Explanations for Autonomous Agents

Alexandre Raymond

Darwin College

This dissertation is submitted on June, 2022 for the degree of Doctor of Philosophy
Declaration
This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or am concurrently submitting, for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or is being concurrently submitted, for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. This dissertation does not exceed the prescribed limit of 60,000 words.
Those people who think they know everything
are a great annoyance
to those of us who do.

– Isaac Asimov
Abstract

Recent years have seen an accelerated development of agents and systems capable of sophisticated autonomous behaviour. As the consequences of such agents’ actions begin to manifest in society, the need for understanding their decisions motivates the study of mechanisms for obtaining explanations that are compatible with human reasoning. However, the design of explainable systems often does not consider the impact that explanations could bring to machine and human agents alike. This thesis explores this challenge.

Our approach begins by looking at decentralised environments with complex regulations, where explanations must be exchanged to ensure orderly interactions between agents. To convert human rule sets into machine-compatible reasoning mechanisms, we propose an argumentation-based human-agent architecture to map human regulations into a culture for artificial agents with explainable behaviour. Our user studies in a hybrid, explainable, human-agent setting show that system complexity is a determining factor for the usefulness of explanations for humans. For autonomous agents, privacy and partial observability can introduce a notion of subjective unfairness in decentralised systems. We show that this effect can also be mitigated with the use of effective explanations.

In like manner, we look at Reinforcement Learning (RL) agents and investigate the possibility to orient the learning mechanism with explainable features. We call this process Explanation-Aware Experience Replay (XAER) and demonstrate that explanation engineering can be used in lieu of reward engineering for environments with explainable features. Further, we extend this concept into multi-agent RL and show how exchanging explanations in environments with partial observability can be used to obtain a more robust and effective collective behaviour.

Our conclusion is that the design of explainable systems should not only consider the generation of explanations, but also their consumption. Explanations can serve as tools for communicating precise and distilled information, and the insights gained by human agents could also be gained by machine agents, especially in systems with decentralised agency or partial knowledge.
Acknowledgements and Dedication

This thesis is the culmination of a lifetime dream. The solitary journey of a PhD feels like signing up for a marathon in the middle of a dense forest, where some trails have barely been set. It is not always obvious in which direction to run, with dead ends and perils aplenty. There is a time limit to reach one of the many possible finish lines, and the only way to get there in time is to commit to the journey and believe in the direction one has chosen.

To reach this destination, I had the invaluable guidance of my supervisor, Professor Amanda Prorok. In Brazilian Portuguese, we refer to academic supervisors as 'orientadores,' i.e. those who orient you towards a specific direction. In many cases, seasoned guides will point you towards a path that is familiar and comfortable