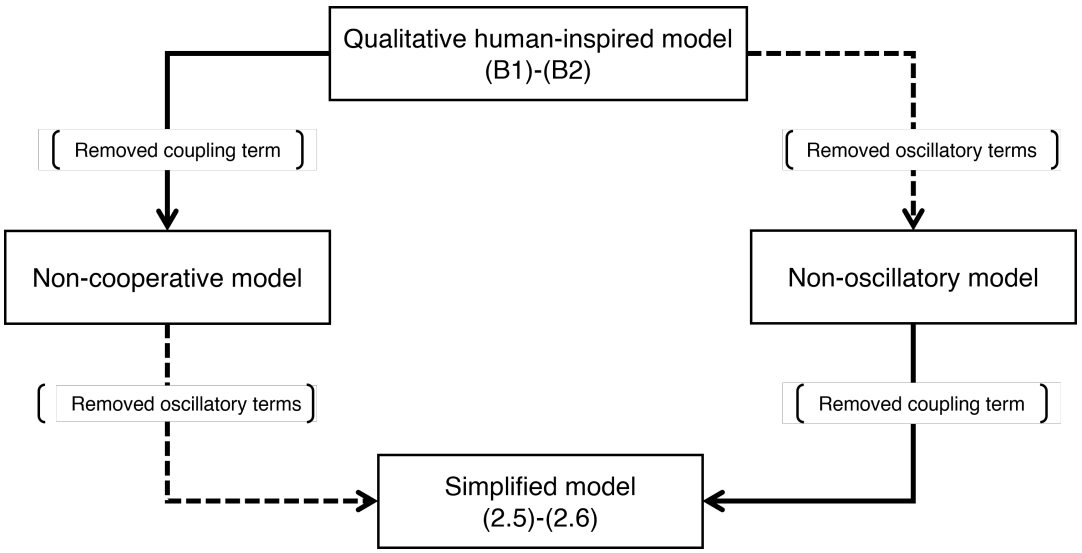


Figure A.1: Hierarchical scheme of the herder models in Section A.2. The arrows represent the logical relations between the models. Specifically, solid arrows stand for the removal of the coupling term O_{HKB} while dashed arrows stand for the removal of Rayleigh-like and Van der Pol-like non-linear terms $O_{\text{R/VP}}$.

4. the *simplified model*, discussed in Chapter 3 and defined by (3.5)-(3.6), obtained by removing the coupling and oscillatory terms to measure herding performance when herders can only chase and corral targets.

The last three models will be also referred to as *bottom-up* designed models. A hierarchical scheme of the herders models above is reported in Figure A.1.

A.3 Numerical validation and comparison

The herding performance of the above models have been compared through the evaluation of the following performance metrics; the gathering time t_{g}, the average length d_{g} of the path travelled by the herders until all targets are contained, the average total length d_{tot} of the path travelled by herders during all the herding trial, the mean distance D_T between the herd's centre of mass and the centre of the containment region, and the herd agents' spread $S_{\%}$. Additionally, as done in [106], the onset of COC was identified by analysing the power spectra of the herders' motion. See Appendix B.1 for their definitions and further details.

For each model we carried out 50 numerical simulations for $N_H = 2$ artificial herders and $N_T = 7$ target agents. See Appendix B.2 for further details on parameters and simulation methods. In Tables A.1, the performance metrics for the four herder models are reported as a function of whether the angular damping parameter b_θ varied according to (A.7) or was kept constant to a positive value.

It clearly appears that all models presented better herd aggregation index values (i.e., in terms of D_T and $S_{\%}$) in the second case with *constant* damping b_θ.

From Table A.1, when $b_{\theta,j} > 0$, we can observe that (i) bottom-up designed models (i.e., models 3 and 4) outperform the complete model in fully successful trials, (ii) models 3 and 4 have overall better scores than the more complex models 1 and 2, that embed the non-linear terms $O_{\mathrm{R/VP}}$, and (iii) the coupling term O_{HKB} in (A.8) between herders has not any significant effect on the herding performance of the simplified models 3 and 4 that differ only by this term. Moreover, from Table A.1

	Human-insp.	Non-coop.	Non-Oscil.	Simpl.
$b_{\theta,j}(t)$ (A.7)				
Successful trials	42	38	49	49
t_{g} [a.u.]	15.2 ± 5.2	18 ± 11.3	10.5 ± 1.3	13.1 ± 11.1
d_{g} [a.u.]	111.4 ± 48	143 ± 125	673 ± 112	2454 ± 2511
d_{tot} [a.u.]	949 ± 19	1023 ± 23	6942 ± 59	3933 ± 2253
D_T [a.u.]	2.8 ± 0.4	2.7 ± 0.3	3.2 ± 0.4	1 ± 0.3
$S_{\%}$ [%]	0.9 ± 0.3	0.9 ± 0.4	0.8 ± 0.3	0.7 ± 0.2
COC pairs	41	37	49	49
S&R pairs	0	1	0	0
$b_{\theta,j} > 0$				
Successful trials	5	17	50	50
t_{g} [a.u.]	32 ± 7.6	31 ± 16.5	9.3 ± 2	9.3 ± 1.9
d_{g} [a.u.]	159 ± 19	147 ± 87	61 ± 38	61 ± 18
d_{tot} [a.u.]	484 ± 42	482 ± 32	569 ± 21	624 ± 22
D_T [a.u.]	1 ± 0.3	1 ± 0.2	0.6 ± 0.1	0.7 ± 0.2
$S_{\%}$ [%]	0.2 ± 0.08	0.2 ± 0.09	0.3 ± 0.04	0.2 ± 0.03
COC pairs	0	0	50	50
S&R pairs	5	16	0	0

Table A.1: Average performance and standard deviation over successful trials of different herder models for $N_T = 7$ target agents and $N_H = 2$ herders.

it appears that the two simplified models also showed a more consistent emergence of the oscillatory behaviour with respect to the other two (see Figure A.2 for a more detailed example of spectral classification). At the same time, it is necessary to stress that, although the human-inspired model (A.1)-(A.2) and model 2 explicitly embed oscillatory terms, this property not necessarily translated in the emergence of COC pairs as in the original work from Nalepka *et al.* [106, 109]. This could be due to the differences in the numerical integration of the models and future experiments could investigate to what degree such models are influenced by integration methods and targets' dynamics.

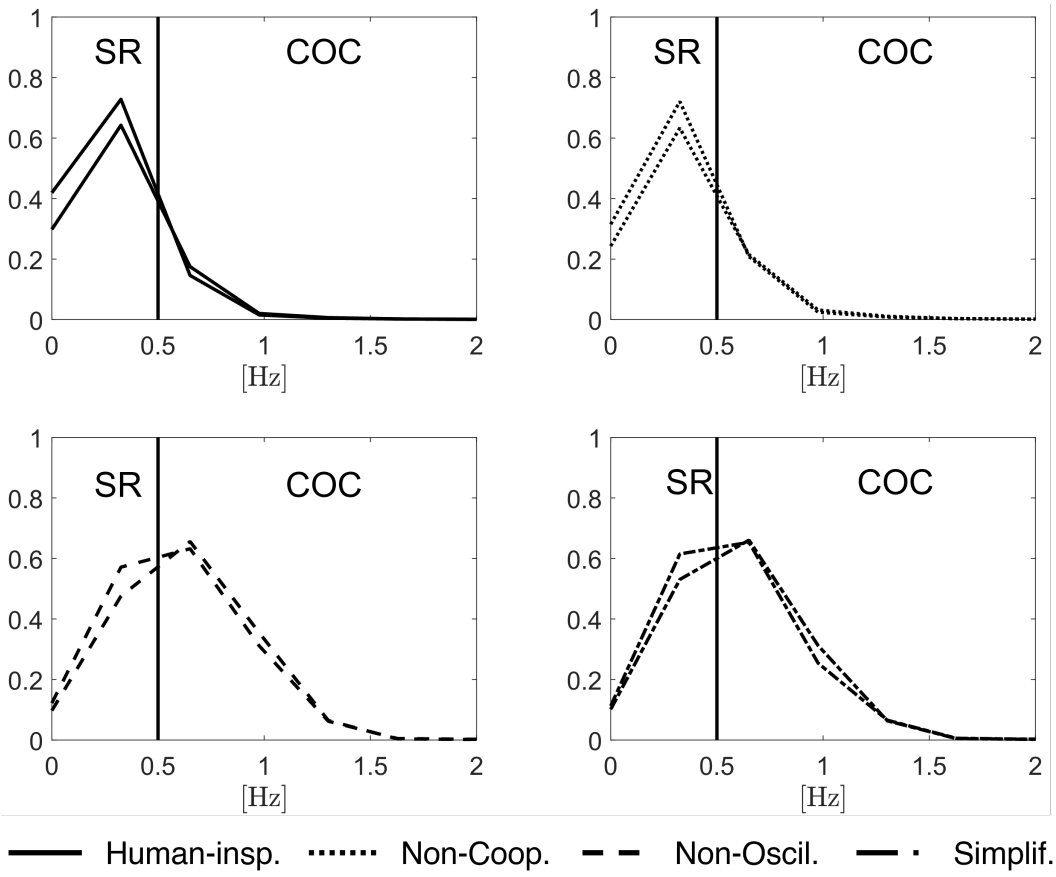


Figure A.2: Comparison of power spectra exhibited by $N_H = 2$ artificial herders in a successful trial with positive constant angular damping when corralling $N_T = 7$ target agents. Peak values of power spectrum are used to classify the coupled behaviour in S&R and COC (see Appendix B.1). Herders driven by the qualitative human-inspired model (A.1)-(A.2) (solid line) have a peak frequency on the left side of the threshold $\omega_c = 0.5\,\text{Hz}$ while herders driven by the simplified model (3.5)-(3.6) (dash-dotted line) have a peak frequency on the right side.

A.4 Summary

In this Section, we have pointed out the three main components of the qualitative human-inspired model presented in [107, 109]; a mass-spring-damper core, a pair of nonlinear oscillatory terms (A.6), and a coupling term (A.8). From a control theory perspective, only the core component is essential, as it is enough to describe the motion of a body (the herder) moving from one point in the plane to another in the process of chasing and steering a target towards the containment region. Therefore, in view of the above considerations, among those presented here, the simplified model (3.5)-(3.6), composed of local control rules, is the most appropriate choice since it guarantees good herding performance with the simplest possible dynamics. Furthermore, these control rules also adequately replicate the emerging behaviours observed in experiments and suggests that the source

of these behaviours could potentially be a function of the target selection strategy employed rather than due to intrinsic oscillatory dynamics being entailed within the herder model. That is, in other words, these behaviours could emerge from the mutual collaboration and division of labour (i.e,. coordinated target selection).

APPENDIX

B

MATERIALS AND METHODS

In this Appendix we detail the methods employed to implement and study the results presented in this thesis. Sections B.1-B.3 describe performance measures and numerical integration process employed to validated and analyse the dynamic target selection strategies investigated in Chapter 3. Sections B.4-B.6, we detailed how relevant system variables were extracted from the experimental data, the learning algorithm employed in Chapters 4-6, and define prediction performance measures employed. Finally, the SHAP algorithm implemented in Chapters 5-6 is described in Section B.7.

B.1 Herding performance measures

Let $X(t) := \left\{ i : \|\mathbf{x}_i(t) - \mathbf{x}^\star\| \leq r^\star \right\}$ denote the set of target agents which are contained within the goal region G at time t (the spatial arrangement of the herding problem is illustrated in Figure 3.1), and $[0, T]$ denote the time interval over which the performance metrics are evaluated.

The *gathering time* is defined as the time instant $t_g \in [0, T]$ that condition (3.2) holds. That is, all the target agents are in the containment region for all $t \geq t_g$.

Distance travelled by the herders measures the mean distance travelled by the herders during the time interval $[0, t]$. It is defined as

$$d(t) := \frac{1}{N_H} \sum_{j=1}^{N_H} \frac{1}{t} \left(\int_0^t \left\| \dot{\mathbf{y}}_j(\tau) \right\| d\tau \right). \tag{B.1}$$

Therefore, $d_g := d(t_g)$, and $d_{tot} := d(T)$.

Herd distance from containment region measures the herders ability to keep the herd close to the containment region, with centre $\mathbf{x}^\star$. It is defined as the mean in time of the Euclidean distance

between the centre of mass of the herd and the centre of the containment region, that is

$$D_T := \frac{1}{T}\int_0^T \left\| \left(\frac{1}{N_T}\sum_{i=1}^{N_T} \mathbf{x}_i(\tau) \right) - \mathbf{x}^\star(\tau) \right\| d\tau. \tag{B.2}$$

Herd spread measures how much scattered the herd is in the game field. Denote as $\mathrm{Pol}(t)$ the convex polygon defined by the convex hull of the points $\mathbf{x}_i$ at time t, that is,

$$\mathrm{Pol}(t) := \mathrm{Conv}\left(\{\mathbf{x}_i(t),\, i = 1,\ldots,N_T\}\right).$$

Then, the herd spread S is defined as the mean in time of the area of this polygon, that is

$$S := \frac{1}{T}\int_0^T \left(\int_{\mathrm{Pol}(\tau)} d\mathbf{x} \right) d\tau. \tag{B.3}$$

The herd spread can also be evaluated with respect to the area of the containment region, $A_{\mathrm{cr}} = \pi(r^\star)^2$, as $S_{\%} = S/A_{\mathrm{cr}} \cdot 100$.

Note that *lower* values of t_{g} correspond to *better* herding performance; herders taking a shorter time to gather all the target agents in the goal region. Also, *lower* values of D_T and $S_{\%}$ correspond to a *tighter* containment of the target agents in the goal region while *lower* values of d_{g} and d_{tot} correspond to a *more efficient* herding capability of the herders during the gathering and containment of the herd.

Containment rate measures the herders' ability to relocate one or more targets inside the desired region. It is defined as the mean in time of the percentage of target agents in the containment region, that is

$$I_{\%} := \frac{1}{\Delta T}\int_{t_i}^{t_f} \frac{|X(\tau)|}{N_T}\, d\tau \cdot 100. \tag{B.4}$$

The emerging behaviour of a herder can be evaluated through its power spectra [106, 109]. The *behavioural-classification index* of the j-th herder is defined as

$$\varphi^{(j)} = \frac{\omega_{\mathrm{freq}}^{(j)} - \omega_c}{|\omega_{\mathrm{freq}}^{(j)} - \omega_c|}\omega_{\mathrm{power}}^{(j)} \tag{B.5}$$

with $\omega_{\mathrm{freq}}^{(j)}$ being the dominant frequency component, $\omega_{\mathrm{power}}^{(j)}$ the corresponding power, and ω_c the frequency threshold empirically determined at 0.5 Hz, as in [106, 109]. A pair of herders is considered to adopt a search and recovery (SR) behaviour if the behavioural-classification index for both herders $\varphi^{(j)} < 0$, or to adopt a coupled oscillatory containment (COC) behaviour if for both herders $\varphi^{(j)} > 0$.

B.2 Numerical integration

In all numerical simulations reported in Section 3.4 we considered the case of $N_H = 2$ or $N_H = 3$ artificial herders and $N_T = 7$ target agents. Moreover, we considered a circular containment region

G with radius $r^\star = 1$, centred in $\mathbf{x}^\star = \mathbf{0}$, and a buffer region of width $\Delta r^\star = 1$. The numerical integration[1] of the differential equations describing the dynamics of target agents and herders was implemented using Euler-Maruyama method [58] in the time interval $[0, T] = [0, 100]$ s with step size $dt = 0.006$ s. With regard to target selection, herder agents computed their next target-to-be-chased each $t_{\text{dwell}} = 50\,dt$ s Note that numerical trials in Chapter 4 have been simulated with a step size of $dt = 0.02$ s to match the sampling time used to collect experimental data [129].

The values of all parameters used in the simulations are reported in Table B.1 and were chosen as in [107].

$[0, T]$	Time interval	[0,100 s]
dt	Step size	0.006 s
N_H	Number of herders	{2, 3}
N_T	Number of targets	{4, 7, 60}
$\mathbf{x}^\star$	Centre of containment region	$\mathbf{0}$
$r^\star$	Radius of containment region	1
$\Delta r^\star$	Width of buffer region	1.0005
r_c	Collision detection radius	0.0001
α_b	Diffusive motion coefficient	0.05
α_r	Repulsive reaction coefficient	$20\,\alpha_b$
b_r	Radial damping coefficient	10.998
ϵ_r	Radial stiffness coefficient	98.706
b_θ	Angular damping coefficient	10.998
ϵ_θ	Angular stiffness coefficient	61.62
δ	Angular damping dynamics parameter	23.089
η	Angular damping dynamics parameter	80.592
β	Rayleigh-like term coefficient	0.161
γ	Van der Pol-like term coefficient	7.2282
A	HKB term parameter	-0.2
B	HKB term parameter	0.2

Table B.1: Parameters' values for numerical simulation of herding agents.

The initial positions of the target agents were set outside the containment region as $\mathbf{x}_i(0) = 2\,r^\star e^{i\phi_i(0)}$, $\forall i = 1, \ldots, N_T$, with $\phi_i(0)$ drawn with uniform distribution in the interval $(-\pi, \pi]$, while the initial positions of herders was defined on a circle with radius $4r^\star$ and with angular displacement $(2\pi)/N_H$, to exclude from the simulations the rearrangement of the herders otherwise occurring at the beginning of each trial. Furthermore, collision avoidance forces between target agents was also considered in the numerical simulations. Specifically, the model (3.3) is extended by adding the term $\mathbf{V}_{c,i}(t)dt$, with

$$\mathbf{V}_{c,i}(t) = \sum_{i' \in X_{c,i}(t)} \frac{\mathbf{x}_{i'}(t) - \mathbf{x}_i(t)}{\|\mathbf{x}_{i'}(t) - \mathbf{x}_i(t)\|^3}, \tag{B.6}$$

[1] Matlab code available at `https://github.com/diBernardoGroup/HerdingProblem`

where $X_{c,i} := \{i' : \|\mathbf{x}_{i'} - \mathbf{x}_i\| \leq r_c\}$ is the set of all target agents at time t inside the closed ball centred in $\mathbf{x}_i$ with radius r_c.

B.3 ROS simulations

The mobile robots used for both target and herder agents were designed as Pioneer 3-DX[2] robots driven by the differential drive controller provided in the set of ROS packages (`gazebo-ros- pkgs`[3]) that allows the integration of Gazebo and ROS[4] .

The environment and the robots share information through an exchange of messages that occurs via publishing and subscribing to one or more of the available topics. A ROS node is attached to each herder and target robots. It subscribes to the `/odom` topic; implements the agent's dynamics; and publishes a personalised `/cmd_vel` topic. The target agents collect odometric information from all the herders in the environment. The herder agents subscribe to the ID of the target agent to-be-chased and collect its position. The published message is a velocity control input w.r.t. the robot's reference system to the differential drive of the robot: a translation v along x-axis and a rotation ω around z-axis of the robot. The reference trajectory $\mathbf{y}^\star(t) = [r^\star \cos\theta^\star, r^\star \sin\theta^\star]^\top$, generated as in Sections 3.2-3.3, is followed by each robot by means of the Cartesian regulator

$$v = -k_v(\mathbf{y} - \mathbf{y}^\star)\,[\cos\Phi, \quad \sin\Phi] \tag{B.7}$$

$$\omega = k_\omega(\theta^\star - \Phi + \pi) \tag{B.8}$$

where $\Phi(t)$ denotes the robot orientation w.r.t. the global reference system. The gains $k_v = 0.125$ and $k_\omega = 0.25$ have been tuned by trial-and-error to achieve smooth robot movements. The initial position of the agents have been set as in Appendix B.2.

The target selection strategies (Section 3.3.2) are processed in an ad hoc ROS node. It subscribes to the odometry topic, computes the user-chosen strategy (i.e. global, static arena partitioning, leader-follower or peer-to-peer), and publishes a custom message with the ID of the targets to-be-chased on the `/herder/chased_target` topic. The custom message is an array of integer numbers, its j-th element corresponds to the target agent chased by the j-th herder robot (Figure B.1).

The Gazebo-ROS simulations were run on Ubuntu 18.0404 LTS hosted on a Virtual Machine with a 10GB RAM with ROS Melodic distribution and Gazebo 9.13.0.

B.4 Features extraction and processed dataset

In Chapter 4, from position and velocity data[5] recorded in the original novice and expert data-sets [129] we extracted and derived the following $N_{sv} = 48$ state variables:

[2] Pioneer 3 - Operations Manual, available at `https://www.inf.ufrgs.br/~prestes/Courses/Robotics/manual_pioneer.pdf` (2020/08/11)

[3] `http://wiki.ros.org/gazebo_ros_pkgs`

[4] `https://www.ros.org`

[5] Raw data for Chapters 4-5 available at `https://github.com/FabLtt/ExplainedDecisions`

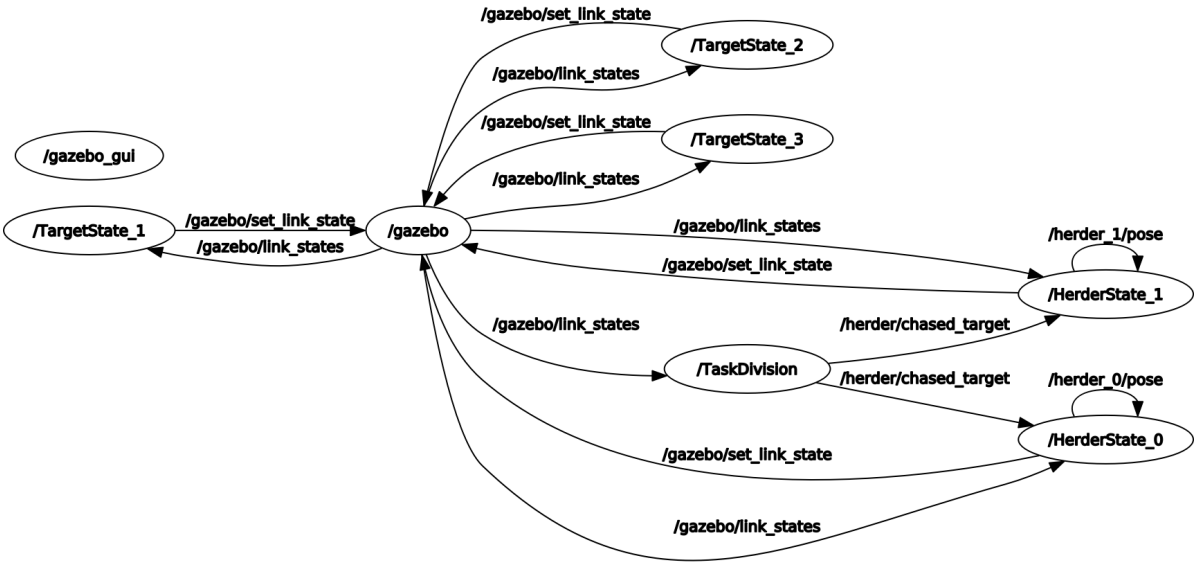


Figure B.1: Overview of the control architecture of the designed Gazebo-ROS application. The robots exchange information on the communication graph through the messages published on the available topics.

- the radial and angular distance (Δ, Ψ) between herders,
- the radial and angular distance ($\Delta_{i,j}$,$\Psi_{i,j}$) of target i from herder j,
- the radial and angular distance ($\delta^{\star}_{i|j}$,$\psi^{\star}_{i|j}$) of herder j or target i from the centre of the containment region.
- the radial velocity and acceleration of herders ($\dot{r}(t)$ $\ddot{r}(t)$) and targets ($\dot{\rho}(t)$ $\ddot{\rho}(t)$),
- the direction of motion of herders and targets

Note that we tested different combinations of state variables as input features for model training (i.e., smaller and larger input feature sets). As detailed in Appendix C.1, regardless the input feature set employed for model training, we observed similar prediction performance as reported in Table 4.1 . The above feature set was employed as it represented a comprehensive description of the relative state of the herding task for a given herder.

All successful trial set data, per level of expertise, was stacked in a common *feature processed* novice or expert data-set[6] along with the corresponding target codes (ID 0 to 4). From the resultant feature processed novice and expert data-set we randomly extracted $N_{train} = 21000$ training samples and $N_{test} = 2000$ test samples. Here, samples refer to pairs of state feature sequences and target label codes. Sequences were composed by $N_{seq} = 25$ consecutive instances of the above listed N_{sv}

[6]The processed datasets discussed in Chapters 4-5 are made available in the public repository https://osf.io/wgk8e/

state variables, sampled at $dt = 0.04$, covering $T_{seq} = 1$ second system evolution, and labels are the ID of the agent being targeted, selected at $T_{hor} = dt$ seconds from the corresponding sequence for the shorter prediction horizon, and at $T_{hor} = 16\,dt$ for the longer prediction horizon.

In Appendix C.2 we also consider values of $T_{seq} = 0.5$s and $T_{seq} = 2$ s varying the sampling time to $dt = 0.02$ and $dt = 0.08$ respectively. As in the default case presented in the Chapters 4-5 the novice and expert models trained with these different T_{seq} lengths also obtain accuracy values greater than 95% when tested on data from the same expertise level (e.g., expert-expert) and closer to 50% when tested on data from the different level of expertise (e.g., novice-expert).

B.5 Learning algorithm

For each combination of expertise or experimental condition, and prediction horizon, we trained[7] a Long-Short Term Memory (LSTM) artificial neural network with Dropout layers as presented in Figure B.2 [60, 61], using Adam optimization. We used Bayesian Optimization to tune the learning rate ($\alpha = 0.0018$) of the Adam optimizer and the hyperparameters of the LSTM$_{NN}$ (i.e., the number of LSTM hidden layers, number of neurons in each layer, and dropout rates). The *InputLayer* and the output *Dense* layer of the optimized LSTM$_{NN}$ had dimensionality (T_{seq}, N_{sv}) and (T_{seq}, $N_T + 1$), respectively, where N_{sv} is the total number of input features and N_T the number of targets ID to be predicted. In the centre, 3 hidden LSTM layers of 253, 25 and 8 neurons were alternated with Dropout layers of equal dimensionality. The dropout rate of each LSTM layers of the novice and expert models in Chapter 4-5 was 0.1145 and 0.1438, respectively. For the dropout layers between each LSTM layer, the dropout rates were 0.0145 and 0.0438 for the novice and expert models, respectively.

In Chapter 6, Bayesian Optimisation indicated that the best fit for predictions at fixed horizons was the LSTM$_{NN}$ structure used for expert models. To predict at variable prediction horizon, instead, 3 hidden LSTM layers of 252, 45 and 8 neurons were alternated with Dropout layers of equal dimensionality; dropout rates of LSTM and Dropout layers being 0.1292 and 0.0292, respectively.

To avoid over-fitting, training was stopped when the logarithmic loss – that penalise false predictions – on the validation set stopped improving; the validation set being a randomly extracted 10% of the training set. The LSTM$_{NN}$ has been built and trained using Python 3.7.1 and Tensorflow[8] library.

B.6 Prediction performance measures

Performance of the LSTM$_{NN}$s were validated using the following measures: *Accuracy* – the fraction of correct predictions outputs among the samples tested; *Precision* – how valid the prediction was, that is the portion of relevant outputs among the predicted ones; *Recall* – how complete the prediction was, that is the portion of relevant outputs that were predicted. Note that when Precision

[7] Code and trained ANNs for results in Chapter 4-5 available at github.com/FabLtt/ExplainedDecisions
[8] https://www.tensorflow.org/ , version 1.15

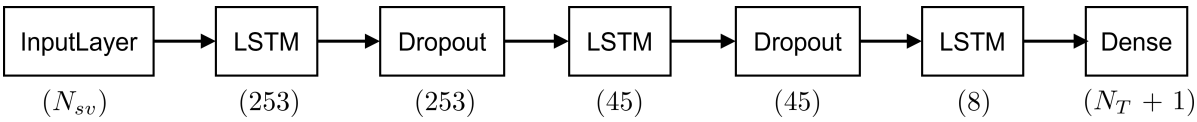


Figure B.2: Model of the Artificial Neural Network built to map the state variables extracted from experiments to the ID of the next target corralled by the current herder player in Chapter 4-5. The *InputLayer* and the output *Dense* layer have dimensionality (T_{seq}, N_{sv}) and (T_{seq}, $N_T + 1$), respectively. In the centre, 3 LSTM layers of 253, 25 and 8 neurons are alternated with Dropout layers of equal dimensionality.

and Recall reach 100%, false positive outputs and false negative outputs are absent, respectively. Additionally, we also report the *F1 score* for model prediction's, with higher values of *F1 score*, the harmonic mean of Precision and Recall, expressing how the model prediction is both precise and robust. Training and performance evaluation were implemented using Scikit-learn[9] Python library.

B.7 SHapley Additive exPlanation

Given a sample, the SHAP[10] algorithm assigns to each input feature an importance value. This is an estimate of the contribution of each feature to the difference between the actual prediction output and the mean prediction output. We randomly selected N^S_{train} = 200 samples from the training set as a prior distribution of the input space. We applied SHAP `DeepExplainer` on N_{test} = 6000 samples used to evaluate performance [92] and obtained the corresponding SHAP values[11] for each state variable. To derive the corresponding approximate global feature importance measure (shown in Figure 5.6) we averaged over the test set, for each class of prediction output (i.e., target ID), the SHAP values associated to each input feature (state variable).

[9] https://scikit-learn.org/stable/index.html, version 0.23.1

[10] https://github.com/slundberg/shap, version 0.31

[11] Computed SHAP values discussed in Chapter 5 are made available in the public repository https://osf.io/wgk8e/. These can be analysed with the code available at github.com/FabLtt/ExplainedDecisions

APPENDIX C

SUPPLEMENTARY INFORMATION TO HUMAN-INSPIRED STRATEGIES

In this appendix we provide further details on inputs sets used to train the human-inspired target selection strategies discussed in Chapters 4-5, as well as the performance of the models implemented as simulated herders.

C.1 Performance of prediction models with different input feature sets

Table C.1 lists the different input feature sets used to evaluate the ability of a Long Short-Term Memory artificial neural network (LSTM_{NN}) to model and predict human target selection decisions using Supervised Machine Learning (SML). Results presented in Chapter 4 were for the *Default* input feature set. The accuracy results for the other input feature sets are shown in Figure C.1.

C.2 Performance of prediction models with different prediction windows and horizons

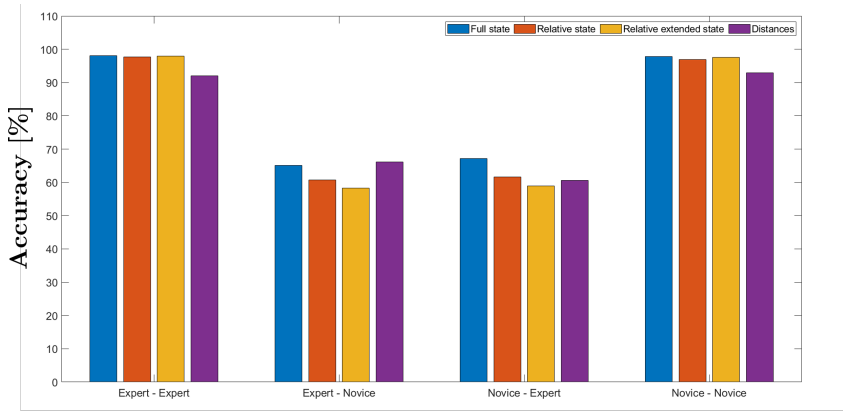
The SML algorithm presented in Chapter 4 can be customised to forecast the ID of the agent targeted that will be corralled by a human herder for different lengths of input state sequences T_{seq} and prediction horizons τ_{hor}. The algorithm's inputs correspond to time-series of relevant state variables, T_{seq}, fixed to $N_{seq} = 25$, such that T_{seq} is scaled by tuning the sampling time dt. The algorithm output prediction is the ID of the next target to be corralled by the herder at τ_{hor} in the future. Accuracy values for multiple combinations of prediction window T_{seq} and horizon τ_{hor} are reported in Figure C.2.

C.3 Performance of human-inspired selection strategies

The time evolution of the average distance from the centre of the goal region over 50 trials are reported in Figures C.3-C.6 for different combinations of $N_H = 2, 3$ artificial herders and $NT = 4, 7$ targets. The herding performance of human-inspired strategies trained using the longer prediction horizon are reported in Table C.2.

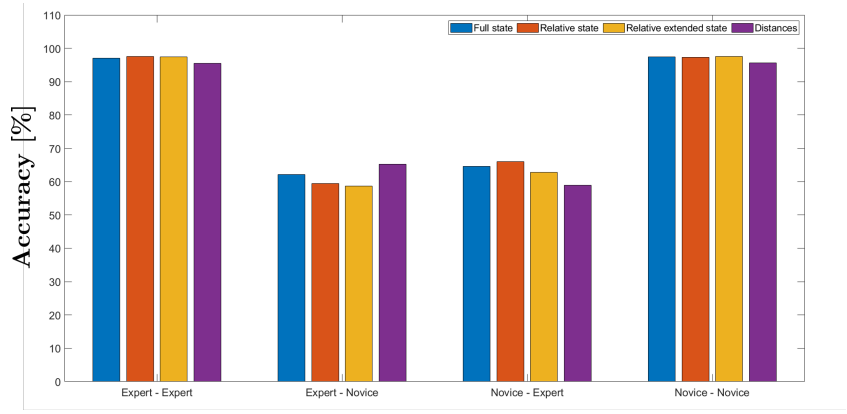
Input features	*Default*	Cartesian	Relative	Extended	Distance
Cartesian position of target		X			
Cartesian position of 1^{st} herder		X			
Cartesian position of the centre of the goal		X			
Radial distance of target i from 1^{st} herder	X	X	X	X	X
Radial distance of target i from 2^{nd} herder	X				
Angular distance of target i from 1^{st} herder	X	X	X	X	X
Angular distance of target i from 2^{nd} herder	X				
Radial distance of target i from goal	X	X	X	X	X
Radial distance of 1^{st} herder from goal	X	X	X	X	X
Radial distance of 2^{nd} herder from goal	X			X	
Angular distance of target i from goal	X				
Angular distance of 1^{st} herder from goal	X				
Angular distance of 2^{nd} herder from goal	X				
Radial distance between herders	X			X	
Angular distance between herders	X			X	
Direction of motion of target agents	X				
Direction of motion 1^{st} herder	X				
Direction of motion 2^{nd} herder	X				
Radial velocity of target agents	X	X	X	X	
Radial velocity 1^{st} herder	X	X	X	X	
Radial velocity 2^{nd} herder	X			X	
Angular velocity of target agents				X	
Angular velocity 1^{st} herder				X	
Angular velocity 2^{nd} herder				X	
Radial acceleration of target agents	X	X			
Radial acceleration 1^{st} herder	X				
Radial acceleration 2^{nd} herder	X				

Table C.1: List of features in each input feature set.



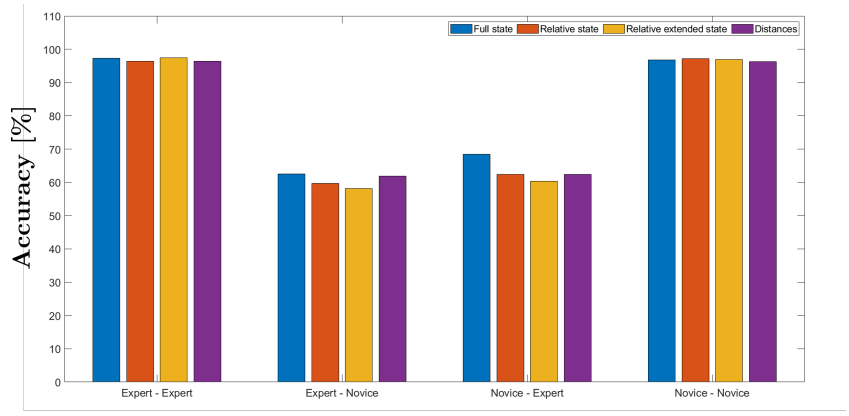
(a)

(a)



(b)

(b)



(c)

(c)

Figure C.1: Accuracy values of the multi-label predictor on $N_{test} = 2000$ samples for different combination of training and test pairs for $\tau_{hor} = dt$ and (a) $T_{seq} = 0.5$ s, (b) $T_{seq} = 1$ s and (c) $T_{seq} = 2$ s.

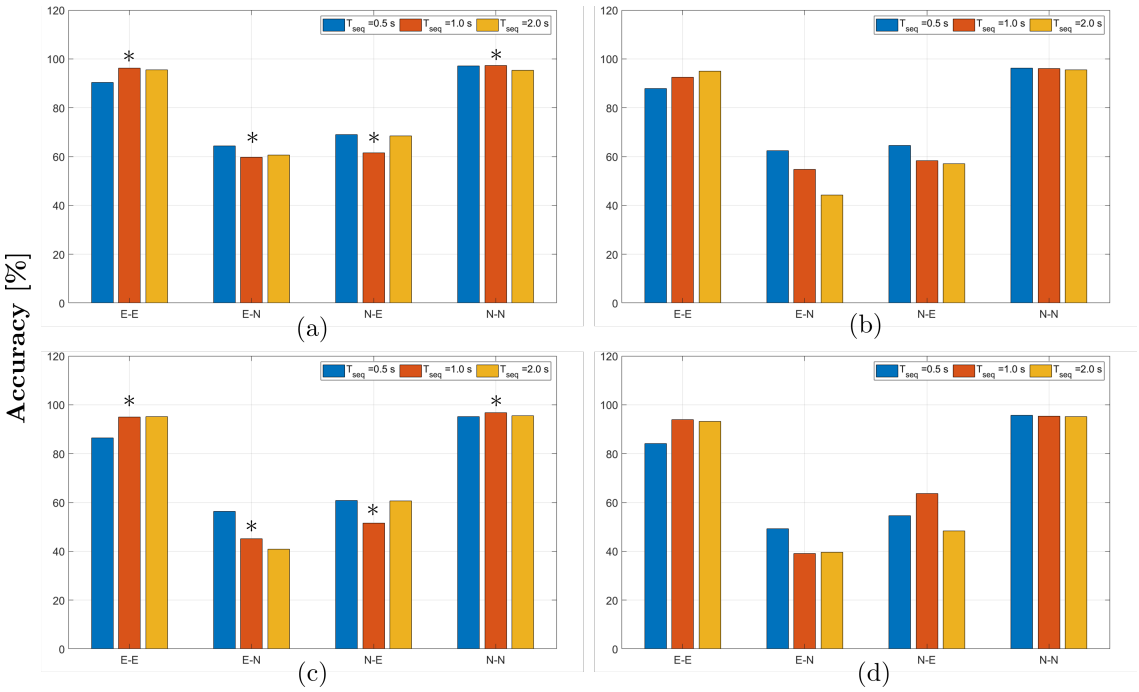


Figure C.2: Accuracy values of the multi-label predictor on $N_{test} = 2000$ samples for different combination of training and test pairs for (a) $\tau_{hor} = dt$, (b) $\tau_{hor} = 8dt$, (c) $\tau_{hor} = 16dt$, (d) $\tau_{hor} = 32dt$. $*$ indicates the accuracy values for the nominal case $T_{seq} = 1$ s reported in Chapter 5.

	$N_T = 4$		$N_T = 7$	
	NoviceLong-insp	ExpertLong-insp	NoviceLong-insp	ExpertLong-insp
$N_H = 2$				
t_g [a.u.]	13.55 ± 4.39	13.69 ± 3.79	18.55 ± 3.16	18.72 ± 4
d_g [a.u.]	130.18 ± 66.26	136.95 ± 63.05	189.55 ± 45.63	195.22 ± 54.63
D_g [a.u.]	2.33 ± 1.5	2.23 ± 1.41	0.99 ± 1.48	1.17 ± 2.07
$S_{g,\%}$ [%]	1.15 ± 0.49	1.12 ± 0.51	2.39 ± 0.48	2.28 ± 0.55
$I_{g,\%}$ [%]	23.34 ± 8.41	24.82 ± 7.78	20.5 ± 6.57	20.7 ± 7.31
$N_H = 3$				
t_g [a.u.]	9.91 ± 3.23	10.57 ± 3.16	12.77 ± 2.72	13.21 ± 2.78
d_g [a.u.]	90.07 ± 43.15	106.2 ± 43	122.46 ± 33.42	137.07 ± 37.2
D_g [a.u.]	2.05 ± 1.12	1.85 ± 1.3	2.2 ± 1.47	3.86 ± 2.23
$S_{g,\%}$ [%]	1.07 ± 0.51	1.14 ± 0.46	2.1 ± 0.44	2.07 ± 0.41
$I_{g,\%}$ [%]	26.48 ± 8.83	26.79 ± 8.88	21.88 ± 7.88	21.1 ± 6.92

Table C.2: Average performance over 50 trials of different herding strategies trained on the long prediction horizon for $N_T = 4$ and $N_T = 7$ passive agents during gathering time t_g. According to Kruskal Wallis test, significant difference between novice- and expert- inspired AA-herders is observed only in distance travelled during gathering phase d_g for $N_H = 2$ and $N_T = 7$ ($\chi^2 = 5.99$, $p < 0.02$), for $N_H = 3$ and $N_T = 4$ ($\chi^2 = 3.83$, $p = 0.05$), for $N_H = 3$ and $N_T =$ ($\chi^2 = 4.33$, $p < 0.04$).

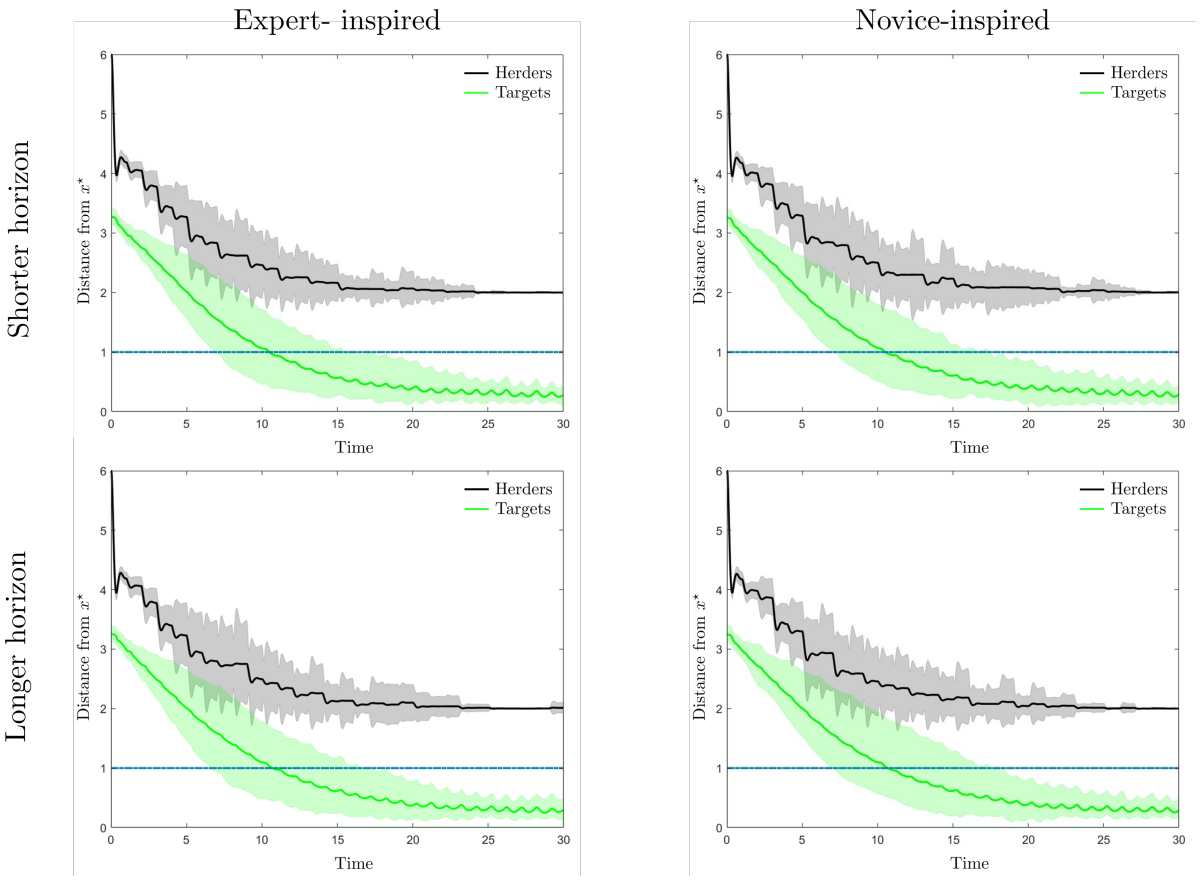


Figure C.3: Time evolution of the average distance from the centre of the containment region over 50 trials for (left column) expert-inspired and (right column) novice-inspired selection strategies for $N_H = 2$ herders and $N_T = 4$ passive agents using LSTM_{NN} trained on short and long prediction horizons. Solid lines are the mean values over agents and trials; the lighter coloured areas span the maximum and lower standard deviations from the mean reached over trials.

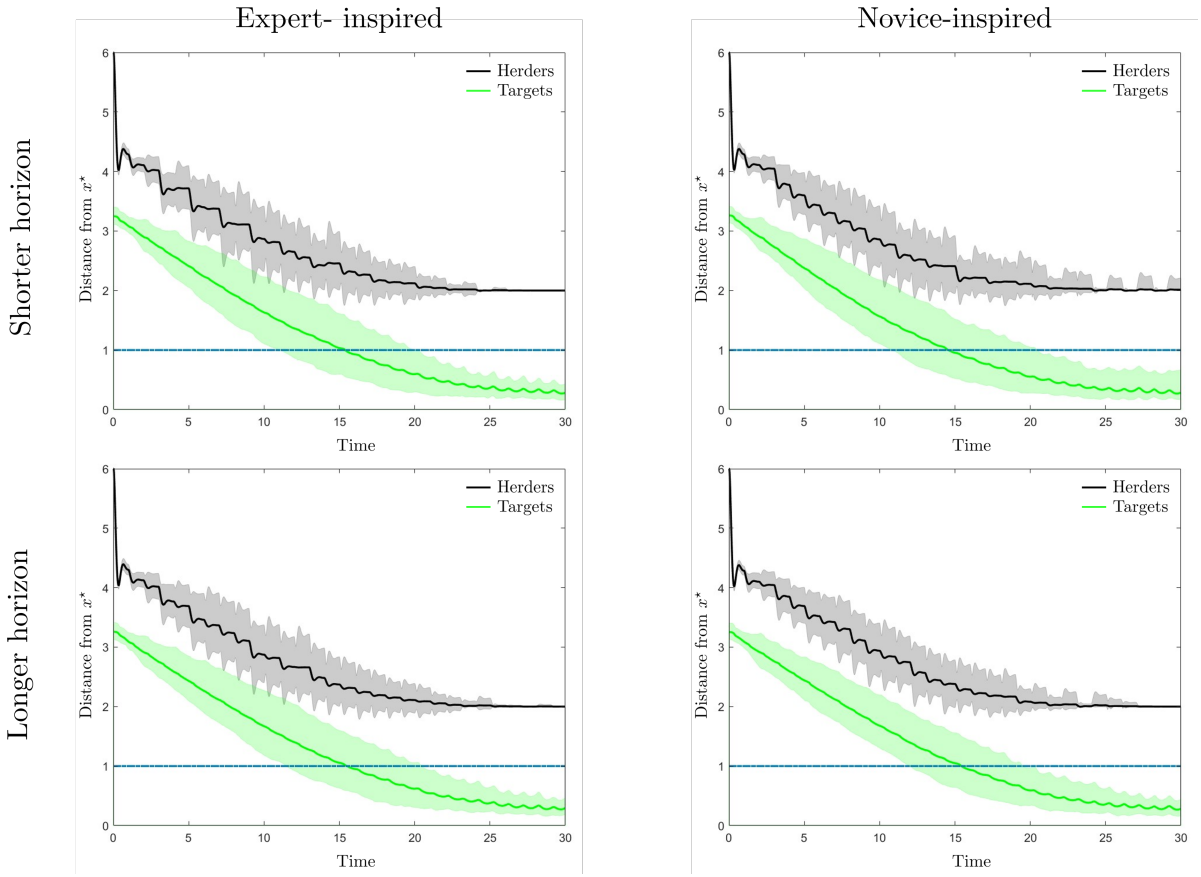


Figure C.4: Time evolution of the average distance from the centre of the containment region over 50 trials for (left column) expert-inspired and (right column) novice-inspired selection strategies for $N_H = 2$ herders and $N_T = 7$ passive agents using LSTM_{NN} trained on short and long prediction horizons. Solid lines are the mean values over agents and trials; the lighter coloured areas span the maximum and lower standard deviations from the mean reached over trials.

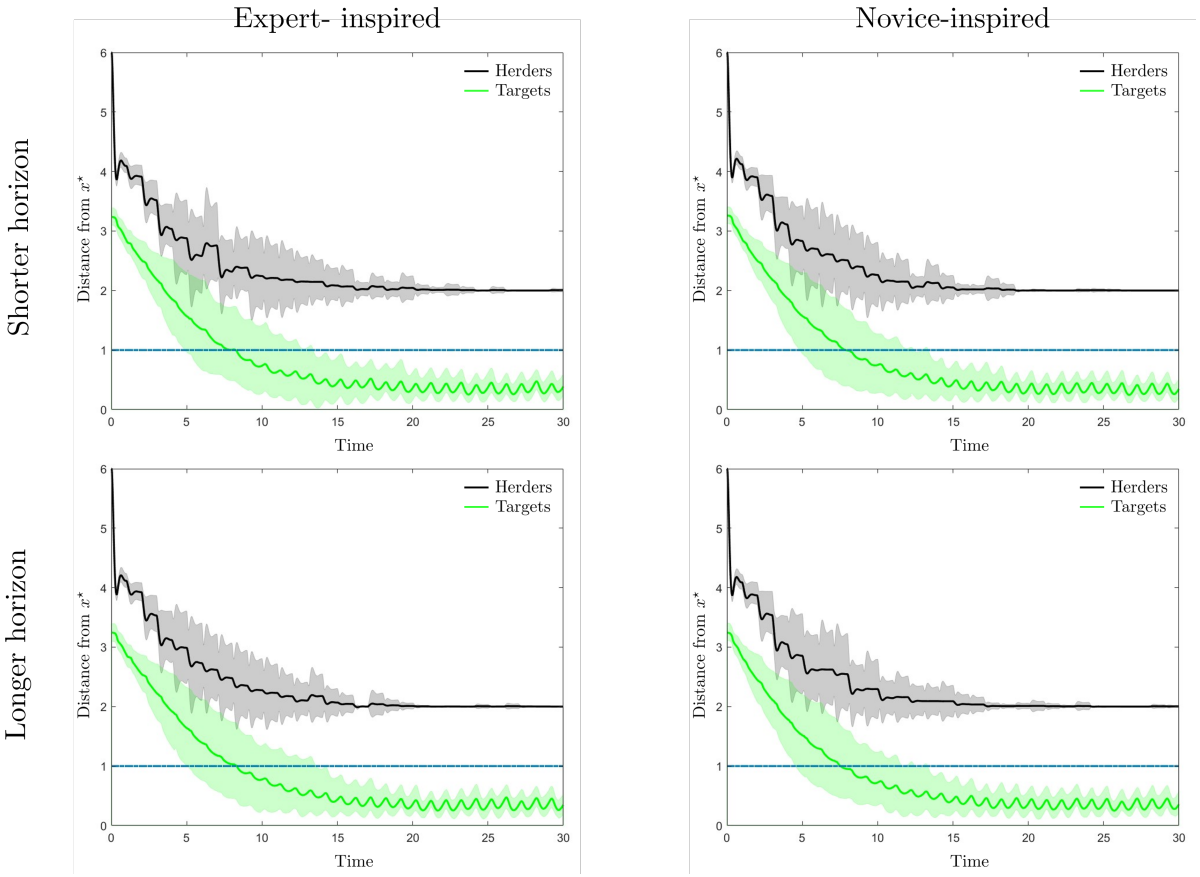


Figure C.5: Time evolution of the average distance from the centre of the containment region over 50 trials for (left column) expert-inspired and (right column) novice-inspired selection strategies for $N_H = 3$ herders and $N_T = 4$ passive agents using LSTM_{NN} trained on short and long prediction horizons. Solid lines are the mean values over agents and trials; the lighter coloured areas span the maximum and lower standard deviations from the mean reached over trials.

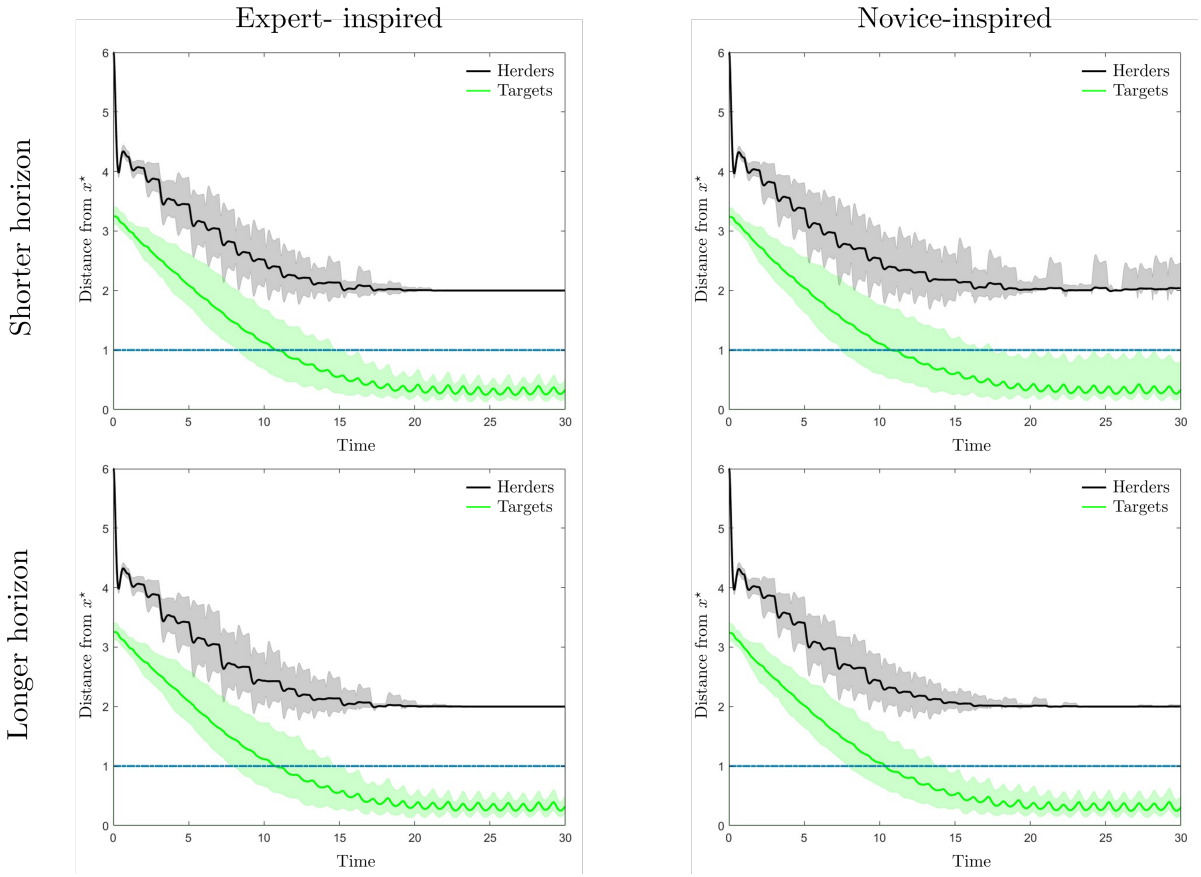


Figure C.6: Time evolution of the average distance from the centre of the containment region over 50 trials for (left column) expert-inspired and (right column) novice-inspired selection strategies for $N_H = 3$ herders and $N_T = 7$ passive agents using LSTM_{NN} trained on short and long prediction horizons. Solid lines are the mean values over agents and trials; the lighter coloured areas span the maximum and lower standard deviations from the mean reached over trials.

APPENDIX

D

SUPPLEMENTARY INFORMATION TO UNDERSTANDING JOINT HUMAN DECISION

In this appendix we provide further details on the prediction performance and explanation of the human target selection policies discussed in Chapter 5.

D.1 Performance of prediction models with different type of samples

In the time interval T_{seq}, a human herder could either continuously corral the same target agent or *transition* between different targets. Here we classified these two types of T_{seq} samples as *non-transitioning* and *transitioning* behavioural sequences, respectively. The second subcategory of sample, corresponded to whether a herder *switched* targets between t_f and $t_{f+T_{hor}}$. That is, at T_{hor}, a herder could be corralling the same target that was being corralled at the end of T_{seq} or switch to a different target, with these two possibilities classified as *non-switching* and *switching* behaviour, respectively (Figure 5.3). In Figure D.1 the distributions of the four corresponding sample types for each prediction horizon,τ_{hor} are illustrated.

Figure D.2 details the performance of the trained LSTM_{NN} models for each samples type as a function of T_{seq} = 0.5, 1 and 2 s and T_{hor} = 1, 8, 16 and 32 time steps in the future. Models were trained on a set of training samples that matched the actual distribution of samples for each T_{hor}. Note that model performance for each sample type is somewhat proportional to the number of samples of a given type in the training set. Recall, that when an even distribution of samples is employed of train $T_{hor} >> 0$ models, these models (i.e., T_{hor} = 16 or 32) perform equally well for all sample types (Chapter 5,Section 5.2.

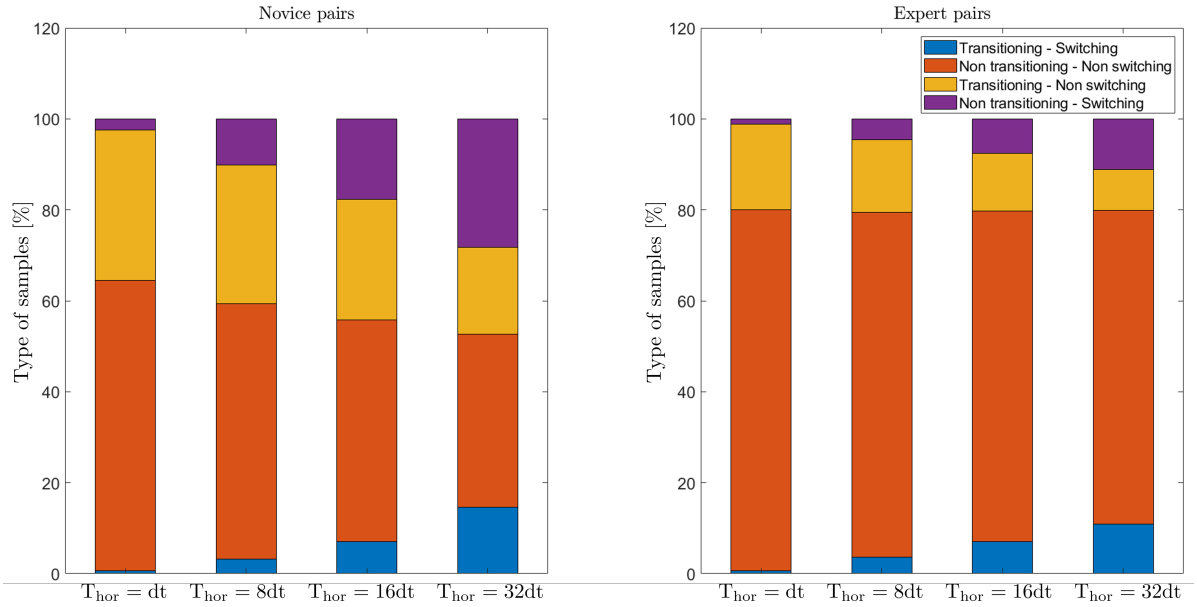


Figure D.1: Percentage different type of samples in the training set for different prediction horizon and T_{seq} = 1s.

D.2 Kendall rank correlation coefficients

The ordinal association of SHAP value rankings (the first top 10 reported in Tables D.3-D.4) was computed using the Kendall rank correlation coefficient (Kendall's τ,[98]) for subgroups of N top ranked input features. Table D.1 includes the Kendall coefficients, and associated p-values, as a function of expertise. Table D.2 includes the Kendall coefficients as a function of the short $\tau_{hor} = 1$ and long $\tau_{hor} = 16$ prediction horizon for each level expertise.

D.3 SHAP value tables

A detailed summary of SHAP feature values for each LSTM$_{NN}$ model, prediction horizon and target ID is provided in Tables D.3-D.4.

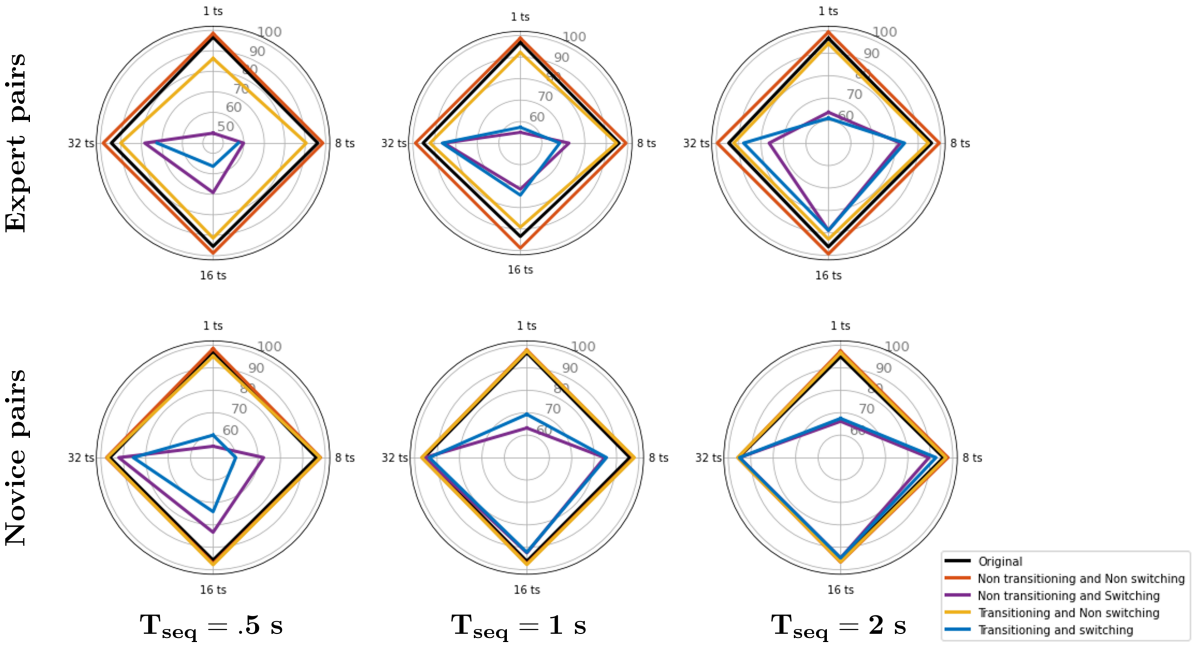


Figure D.2: Vertexes are the accuracy of the trained models for each decision time interval T_{seq} and horizon τ_{hor}. Accuracy is scored on N_{test}= 2000 of (black) mixed, (grey) non transitioning and (red) transitioning samples.

	Short timescale							
	All features		Top 10		Top 5		Top 3	
	Kendall tau	p-value	Kendall tau	p-value	Kendall tau	p-value	Kendall tau	p-value
Label 0	0.1578	0.1136	0.3778	0.1557	-0.4	0.4834	0.3334	1
Label 1	0.1241	0.2134	0.2	0.4843	0.4	0.4834	1	0.3334
Label 2	0.0177	0.8589	0.2	0.4843	-0.4	0.4834	0.3334	1
Label 3	0.1418	0.1550	0.2889	0.2912	0	1	-1	0.3334
Label 4	0.1294	0.1944	0.2889	0.2912	0	1	0.3334	1
	Long timescale							
	All features		Top 10		Top 5		Top 3	
	Kendall tau	p-value	Kendall tau	p-value	Kendall tau	p-value	Kendall tau	p-value
Label 0	0.1188	0.2336	0.1555	0.6006	0.4	0.4834	0.3334	1
Label 1	-0.0372	0.7089	0.0222	1	0	1.	-0.3334	1
Label 2	-0.0957	0.3371	-0.1555	0.6006	-0.2	0.8167	-0.3334	1
Label 3	0.1347	0.1767	0.0667	0.8618	-0.6	0.2334	-0.3334	1
Label 4	-0.0301	0.7625	-0.4222	0.1083	-0.4	0.4834	-0.3334	1

Table D.1: Kendall τ's values and corresponding p-values between novice and expert SHAP ranking for both short and long timescale.

	expert models							
	All features		Top 10		Top 5		Top 3	
	Kendall tau	p-value	Kendall tau	p-value	Kendall tau	p-value	Kendall tau	p-value
Label 0	0.2801	0.0049	0.1111	0.7275	0	1	0.3334	1
Label 1	0.0851	0.3935	0.4667	0.0725	0.8	0.0834	0.3334	1
Label 2	0.0904	0.3646	0.5555	0.0286	1	0.0167	1	0.3334
Label 3	0.0745	0.4553	0.2	0.4843	-0.2	0.8167	0.3334	1
Label 4	0.0106	0.9151	0.2889	0.2912	0.4	0.4834	0.3334	1
	novice models							
	All features		Top 10		Top 5		Top 3	
	Kendall tau	p-value	Kendall tau	p-value	Kendall tau	p-value	Kendall tau	p-value
Label 0	0.0567	0.5695	0.0667	0.8618	0	1	0.3334	1
Label 1	-0.0514	0.6062	-0.2444	0.3807	-0.8	0.0834	-1	0.3334
Label 2	-0.1862	0.0619	-0.4222	0.1083	-0.4	0.4834	-1	0.3334
Label 3	0.1844	0.0645	0.5111	0.0466	0.4	0.4834	1	0.3334
Label 4	0.0284	0.7761	-0.1555	0.6006	-0.4	0.4834	-0.3334	1

Table D.2: Kendall τ's values and corresponding p-values between short and long timescale SHAP ranking for both novice and expert models.

	Short timescale									
	Label 0		Label 1		Label 2		Label 3		Label 4	
	Features	SHAP values	Features	SHAP values	Features	SHAP values	Features	SHAP values	Features	SHAP values
1	herd. velocity	0.0405	herd.targ.0 dist.	0.0447	herd.targ.1 dist.	0.0530	herd.targ.2 dist.	0.0465	herd.targ.3 dist.	0.0454
2	herd.targ.1 dist.	0.0376	targ.0 acceler.	0.0278	targ.1 acceler.	0.0348	targ.2 acceler.	0.0267	targ.3 acceler.	0.0325
3	herd.targ.2 dist.	0.0328	herd.1 targ.0 dist.	0.0211	herd.1 targ.1 dist.	0.0232	herd.1 targ.2 dist.	0.0188	herd.1 targ.3 dist.	0.0249
4	herd.targ.3 dist.	0.0321	herd. velocity	0.0189	herd. velocity	0.0189	herd. velocity	0.0171	herd.targ.2 dist.	0.0203
5	herd.targ.0 dist.	0.0319	herd.targ.3 dist.	0.0161	herd.targ.0 dist.	0.0174	herd.targ.1 dist.	0.0170	herd. velocity	0.0198
6	herd. acceler.	0.0306	herd.targ.1 dist.	0.0159	herd.targ.3 dist.	0.0141	herd.targ.3 dist.	0.0167	herd.targ.1 dist.	0.0168
7	targ.3 acceler.	0.0257	herd.targ.2 dist.	0.0127	herd.targ.2 dist.	0.0139	herd.targ.0 dist.	0.0149	targ.3 goal dist.	0.0151
8	targ.1 acceler.	0.0255	herd. acceler.	0.0124	targ.1 goal dist.	0.0126	targ.1 acceler.	0.0113	herd.targ.0 dist.	0.0145
9	targ.0 acceler.	0.0227	targ.0 goal dist.	0.0109	h acceler.	0.0125	targ.2 goal dist.	0.0099	herd. acceler.	0.0142
10	targ.2 acceler.	0.0227	herd.1 targ.3 dist.	0.0103	herd.1 targ.3 dist.	0.0102	herd. acceler.	0.0099	herd.1 targ.0 dist.	0.0113
	Long timescale									
	Label 0		Label 1		Label 2		Label 3		Label 4	
	Features	SHAP values	Features	SHAP values	Features	SHAP values	Features	SHAP values	Features	SHAP values
1	herd. acceler.	0.0322	herd.targ.0 dist.	0.0356	herd.targ.1 dist.	0.0297	herd.targ.2 dist.	0.0249	herd.targ.3 dist.	0.0340
2	herd.targ.3 dist.	0.0307	herd.1 targ.0 dist.	0.0253	herd.1 targ.1 dist.	0.0212	targ.2 goal dist.	0.0181	herd.1 targ.3 dist.	0.0306
3	herd.targ.0 dist.	0.0285	herd.targ.3 dist.	0.0216	targ.1goal dist.	0.0177	herd.1 targ.2 dist.	0.0160	herd.targ.0 dist.	0.0163
4	herd.targ.1 dist.	0.0263	targ.0 goal dist.	0.0173	herd.targ.0 dist.	0.0176	herd.targ.3 dist.	0.0159	targ.3goal dist.	0.0151
5	targ.3 velocity	0.0257	herd.targ.1 dist.	0.0163	herd.targ.3 dist.	0.0139	herd.targ.1 dist.	0.0143	targ.3 acceler.	0.0144
6	herd.goal dist.	0.0248	herd.1 targ.3 dist.	0.0149	herd. acceler.	0.0124	herd.1 targ.3 dist.	0.0137	herd. acceler.	0.0141
7	herd.1 targ.3 dist.	0.0239	targ.0 direction	0.0135	targ.1 acceler.	0.0108	herd.targ.0 dist.	0.0118	herd.targ.1 dist.	0.0136
8	targ.1 velocity	0.0227	herd.1 targ.2 dist.	0.0134	herd.1 targ.3 dist.	0.0107	targ.2 acceler.	0.0113	herd.1 targ.0 dist.	0.0130
9	herd. velocity	0.0216	herd.goal dist.	0.0115	targ.1 direction	0.0102	targ.2 direction	0.0109	herd.targ.2 dist.	0.0125
10	targ.0 velocity	0.0200	targ.0 acceler.	0.0100	herd.targ.2 dist.	0.0098	herd. goal dist.	0.0106	targ.3 velocity	0.0107

Table D.3: Top 10 ranked features and corresponding SHAP values for each class predicted by the model trained on *novice* pairs with a sequence T_{seq} = 1s

	Short timescale									
	Label 0		Label 1		Label 2		Label 3		Label 4	
	Features	SHAP values	Features	SHAP values	Features	SHAP values	Features	SHAP values	Features	SHAP values
1	herd.targ.2 dist.	0.0399	herd.targ.0 dist.	0.0178	herd.1 targ.1 dist.	0.0174	herd.targ.2 dist.	0.0148	herd.1 targ.3 dist.	0.0267
2	herd.targ.0 dist.	0.0396	herd.targ.3 dist.	0.0158	herd.targ.1 dist.	0.0169	herd.targ.1 dist.	0.0095	herd.targ.3 dist.	0.0231
3	herd.targ.3 dist.	0.0337	herd.1 targ.0 dist.	0.0129	herd.targ.0 dist.	0.0153	herd.targ.3 dist.	0.0094	herd.targ.2 dist.	0.0211
4	herd.1 targ.3 dist.	0.0333	herd.targ.2 dist.	0.0117	herd.targ.3 dist.	0.0133	herd.targ.0 dist.	0.0094	herd.targ.0 dist.	0.0196
5	herd.targ.1 dist.	0.0277	targ.0 acceler.	0.0114	herd.targ.2 dist.	0.0126	herd.1 targ.2 dist.	0.0087	herd.targ.1 dist.	0.0122
6	herd.1 targ.1 dist.	0.0273	herd.targ.1 dist.	0.0106	targ.1 acceler.	0.0112	targ.2 acceler.	0.0080	targ.3 goal dist.	0.0104
7	hgoal dist.	0.0244	targ.0 goal dist.	0.0074	targ.1 direction	0.0100	targ.2 velocity	0.0058	targ.3 acceler.	0.0098
8	targ.1 direction	0.0230	herd. goal dist.	0.0071	targ.1 velocity	0.0087	targ.1 direction	0.0046	targ.3 velocity	0.0090
9	targ.1 velocity	0.0212	herd. velocity	0.0061	herd.goal dist.	0.0074	targ.1 velocity	0.0042	herd. goal dist.	0.0084
10	targ.2 velocity	0.0197	herd. acceler.	0.0057	targ.2 velocity	0.0068	herd. velocity	0.0042	herd. velocity	0.0081
	Long timescale									
	Label 0		Label 1		Label 2		Label 3		Label 4	
	Features	SHAP values	Features	SHAP values	Features	SHAP values	Features	SHAP values	Features	SHAP values
	Features	SHAP values	Features	SHAP values	Features	SHAP values	Features	SHAP values	Features	SHAP values
1	herd.1 targ.1 dist.	0.0351	herd.1 targ.0 dist.	0.0159	herd.1 targ.1 dist.	0.0233	herd.1 targ.2 dist.	0.0184	herd.1targ.3 dist.	0.0242
2	herd.1 targ.3 dist.	0.0322	herd.targ.3 dist.	0.0092	herd.targ.2 dist.	0.0137	herd.targ.0 dist.	0.0121	herd.targ.2 dist.	0.0126
3	herd. targ.2 dist.	0.0279	herd.targ.0 dist.	0.0091	herd.targ.0 dist.	0.0106	herd.targ.3 dist.	0.0106	targ.3 goal dist.	0.0100
4	herd. targ.0 dist.	0.0250	herd.targ.2 dist.	0.0086	targ.1 goal dist.	0.0098	herd.targ.1 dist.	0.0091	herd.targ.0 dist.	0.0098
5	targ.2 velocity	0.0242	targ.2 direction	0.0074	herd.targ.3 dist.	0.0094	herd.targ.2 dist.	0.0097	herd.1targ.1 dist.	0.0091
6	herd.targ.3 dist.	0.0230	targ.2 velocity	0.0074	herd.targ.1 dist.	0.0093	targ.2 velocity	0.0088	herd.targ.3 dist.	0.0076
7	herd. goal dist.	0.0223	targ.0 goal dist.	0.0072	herd. goal dist.	0.0084	herd.1 targ.1 dist.	0.0087	herd.targ.1 dist.	0.0072
8	herd.herd.1 dist.	0.0197	herd.1targ.1 dist.	0.0069	targ.1 velocity	0.0069	targ.1 velocity	0.0086	herd.herd.1 dist.	0.0071
9	targ.2 direction	0.0196	herd.targ.1 dist.	0.0066	herd.1 targ.3 dist.	0.0067	targ.2 goal dist.	0.0084	targ.2 direction	0.0067
10	targ.1 velocity	0.0195	herd.1targ.3 dist.	0.0056	targ.2 velocity	0.0062	herd.herd.1 dist.	0.0074	herd.1targ.0 dist.	0.0063

Table D.4: Top 10 ranked features and corresponding SHAP values for each class predicted by the model trained on *expert* pairs with a sequence T_{seq} = 1s

APPENDIX E

SUPPLEMENTARY INFORMATION TO UNDERSTANDING TEAM HUMAN DECISION

In this appendix we provide further details on the prediction performance and explanation of the human target selection policies discussed in Chapter 6.

E.1 Prediction performance for fixed and variable prediction horizons

In chapter 6, target selection strategies of team members tasked to corral N_T target agents in a large-scale slow-paced multi-agent environment were modelled by training Long-Short-Term-Memory neural networks (LSTM_{NN}) through supervised machine learning. We fed the LSTM_{NN} with sequences of system evolution covering intervals of $t_f - t_i = T_{seq}$ and selected the correct prediction horizon at a fixed ($t_{f+\tau_{hor}}$) and variable ($\tau_{hor} \in (\tau_{hor,min}, \tau_{hor,max})$) timescales. Prediction performance for $T_{seq} = 2$ s are reported in Table E.1, showing performance comparable to the case of $T_{seq} = 5$ s discussed in chapter 6.

E.2 Kendall rank correlation coefficients

The ordinal association of SHAP value rankings (the first top 10 reported in Tables E.3-E.4) was computed using the Kendall rank correlation coefficient (Kendall's τ,[98]) for subgroups of N top ranked input features. Table E.2 includes the Kendall coefficients, and associated p-values, as a function of the visibility condition.

	Accuracy	Precision	Recall	F1 score
Shorter timescale				
Full visibility	98.36 ± 0.3	98.28 ± 0.34	98.52 ± 0.28	98.38 ± 0.31
Partial visibility	97.77 ± 0.32	97.87 ± 0.31	97.79 ± 0.32	97.82 ± 0.31
Longer timescale				
Full visibility	97.72 ± 0.32	97.63 ± 0.33	97.89 ± 0.31	97.74 ± 0.32
Partial visibility	98 ± 0.29	98.12 ± 0.26	98 ± 0.29	98.05 ± 0.28
Variable timescale				
Full visibility	94.44 ± 0.47	94.60 ± 0.47	95.04 ± 0.45	94.79 ± 0.45
Partial visibility	94.56 ± 0.48	94.90 ± 0.44	95.15 ± 0.47	94.96 ± 0.45

Table E.1: Average performance [%] of the multi-label predictor trained on time-series of length T_{seq} = 2 s and tested on 100 sets of N_{test}= 2000 samples for shorter (T_{hor} = 2s) and longer (T_{hor} = 5.2s) and variable timescales.

	All features		Top 20		Top 10		Top 5	
	Kendall tau	p-value	Kendall tau	p-value	Kendall tau	p-value	Kendall tau	p-value
Label 0	0.12	0.0120	0.06	0.73	0.22	0.48	0.33	0.75
Label 1	0.22	0.0003	0.27	0.12	0.61	0.02	0.67	0.33
Label 2	0.2	0.0013	0.34	0.05	0.39	0.18	1	0.08
Label 3	0.13	0.0325	-0.37	0.03	-0.17	0.61	0	1
Label 4	0.16	0.0095	0.17	0.33	0.22	0.48	0.67	0.33
Label 5	0.29	0	0.18	0.29	0.28	0.36	0.33	0.75
Label 6	0.12	0.0580	0.16	0.37	0.17	0.61	-0.33	0.75
Label 7	0.23	0.0002	0.1	0.58	0	1	0	1
Label 8	0.06	0.3681	-0.37	0.03	-0.33	0.26	0	1
Label 9	0.19	0.0023	-0.12	0.49	-0.17	0.61	-1.00	0.08

Table E.2: Kendall τ's values and corresponding p-values between full and partial visibility SHAP ranking.

E.3 SHAP value tables

A detailed summary of SHAP feature values for each LSTM$_{NN}$ model of target switching decision, discussed in chapter 6, is provided in Tables E.3-E.4.

	Label 0		Label 1		Label 2		Label 3		Label 4	
	Features	SHAP	Features	SHAP	Features	SHAP	Features	SHAP	Features	SHAP
1	herd. ang. vel	0.0346	herd.targ.0 dist	0.0070	herd.targ.1 dist	0.01193	herd.targ.2 dist	0.0114	herd.targ.3 dist	0.0104
2	herd.targ.6 dist	0.0128	herd.targ.0 angle	0.0063	herd.0 targ.1 dist	0.0070	herd.0 targ.2 dist	0.0071	herd. ang. vel	0.0097
3	herd. goal dist	0.0128	herd.targ.6 dist	0.0058	herd. ang. vel	0.0067	herd. ang. vel	0.0051	herd.0 targ.3 dist	0.0054
4	targ.8 acceler.	0.0119	herd. ang. vel	0.0058	herd.1 targ.1 dist	0.0053	herd.targ.4 dist	0.0045	herd.0 targ.5 angle	0.0047
5	targ.6 acceler.	0.0116	targ.8 acceler.	0.0058	herd.0 targ.3 dist	0.0047	herd.1 targ.1 dist	0.0045	herd.1 targ.7 angle	0.0047
6	herd.0 targ.5 dist	0.0116	herd.0 targ.5 dist	0.0058	targ.6 acceler.	0.0044	targ.2 acceler.	0.0042	targ.7 acceler.	0.0045
7	herd.0 targ.5 angle	0.0113	targ.0 acceler.	0.0058	herd.0 targ.5 angle	0.0042	herd.0 targ.7 angle	0.0042	herd.1 targ.3 dist	0.0044
8	herd.targ.0 angle	0.0103	targ.0 vel	0.0050	targ.1 acceler.	0.0041	targ.2 goal dist	0.0041	targ.3 vel	0.0044
9	herd.0 targ.1 angle	0.012	targ.7 goal dist	0.0046	herd.0 targ.1 angle	0.0040	targ.8 goal dist	0.0041	targ.3 acceler.	0.0044
10	targ.0 acceler.	0.0098	herd.1 targ.7 angle	0.0043	herd.1 targ.1 angle	0.0039	herd.targ.5 dist	0.0040	herd.0 targ.6 dist	0.0043
	Label 5		Label 6		Label 7		Label 8		Label 9	
	Features	SHAP	Features	SHAP	Features	SHAP	Features	SHAP	Features	SHAP
1	herd.targ.4 dist	0.0108	herd.targ.5 dist	0.0086	herd.targ.6 dist	0.0101	herd. ang. vel	0.0091	herd.targ.8 dist	0.0145
2	herd. ang. vel	0.0075	herd. ang. vel	0.0069	targ.6 acceler.	0.0071	herd.targ.7 dist	0.0076	targ.8 acceler.	0.0085
3	herd.targ.7 dist	0.0057	herd.1 targ.5 dist	0.0056	herd. ang. vel	0.0065	herd.1 targ.7 dist	0.0063	herd.1 targ.8 dist	0.0079
4	targ.4 acceler.	0.0054	targ.5 acceler.	0.0048	herd.targ.8 dist	0.0054	targ.7 acceler.	0.0053	herd.0 targ.8 dist	0.0078
5	herd.1 targ.4 dist	0.0049	targ.8 acceler.	0.0043	herd.1 targ.6 dist	0.0053	herd.0 targ.7 dist	0.0049	herd. ang. vel	0.0075
6	herd.0 targ.4 dist	0.0044	herd.0 targ.4 dist	0.0043	herd.0 targ.6 dist	0.0052	targ.7 goal dist	0.0048	herd. goal dist	0.0064
7	targ.4 goal dist	0.0044	herd.targ.3 dist	0.0042	herd.1 targ.7 rel angle	0.0042	herd.targ.4 dist	0.0044	herd.targ.5 dist	0.0056
8	herd.0 targ.0 dist	0.0041	herd.0 goal angle	0.0040	targ.6 direction	0.0040	herd.0 targ.5 rel angle	0.0043	herd.1 targ.7 angle	0.0056
9	herd.1 targ.7 dist	0.0041	herd.1 targ.3 dist	0.0039	targ.6 goal dist	0.0040	herd.0 targ.1 angle	0.0041	targ.8 direction	0.0055
10	herd.1 targ.4 angle	0.0040	herd. goal dist	0.0038	herd.1 targ.6 angle	0.0040	herd.targ.6 dist	0.0040	herd.1 targ.5 dist	0.0053

Table E.3: Top 10 ranked features and corresponding SHAP values for each class of N_{test} = 6000 full visibility samples predicted by the model trained with a sequence T_{seq} = 5 s and variable prediction horizon ($t_{\tau_{hor,min}}, t_{\tau_{hor,max}}$ = (0.2s, 2s)).

	Label 0		Label 1		Label 2		Label 3		Label 4	
	Features	SHAP	Features	SHAP	Features	SHAP	Features	SHAP	Features	SHAP
1	herd.ang. vel	0.0294	herd. targ.0 dist	0.0152	herd. targ.1 dist	0.0082	herd. targ.2 dist	0.0118	herd. targ.3 dist	0.0121
2	targ.8 acceler.	0.0145	herd.0 targ.0 dist	0.0099	herd.0 targ.1 dist	0.0060	herd. targ.0 dist	0.0074	herd.0 targ.2 rel angle	0.0057
3	targ.1 goal angle	0.0145	herd.1 targ.0 dist	0.0089	herd.ang. vel	0.0047	herd.1 targ.2 dist	0.0067	herd. targ.0 dist	0.0049
4	herd.goal dist	0.0142	herd.ang. vel	0.0074	herd.1 targ.7 rel angle	0.0043	herd.0 targ.3 rel angle	0.0065	herd.0 targ.3 dist	0.0047
5	targ.2 acceler.	0.0135	herd.0 targ.3 rel angle	0.0074	herd. targ.0 dist	0.0043	targ.2 goal dist	0.0062	herd.1 targ.3 dist	0.0046
6	targ.0 acceler.	0.0131	herd.0 targ.2 rel angle	0.0069	herd.0 targ.0 dist	0.0042	herd.ang. vel	0.0060	herd.0 targ.6 dist	0.0044
7	targ.4 acceler.	0.0129	herd. targ.2 dist	0.0067	herd.0 targ.8 dist	0.0039	targ.1 goal angle	0.0055	herd.0 targ.0 dist	0.0044
8	herd. targ.8 dist	0.0129	targ.0 acceler.	0.0067	herd. targ.6 dist	0.0035	targ.2 goal angle	0.0054	herd.1goal angle	0.0039
9	herd. targ.6 dist	0.0129	targ.0 goal dist	0.0066	herd.1 targ.2 dist	0.0034	herd.0 targ.2 rel angle	0.0053	targ.3 acceler.	0.0038
10	targ.8 vel_r	0.0127	herd.0 targ.1 dist	0.0062	herd. targ.4 angle	0.0034	herd.0 targ.0 dist	0.0048	herd.ang. vel	0.0038
	Label 5		Label 6		Label 7		Label 8		Label 9	
	Features	SHAP	Features	SHAP	Features	SHAP	Features	SHAP	Features	SHAP
1	herd. targ.4 dist	0.0080	herd. targ.5 dist	0.0115	herd. targ.6 dist	0.0096	herd. targ.7 dist	0.0121	herd. targ.8 dist	0.0101
2	h ang. vel	0.0068	herd.0 targ.5 dist	0.0081	herd.0 targ.6 dist	0.0060	herd.1 targ.7 dist	0.0087	h ang. vel	0.0066
3	herd. targ.8 dist	0.0057	herd. targ.8 dist	0.0078	h ang. vel	0.0058	h ang. vel	0.0071	targ.8 acceler.	0.0065
4	herd.0 targ.6 dist	0.0052	h ang. vel	0.0075	herd.1 targ.7 dist	0.0049	herd. targ.1 dist	0.0055	targ.8 vel_r	0.0051
5	herd.h1 rel angle	0.0051	herd. targ.6 dist	0.0065	targ.6 acceler.	0.0043	herd. targ.8 angle	0.0054	herd.0 targ.8 dist	0.0047
6	targ.4 acceler.	0.0050	herd.1 targ.5 dist	0.0065	herd. targ.1 dist	0.0042	herd.0 targ.2 rel angle	0.0053	targ.4 acceler.	0.0046
7	targ.1 goal dist	0.0049	targ.5 goal dist	0.0062	herd. targ.8 dist	0.0039	targ.1 goal angle	0.0052	herd. targ.0 dist	0.0045
8	herd.1 targ.7 rel angle	0.0049	herd. targ.0 dist	0.0059	herd.1 targ.2 dist	0.0038	targ.8 acceler.	0.0051	herd.1 targ.7 dist	0.0044
9	herd.1 targ.4 dist	0.0048	herd.0 targ.3 rel angle	0.0058	targ.0 acceler.	0.0037	targ.0 goal dist	0.0049	herd. targ.6 dist	0.0044
10	herd.0 targ.2 rel angle	0.0047	herd.0 targ.0 dist	0.0057	herd. targ.7 dist	0.0037	targ.7 acceler.	0.0049	targ.8 dir_motion	0.0043

Table E.4: Top 10 ranked features and corresponding SHAP values for each class of N_{test} = 6000 partial visibility samples predicted by the model trained with a sequence T_{seq} = 5 s and variable prediction horizon ($t_{\tau_{hor,min}}, t_{\tau_{hor,max}}$ = (0.2s, 2s)).

BIBLIOGRAPHY

[1] P. AJIT, *Prediction of employee turnover in organizations using machine learning algorithms*, algorithms, 4 (2016), p. C5.

[2] F. ALDERISIO, G. FIORE, R. N. SALESSE, B. G. BARDY, AND M. DI BERNARDO, *Interaction patterns and individual dynamics shape the way we move in synchrony*, Scientific reports, 7 (2017), pp. 1–10.

[3] S. ALHAGRY, A. A. FAHMY, AND R. A. EL-KHORIBI, *Emotion recognition based on eeg using lstm recurrent neural network*, Emotion, 8 (2017), pp. 355–358.

[4] Y. ALUFAISAN, L. R. MARUSICH, J. Z. BAKDASH, Y. ZHOU, AND M. KANTARCIOGLU, *Does explainable artificial intelligence improve human decision-making?*, in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35,8, 2021, pp. 6618–6626.

[5] D. AMODEI, S. ANANTHANARAYANAN, R. ANUBHAI, J. BAI, E. BATTENBERG, C. CASE, J. CASPER, B. CATANZARO, Q. CHENG, G. CHEN, ET AL., *Deep speech 2: End-to-end speech recognition in english and mandarin*, in International conference on machine learning, PMLR, 2016, pp. 173–182.

[6] E. ANGELINI, G. DI TOLLO, AND A. ROLI, *A neural network approach for credit risk evaluation*, The quarterly review of economics and finance, 48 (2008), pp. 733–755.

[7] D. ARAUJO, K. DAVIDS, AND R. HRISTOVSKI, *The ecological dynamics of decision making in sport*, Psychology of sport and exercise, 7 (2006), pp. 653–676.

[8] F. AULETTA, M. DI BERNARDO, AND M. J. RICHARDSON, *Human-inspired strategies to solve complex joint tasks in multi agent systems*, IFAC-PapersOnLine, 54 (2021), pp. 105–110.

[9] F. AULETTA, D. FIORE, M. J. RICHARDSON, AND M. DI BERNARDO, *Herding stochastic autonomous agents via local control rules and online target selection strategies*, Autonomous Robots, (2022), pp. 1–13.

[10] I. A. BASHEER AND M. HAJMEER, *Artificial neural networks: fundamentals, computing, design, and application*, Journal of microbiological methods, 43 (2000), pp. 3–31.

[11] B. BENNET AND M. TRAFANKOWSKI, *A Comparative Investigation of Herding Algorithms*, in Proc. Symp. on Understanding and Modelling Collective Phenomena (UMoCoP), 2012, pp. 33–38.

[12] G. B. BERIKOL, O. YILDIZ, AND İ. T. ÖZCAN, *Diagnosis of acute coronary syndrome with a support vector machine*, Journal of medical systems, 40 (2016), p. 84.

[13] A. A. BERRYMAN, *The orgins and evolution of predator-prey theory*, Ecology, 73 (1992), pp. 1530–1535.

[14] E. BICHO, W. ERLHAGEN, L. LOURO, AND E. C. E SILVA, *Neuro-cognitive mechanisms of decision making in joint action: A human–robot interaction study*, Human movement science, 30 (2011), pp. 846–868.

[15] S. BORDT AND U. VON LUXBURG, *When humans and machines make joint decisions: A non-symmetric bandit model*, 2020.

[16] A. BOUKERCHE AND J. WANG, *Machine learning-based traffic prediction models for intelligent transportation systems*, Computer Networks, 181 (2020), p. 107530.

[17] L. BREIMAN, *Random forests*, Machine learning, 45 (2001), pp. 5–32.

[18] M. BÜRGER, G. NOTARSTEFANO, F. ALLGÖWER, AND F. BULLO, *A distributed simplex algorithm and the multi-agent assignment problem*, in Proc. of the American Control Conference, 2011, pp. 2639–2644.

[19] M. CARROLL, R. SHAH, M. K. HO, T. GRIFFITHS, S. SESHIA, P. ABBEEL, AND A. DRAGAN, *On the utility of learning about humans for human-ai coordination*, Advances in Neural Information Processing Systems, 32 (2019), pp. 5174–5185.

[20] R. CARUANA AND A. NICULESCU-MIZIL, *An empirical comparison of supervised learning algorithms*, in Proceedings of the 23rd international conference on Machine learning, 2006, pp. 161–168.

[21] V. K. R. CHIMMULA AND L. ZHANG, *Time series forecasting of covid-19 transmission in canada using lstm networks*, Chaos, Solitons & Fractals, 135 (2020), p. 109864.

[22] V. S. CHIPADE, V. S. A. MARELLA, AND D. PANAGOU, *Aerial swarm defense by stringnet herding: Theory and experiments*, Frontiers in Robotics and AI, 8 (2021), p. 640446.

[23] V. S. CHIPADE AND D. PANAGOU, *Herding an adversarial swarm in an obstacle environment*, in Proc. of the IEEE Conference on Decision and Control, IEEE, 2019, pp. 3685–3690.

[24] W. CHRISTENSEN, J. SUTTON, AND K. BICKNELL, *Memory systems and the control of skilled action*, Philosophical Psychology, 32 (2019), pp. 692–718.

[25] W. CHRISTENSEN, J. SUTTON, AND D. J. MCILWAIN, *Cognition in skilled action: Meshed control and the varieties of skill experience*, Mind & Language, 31 (2016), pp. 37–66.

[26] T. H. CHUNG, G. A. HOLLINGER, AND V. ISLER, *Search and pursuit-evasion in mobile robotics*, Autonomous Robots, 31 (2011), p. 299.

[27] N. R. CLAYTON AND H. ABBASS, *Machine teaching in hierarchical genetic reinforcement learning: Curriculum design of reward functions for swarm shepherding*, in 2019 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2019, pp. 1259–1266.

[28] S. CONVERSE, J. CANNON-BOWERS, AND E. SALAS, *Shared mental models in expert team decision making*, Individual and group decision making: Current issues, 221 (1993), pp. 221–46.

[29] C. CORTES AND V. VAPNIK, *Support-vector networks*, Machine learning, 20 (1995), pp. 273–297.

[30] I. D. COUZIN, J. KRAUSE, R. JAMES, G. D. RUXTON, AND N. R. FRANKS, *Collective memory and spatial sorting in animal groups*, Journal of theoretical biology, 218 (2002), pp. 1–11.

[31] N. CRISTIANINI, J. SHAWE-TAYLOR, ET AL., *An introduction to support vector machines and other kernel-based learning methods*, Cambridge university press, 2000.

[32] A. CURIONI, G. KNOBLICH, N. SEBANZ, A. GOSWAMI, AND P. VADAKKEPAT, *Joint action in humans: A model for human-robot interactions*, Humanoid Robotics: A Reference, eds Goswami A, Vadakkepat P (Springer, Dordrecht, The Netherlands), (2019), pp. 2149–2167.

[33] X. DA, R. HARTLEY, AND J. W. GRIZZLE, *Supervised learning for stabilizing underactuated bipedal robot locomotion, with outdoor experiments on the wave field*, in 2017 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2017, pp. 3476–3483.

[34] R. DALE, R. FUSAROLI, N. D. DURAN, AND D. C. RICHARDSON, *The self-organization of human interaction*, in Psychology of learning and motivation, vol. 59, Elsevier, 2013, pp. 43–95.

[35] K. DAVIDS, D. ARAÚJO, L. SEIFERT, AND D. ORTH, *Expert performance in sport: An ecological dynamics perspective*, in Routledge handbook of sport expertise, Routledge, 2015, pp. 130–144.

[36] P. DELELLIS, G. POLVERINO, G. USTUNER, N. ABAID, S. MACRÌ, E. M. BOLLT, AND M. PORFIRI, *Collective behaviour across animal species*, Scientific reports, 4 (2014), pp. 1–6.

[37] L. DENG AND X. LI, *Machine learning paradigms for speech recognition: An overview*, IEEE Transactions on Audio, Speech, and Language Processing, 21 (2013), pp. 1060–1089.

[38] P. DEPTULA, Z. I. BELL, F. M. ZEGERS, R. A. TRA, AND W. E. DIXON, *Single Agent Indirect Herding via Approximate Dynamic Programming*, in IEEE Conference on Decision and Control (CDC), 2018, pp. 7136–7141.

[39] J. T. EMLEN, *Flocking behavior in birds*, The Auk, 69 (1952), pp. 160–170.

[40] K. A. ERICSSON AND N. CHARNESS, *Expert performance: Its structure and acquisition.*, American psychologist, 49 (1994), p. 725.

[41] W. ERLHAGEN AND E. BICHO, *The dynamic neural field approach to cognitive robotics*, Journal of neural engineering, 3 (2006), p. R36.

[42] M. EVERED, P. BURLING, AND M. TROTTER, *An Investigation of Predator Response in Robotic Herding of Sheep*, in International Conference on Intelligent Agriculture, vol. 63, Singapore, 2014.

[43] X. FAN, M. MCNEESE, AND J. YEN, *Ndm-based cognitive agents for supporting decision-making teams*, Human–Computer Interaction, 25 (2010), pp. 195–234.

[44] A. GEE AND H. ABBASS, *Transparent machine education of neural networks for swarm shepherding using curriculum design*, in 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, 2019, pp. 1–8.

[45] F. A. GERS, D. ECK, AND J. SCHMIDHUBER, *Applying lstm to time series predictable through time-window approaches*, in Neural Nets WIRN Vietri-01, Springer, 2002, pp. 193–200.

[46] A. GOLESTANI, M. MASLI, N. S. SHAMI, J. JONES, A. MENON, AND J. MONDAL, *Real-time prediction of employee engagement using social media and text mining*, in 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2018, pp. 1383–1387.

[47] I. GOODFELLOW, Y. BENGIO, AND A. COURVILLE, *Machine learning basics*, Deep learning, 1 (2016), pp. 98–164.

[48] H. HAGRAS, *Toward human-understandable, explainable ai*, Computer, 51 (2018), pp. 28–36.

[49] H. HAKEN, J. A. S. KELSO, AND H. BUNZ, *A theoretical model of phase transitions in human hand movements*, Biological Cybernetics, 51 (1985), pp. 347–356.

[50] L. HAMADEH, S. IMRAN, M. BENCSIK, G. R. SHARPE, M. A. JOHNSON, AND D. J. FAIRHURST, *Machine learning analysis for quantitative discrimination of dried blood droplets*, Scientific reports, 10 (2020), pp. 1–13.

[51] W. HAMILTON, *Geometry for the selfish herd*, Journal of Theoretical Biology, 31 (1971), pp. 295–311.

[52] M. HAQUE, A. RAHMANI, AND M. EGERSTEDT, *A hybrid, multi-agent model of foraging bottlenose dolphins*, IFAC Proceedings Volumes, 42 (2009), pp. 262–267.
3rd IFAC Conference on Analysis and Design of Hybrid Systems.

[53] M. A. HAQUE, A. R. RAHMANI, AND M. B. EGERSTEDT, *Biologically inspired confinement of multi-robot systems*, International Journal of Bio-Inspired Computation, 3 (2011), pp. 213–224.

[54] S. J. HARRISON AND N. STERGIOU, *Complex adaptive behavior and dexterous action*, Nonlinear dynamics, psychology, and life sciences, 19 (2015), p. 345.

[55] B. D. HASSARD, B. HASSARD, N. D. KAZARINOFF, Y.-H. WAN, AND Y. W. WAN, *Theory and applications of Hopf bifurcation*, vol. 41, CUP Archive, 1981.

[56] T. HASTIE, R. TIBSHIRANI, AND J. FRIEDMAN, *Overview of supervised learning*, in The elements of statistical learning, Springer, 2009, pp. 9–41.

[57] K. HE, X. ZHANG, S. REN, AND J. SUN, *Deep residual learning for image recognition*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[58] D. J. HIGHAM, *An algorithmic introduction to numerical simulation of stochastic differential equations*, SIAM Review, 43 (2001), pp. 525–546.

[59] T. HIMBERG AND M. THOMPSON, *Group synchronization of coordinated movements in a cross-cultural choir workshop*, in ESCOM 2009: 7th Triennial Conference of European Society for the Cognitive Sciences of Music, 2009.

[60] G. E. HINTON, N. SRIVASTAVA, A. KRIZHEVSKY, I. SUTSKEVER, AND R. R. SALAKHUTDINOV, *Improving neural networks by preventing co-adaptation of feature detectors*, 2012.

[61] S. HOCHREITER AND J. SCHMIDHUBER, *Long short-term memory*, Neural Computation, 9 (1997), pp. 1735–1780.

[62] J. HOOK, S. EL-SEDKY, V. DE SILVA, AND A. KONDOZ, *Learning data-driven decision-making policies in multi-agent environments for autonomous systems*, Cognitive Systems Research, 65 (2021), pp. 40–49.

[63] R. J. HUTTON AND G. KLEIN, *Expert decision making*, Systems Engineering: The Journal of The International Council on Systems Engineering, 2 (1999), pp. 32–45.

[64] T. IQBAL AND L. D. RIEK, *Human-robot teaming: Approaches from joint action and dynamical systems*, Humanoid robotics: A reference, 1 (2019), pp. 2293–2312.

[65] J.-M. LIEN, O. BAYAZIT, R. SOWELL, S. RODRIGUEZ, AND N. AMATO, *Shepherding behaviors*, in IEEE International Conference on Robotics and Automation, 2004, pp. 4159–4164.

[66] J.-M. LIEN AND E. PRATT, *Interactive planning for shepherd motion*, in AAAI Spring Symposium: Agents that Learn from Human Teachers, 2009.

[67] J.-M. LIEN, S. RODRIGUEZ, J. MALRIC, AND N. AMATO, *Shepherding Behaviors with Multiple Shepherds*, in IEEE International Conference on Robotics and Automation, 2005, pp. 3402–3407.

[68] D. M. JACOBS AND C. F. MICHAELS, *Direct learning*, Ecological psychology, 19 (2007), pp. 321–349.

[69] P. KACHROO, S. SHEDIED, J. BAY, AND H. VANLANDINGHAM, *Dynamic programming solution for a class of pursuit evasion problems: the herding problem*, IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), 31 (2001), pp. 35–41.

[70] J. A. S. KELSO, *Dynamic patterns: The self-organization of brain and behavior*, Cambridge, MA: MIT Press, 1995.

[71] H. K. KHALIL, *Nonlinear Systems*, NJ: Prentice Hall, 3rd edition. upper saddle river ed., 2002.

[72] A. J. KING, A. M. WILSON, S. D. WILSHIN, J. LOWE, H. HADDADI, S. HAILES, AND A. J. MORTON, *Selfish-herd behaviour of sheep under threat*, Current Biology, 22 (2012), pp. R561–R562.

[73] G. KLEIN, *The recognition-primed decision (rpd) model: Looking back, looking forward*, 1997.

[74] ——, *Naturalistic decision making*, Human factors, 50 (2008), pp. 456–460.

[75] G. A. KLEIN, *A recognition-primed decision (rpd) model of rapid decision making*, Decision making in action: Models and methods, 5 (1993), pp. 138–147.

[76] J. KLEINBERG, H. LAKKARAJU, J. LESKOVEC, J. LUDWIG, AND S. MULLAINATHAN, *Human decisions and machine predictions*, The quarterly journal of economics, 133 (2018), pp. 237–293.

[77] J. KRAUSE, V. GULSHAN, E. RAHIMY, P. KARTH, K. WIDNER, G. S. CORRADO, L. PENG, AND D. R. WEBSTER, *Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy*, Ophthalmology, 125 (2018), pp. 1264–1272.

[78] M. KRSTIC, P. V. KOKOTOVIC, AND I. KANELLAKOPOULOS, *Nonlinear and Adaptive Control Design*, John Wiley & Sons, Inc., New York, NY, USA, 1st ed., 1995.

[79] W. H. KRUSKAL AND W. A. WALLIS, *Use of ranks in one-criterion variance analysis*, Journal of the American statistical Association, 47 (1952), pp. 583–621.

[80] E. LAKSHIKA, M. BARLOW, AND A. EASTON, *Co-evolving semi-competitive interactions of sheepdog herding behaviors utilizing a simple rule-based multi agent framework*, in 2013 IEEE Symposium on Artificial Life (ALife), April 2013, pp. 82–89.

[81] M. LAMB, P. NALEPKA, R. W. KALLEN, T. LORENZ, S. J. HARRISON, A. A. MINAI, AND M. J. RICHARDSON, *A hierarchical behavioral dynamic approach for naturally adaptive human-agent pick-and-place interactions*, Complexity, 2019 (2019).

[82] P. LANGLEY, *Elements of machine learning*, Morgan Kaufmann, 1996.

[83] M. LE MENN, C. BOSSARD, B. TRAVASSOS, R. DUARTE, AND G. KERMARREC, *Handball goal-keeper intuitive decision-making: A naturalistic case study*, Journal of human kinetics, 70 (2019), p. 297.

[84] W. LEE AND D. KIM, *Autonomous shepherding behaviors of multiple target steering robots*, Sensors, 17 (2017), p. 2729.

[85] N. F. LEPORA AND G. PEZZULO, *Embodied choice: How action influences perceptual decision making*, PLoS computational biology, 11 (2015), p. e1004110.

[86] R. A. LICITRA, Z. I. BELL, AND W. E. DIXON, *Single-agent indirect herding of multiple targets with uncertain dynamics*, IEEE Trans. on Robotics, 35 (2019), pp. 847–860.

[87] R. A. LICITRA, Z. I. BELL, E. A. DOUCETTE, AND W. E. DIXON, *Single Agent Indirect Herding of Multiple Targets: A Switched Adaptive Control Approach*, IEEE Control Systems Letters, 2 (2018), pp. 127–132.

[88] R. A. LICITRA, Z. D. HUTCHESON, E. A. DOUCETTE, AND W. E. DIXON, *Single Agent Herding of n-Agents: A Switched Systems Approach*, IFAC-PapersOnLine, 50 (2017), pp. 14374–14379.

[89] W.-Y. LIN, Y.-H. HU, AND C.-F. TSAI, *Machine learning in financial crisis prediction: a survey*, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42 (2011), pp. 421–436.

[90] N. K. LONG, K. SAMMUT, D. SGARIOTO, M. GARRATT, AND H. A. ABBASS, *A comprehensive review of shepherding as a bio-inspired swarm-robotics guidance approach*, IEEE Trans on Emerging Topics in Computational Intelligence, 4 (2020), pp. 523–537.

[91] S. M. LUNDBERG, G. ERION, H. CHEN, A. DEGRAVE, J. M. PRUTKIN, B. NAIR, R. KATZ, J. HIMMELFARB, N. BANSAL, AND S.-I. LEE, *From local explanations to global understanding with explainable ai for trees*, Nature machine intelligence, 2 (2020), pp. 56–67.

[92] S. M. LUNDBERG AND S.-I. LEE, *A unified approach to interpreting model predictions*, Advances in Neural Information Processing Systems, 30 (2017), pp. 4765–4774.

[93] S. M. LUNDBERG, B. NAIR, M. S. VAVILALA, M. HORIBE, M. J. EISSES, T. ADAMS, D. E. LISTON, D. K.-W. LOW, S.-F. NEWMAN, J. KIM, ET AL., *Explainable machine-learning predictions for the prevention of hypoxaemia during surgery*, Nature biomedical engineering, 2 (2018), pp. 749–760.

[94] Y. MA, R. K. K. YUEN, AND E. W. M. LEE, *Effective leadership for crowd evacuation*, Physica A: Statistical Mechanics and its Applications, 450 (2016), pp. 333–341.

[95] A. J. MARASCO, S. N. GIVIGI, C. A. RABBATH, AND A. BEAULIEU, *Dynamic encirclement of a moving target using decentralized nonlinear Model Predictive Control*, in 2013 American Control Conference, Washington, DC, 2013, IEEE, pp. 3960–3966.

[96] J. MARTENS, *Doing Things Together: A Theory of Skillful Joint Action*, vol. 41, Walter de Gruyter GmbH & Co KG, 2020.

[97] J. H. MARTENS, *Habit and skill in the domain of joint action*, Topoi, 1 (2020), pp. 1–13.

[98] A. I. MCLEOD, *Kendall rank correlation and Mann-Kendall trend test*, Western Univ., 2005.

[99] T. MIKI AND T. NAKAMURA, *An Effective Simple Shepherding Algorithm Suitable for Implementation to a Multi-Mobile Robot System*, in First International Conference on Innovative Computing, Information and Control, 2006, pp. 161–165.

[100] D. MOBBS, T. WISE, N. SUTHANA, N. GUZMÁN, N. KRIEGESKORTE, AND J. Z. LEIBO, *Promises and challenges of human computational ethology*, Neuron, 109 (2021), pp. 2224–2238.

[101] K. E. MOKHTARI, B. P. HIGDON, AND A. BAŞAR, *Interpreting financial time series with shap values*, in Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering, 2019, pp. 166–172.

[102] E. MONTIJANO, A. PRIOLO, A. GASPARRI, AND C. SAGUES, *Distributed entrapment for multi-robot systems with uncertainties*, in Proc. of the IEEE Conference on Decision and Control, IEEE, 2013, pp. 5403–5408.

[103] A. MÖRTL, T. LORENZ, B. N. VLASKAMP, A. GUSRIALDI, A. SCHUBÖ, AND S. HIRCHE, *Modeling inter-human movement coordination: synchronization governs joint task dynamics*, Biological cybernetics, 106 (2012), pp. 241–259.

[104] S. J. MOTOWIDLO AND J. R. VAN SCOTTER, *Evidence that task performance should be distinguished from contextual performance.*, Journal of Applied psychology, 79 (1994), p. 475.

[105] R. R. MURPHY, *Human-robot interaction in rescue robotics*, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34 (2004), pp. 138–153.

[106] P. NALEPKA, R. W. KALLEN, A. CHEMERO, E. SALTZMAN, AND M. J. RICHARDSON, *Herd Those Sheep: Emergent Multiagent Coordination and Behavioral-Mode Switching*, Psychological Science, 28 (2017), pp. 630–650.

[107] P. NALEPKA, M. LAMB, R. W. KALLEN, E. SALTZMAN, A. CHEMERO, AND M. J. RICHARDSON, *First step is to group them: Task-dynamic model validation for human multiagent herding in a less constrained task*, in Proceedings of the Annual Meeting of the Cognitive Science Society, 2017, pp. 2784–2789.

[108] P. NALEPKA, M. LAMB, R. W. KALLEN, K. SHOCKLEY, A. CHEMERO, AND M. J. RICHARDSON, *A Bio-Inspired Artificial Agent to Complete a Herding Task with Novices*, in Proceedings of the Artificial Life Conference 2016, Cancun, Mexico, 2016, MIT Press, pp. 656–663.

[109] P. NALEPKA, M. LAMB, R. W. KALLEN, K. SHOCKLEY, A. CHEMERO, E. SALTZMAN, AND M. J. RICHARDSON, *Human social motor solutions for human–machine interaction in dynamical task contexts*, Proceedings of the National Academy of Sciences, 116 (2019), pp. 1437–1446.

[110] P. NALEPKA, C. RIEHM, C. B. MANSOUR, A. CHEMERO, AND M. J. RICHARDSON, *Investigating strategy discovery and coordination in a novel virtual sheep herding game among dyads*, in COGSCI 2015: Proceedings of the 37th Annual Meeting of the Cognitive Science Society, 2015, pp. 1703–1708.

[111] F. NARETTO, R. PELLUNGRINI, A. MONREALE, F. M. NARDINI, AND M. MUSOLESI, *Predicting and explaining privacy risk exposure in mobility data*, in International Conference on Discovery Science, Springer, 2020, pp. 403–418.

[112] F. NARETTO, R. PELLUNGRINI, F. M. NARDINI, AND F. GIANNOTTI, *Prediction and explanation of privacy risk on mobility data with neural networks*, in Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2020, pp. 501–516.

[113] A. A. PARANJAPE, S. CHUNG, K. KIM, AND D. H. SHIM, *Robotic herding of a flock of birds using an unmanned aerial vehicle*, IEEE Transactions on Robotics, 34 (2018), pp. 901–915.

[114] S. PARK AND J.-Y. CHOI, *Malware detection in self-driving vehicles using machine learning algorithms*, Journal of advanced transportation, 2020 (2020).

[115] Y.-J. PARK AND K.-N. CHANG, *Individual and group behavior-based customer profile model for personalized product recommendation*, Expert Systems with Applications, 36 (2009), pp. 1932–1939.

[116] A. B. PARSA, A. MOVAHEDI, H. TAGHIPOUR, S. DERRIBLE, AND A. K. MOHAMMADIAN, *Toward safer highways, application of xgboost and shap for real-time accident detection and feature analysis*, Accident Analysis & Prevention, 136 (2020), p. 105405.

[117] G. PATIL, P. NALEPKA, R. W. KALLEN, AND M. J. RICHARDSON, *Hopf bifurcations in complex multiagent activity: the signature of discrete to rhythmic behavioral transitions*, Brain Sciences, 10 (2020), p. 536.

[118] D. PICKEM, P. GLOTFELTER, L. WANG, M. MOTE, A. AMES, E. FERON, AND M. EGERSTEDT, *The Robotarium: A remotely accessible swarm robotics research testbed*, in Proc. of the IEEE International Conference on Robotics and Automation, 2017, pp. 1699–1706.

[119] D. PICKEM, M. LEE, AND M. EGERSTEDT, *The GRITSBot in its natural habitat - a multi-robot testbed*, in Proc. of the IEEE International Conference on Robotics and Automation, 2015, pp. 4062–4067.

[120] A. PIERSON AND M. SCHWAGER, *Controlling Noncooperative Herds with Robotic Herders*, IEEE Transactions on Robotics, 34 (2018), pp. 517–525.

[121] A. PIERSON, Z. WANG, AND M. SCHWAGER, *Intercepting rogue robots: An algorithm for capturing multiple evaders with multiple pursuers*, IEEE Robotics and Automation Letters, 2 (2017), pp. 530–537.

[122] P. M. PILARSKI, A. BUTCHER, M. JOHANSON, M. M. BOTVINICK, A. BOLT, AND A. S. PARKER, *Learned human-agent decision-making, communication and joint action in a virtual reality environment*, 2019.

[123] M. J. PRANTS, J. SIMPSON, P. NALEPKA, R. W. KALLEN, M. DRAS, E. D. REICHLE, S. HOSKING, C. J. BEST, AND M. J. RICHARDSON, *The structure of team search behaviors with varying access to information*, in Proceedings of the Annual Meeting of the Cognitive Science Society, vol. 43, 2021.

[124] J. R. QUINLAN, *Induction of decision trees*, Machine learning, 1 (1986), pp. 81–106.

[125] C. W. REYNOLDS, *Flocks, herds and schools: A distributed behavioral model*, in Proceedings of the 14th annual conference on Computer graphics and interactive techniques, 1987, pp. 25–34.

[126] M. T. RIBEIRO, S. SINGH, AND C. GUESTRIN, *"why should i trust you?" explaining the predictions of any classifier*, in Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144.

[127] M. J. RICHARDSON AND R. W. KALLEN, *Symmetry-breaking and the contextual emergence of human multiagent coordination and social activity*, in Contextuality from quantum physics to psychology, World Scientific, 2016, pp. 229–286.

[128] M. J. RICHARDSON, K. L. MARSH, AND R. M. BARON, *Judging and actualizing intrapersonal and interpersonal affordances.*, Journal of experimental psychology: Human Perception and Performance, 33 (2007), p. 845.

[129] L. M. RIGOLI, P. NALEPKA, H. DOUGLAS, R. W. KALLEN, S. HOSKING, C. BEST, E. SALTZMAN, AND M. J. RICHARDSON, *Employing models of human social motor behavior for artificial agent trainers*, in Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, 2020, pp. 1134–1142.

[130] K. W. RIO, G. C. DACHNER, AND W. H. WARREN, *Local interactions underlying collective motion in human crowds*, Proceedings of the Royal Society B: Biological Sciences, 285 (2018), p. 20180611.

[131] A. ROSENFELD AND S. KRAUS, *Predicting human decision-making: From prediction to action*, Synthesis Lectures on Artificial Intelligence and Machine Learning, 12 (2018), pp. 1–150.

[132] S. R. SAFAVIAN AND D. LANDGREBE, *A survey of decision tree classifier methodology*, IEEE transactions on systems, man, and cybernetics, 21 (1991), pp. 660–674.

[133] E. SALAS, J. A. CANNON-BOWERS, AND J. H. JOHNSTON, *How can you turn a team of experts into an expert team?: Emerging training strategies*, Naturalistic decision making, 1 (1997), pp. 359–370.

[134] E. SALTZMAN AND D. CAPLAN, *A graph-dynamic perspective on coordinative structures, the role of affordance-effectivity relations in action selection, and the self-organization of complex activities*, Ecological Psychology, 27 (2015), pp. 300–309.

[135] E. SALTZMAN AND J. KELSO, *Skilled actions: a task-dynamic approach.*, Psychological review, 94 (1987), p. 84.

[136] R. C. SCHMIDT, P. FITZPATRICK, R. CARON, AND J. MERGECHE, *Understanding social motor coordination*, Human movement science, 30 (2011), pp. 834–845.

[137] N. SEBANZ, H. BEKKERING, AND G. KNOBLICH, *Joint action: bodies and minds moving together*, Trends in cognitive sciences, 10 (2006), pp. 70–76.

[138] N. SEBANZ AND G. KNOBLICH, *Prediction in joint action: what, when, and where.*, Topics in Cognitive Science, 1 (2009), pp. 353–367.

[139] N. SEBANZ AND G. KNOBLICH, *Progress in joint-action research*, Current Directions in Psychological Science, 30 (2021), pp. 138–143.

[140] E. SEBASTIÁN AND E. MONTIJANO, *Multi-robot implicit control of herds*, in Proc. of the IEEE International Conference on Robotics and Automation, 2021.
to appear.

[141] R. SHARMA, M. KOTHARI, C. N. TAYLOR, AND I. POSTLETHWAITE, *Cooperative target-capturing with inaccurate target information*, in Proceedings of the 2010 American Control Conference, Baltimore, MD, June 2010, IEEE, pp. 5520–5525.

[142] S. A. SHEDIED, *Optimal trajectory planning for the herding problem: a continuous time model*, International Journal Machine Learning & Cybernetics, 4 (2013), pp. 25–30.

[143] M. A. SHIPP, K. N. ROSS, P. TAMAYO, A. P. WENG, J. L. KUTOK, R. C. AGUIAR, M. GAASENBEEK, M. ANGELO, M. REICH, G. S. PINKUS, ET AL., *Diffuse large b-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning*, Nature medicine, 8 (2002), pp. 68–74.

[144] A. SHRIKUMAR, P. GREENSIDE, A. SHCHERBINA, AND A. KUNDAJE, *Not just a black box: Learning important features through propagating activation differences*, 2017.

[145] K. SIMONYAN AND A. ZISSERMAN, *Very deep convolutional networks for large-scale image recognition*, 2014.

[146] H. SINGH, B. CAMPBELL, S. ELSAYED, A. PERRY, R. HUNJET, AND H. ABBASS, *Modulation of force vectors for effective shepherding of a swarm: A bi-objective approach*, in 2019 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2019, pp. 2941–2948.

[147] D. SLACK, S. HILGARD, E. JIA, S. SINGH, AND H. LAKKARAJU, *Fooling lime and shap: Adversarial attacks on post hoc explanation methods*, in Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 2020, pp. 180–186.

[148] P. E. SMOUSE, S. FOCARDI, P. R. MOORCROFT, J. G. KIE, J. D. FORESTER, AND J. M. MORALES, *Stochastic modelling of animal movement*, Philosophical Transactions of the Royal Society B: Biological Sciences, 365 (2010), pp. 2201–2211.

[149] H. SONG, A. VARAVA, O. KRAVCHENKO, D. KRAGIC, M. Y. WANG, F. T. POKORNY, AND K. HANG, *Herding by caging: a formation-based motion planning framework for guiding mobile agents*, Autonomous Robots, 45 (2021), pp. 613–631.

[150] S. STACY, C. LI, M. ZHAO, Y. YUN, Q. ZHAO, M. KLEIMAN-WEINER, AND T. GAO, *Modeling communication to coordinate perspectives in cooperation*, arXiv preprint arXiv:2106.02164, (2021).

[151] D. STROMBOM, R. P. MANN, A. M. WILSON, S. HAILES, A. J. MORTON, D. J. T. SUMPTER, AND A. J. KING, *Solving the shepherding problem: heuristics for herding autonomous, interacting agents*, Journal of The Royal Society Interface, 11 (2014), pp. 20140719–20140719.

[152] E. TOGNOLI AND J. S. KELSO, *The metastable brain*, Neuron, 81 (2014), pp. 35–48.

[153] P. TRAUTMAN, J. MA, R. M. MURRAY, AND A. KRAUSE, *Robot navigation in dense human crowds: Statistical models and experimental studies of human robot cooperation*, The International Journal of Robotics Research, 34 (2015), pp. 335–356.

[154] Y. TSUNODA, Y. SUEOKA, Y. SATO, AND K. OSUKA, *Analysis of local-camera-based shepherding navigation*, Advanced Robotics, 32 (2018), pp. 1217–1228.

[155] H. TUINHOF, C. PIRKER, AND M. HALTMEIER, *Image-based fashion product recommendation with deep learning*, in International Conference on Machine Learning, Optimization, and Data Science, Springer, 2018, pp. 472–481.

[156] C. E. TUNCALI, G. FAINEKOS, H. ITO, AND J. KAPINSKI, *Simulation-based adversarial test generation for autonomous vehicles with machine learning components*, in 2018 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2018, pp. 1555–1562.

[157] M. T. TURVEY, *Action and perception at the level of synergies*, Human movement science, 26 (2007), pp. 657–697.

[158] J. VAN DER KAMP AND I. RENSHAW, *Information-movement coupling as a hallmark of sport expertise*, Routledge, 2015.

[159] A. VARAVA, K. HANG, D. KRAGIC, AND F. T. POKORNY, *Herding by caging: a topological approach towards guiding moving agents via mobile robots*, in Proc. of the Robotics: Science and Systems XIII, 2017.

[160] R. VAUGHAN, N. SUMPTER, J. HENDERSON, A. FROST, AND S. CAMERON, *Experiments in automatic flock control*, Robotics and Autonomous Systems, 31 (2000), pp. 109–117.

[161] C. VESPER, E. ABRAMOVA, J. BÜTEPAGE, F. CIARDO, B. CROSSEY, A. EFFENBERG, D. HRISTOVA, A. KARLINSKY, L. MCELLIN, S. R. NIJSSEN, ET AL., *Joint action: mental representations, shared information and general mechanisms for coordinating with others*, Frontiers in psychology, 7 (2017), p. 2039.

[162] C. VO, J. F. HARRISON, AND J. LIEN, *Behavior-based motion planning for group control*, in 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2009, pp. 3768–3773.

[163] P. VOIGT AND A. V. D. BUSSCHE, *The EU General Data Protection Regulation (GDPR): A Practical Guide*, Springer Publishing Company, Incorporated, 1st ed., 2017.

[164] H. WADE AND H. A. ABBASS, *Cyber-shepherd: A smartphone-based game for human and autonomous swarm control*, in 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), IEEE, 2019, pp. 198–204.

[165] D. WANG, Q. YANG, A. ABDUL, AND B. Y. LIM, *Designing theory-driven user-centric explainable ai*, in Proceedings of the 2019 CHI conference on human factors in computing systems, 2019, pp. 1–15.

[166] R. E. WANG, S. A. WU, J. A. EVANS, J. B. TENENBAUM, D. C. PARKES, AND M. KLEIMAN-WEINER, *Too many cooks: Coordinating multi-agent collaboration through inverse planning*, in Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, 2020, pp. 2032–2034.

[167] Y. WANG, M. HUANG, X. ZHU, AND L. ZHAO, *Attention-based lstm for aspect-level sentiment classification*, in Proceedings of the 2016 conference on empirical methods in natural language processing, 2016, pp. 606–615.

[168] W. H. WARREN, *The dynamics of perception and action.*, Psychological Review, 113 (2006), pp. 358–389.

[169] W. WELFORD, J. M. BREBNER, AND N. KIRBY, *Reaction times*, Stanford University, 1980.

[170] S. A. WILMARTH, N. M. AMATO, AND P. F. STILLER, *Maprm: a probabilistic roadmap planner with sampling on the medial axis of the free space*, in Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C), vol. 2, 1999, pp. 1024–1031 vol.2.

[171] S. WILSON, P. GLOTFELTER, L. WANG, S. MAYYA, G. NOTOMISTA, M. MOTE, AND M. EGERSTEDT, *The Robotarium: Globally impactful opportunities, challenges, and lessons learned in remote-access, distributed control of multirobot systems*, IEEE Control Systems Magazine, 40 (2020), pp. 26–44.

[172] D. M. WOLPERT AND M. S. LANDY, *Motor control is decision-making*, Current opinion in neurobiology, 22 (2012), pp. 996–1003.

[173] A. WORM, *Joint tactical cognitive systems: modeling, analysis, and performance assessment*, in Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 42, SAGE Publications Sage CA: Los Angeles, CA, 1998, pp. 315–319.

[174] Y. YAMAMOTO, K. YOKOYAMA, M. OKUMURA, A. KIJIMA, K. KADOTA, AND K. GOHARA, *Joint action syntax in japanese martial arts*, Plos One, 8 (2013), p. e72436.

[175] L. YU, W. YUE, S. WANG, AND K. K. LAI, *Support vector machine based multiagent ensemble learning for credit risk evaluation*, Expert Systems with Applications, 37 (2010), pp. 1351–1360.

[176] H. ZHAO AND W. H. WARREN, *On-line and model-based approaches to the visual control of action*, Vision research, 110 (2015), pp. 190–202.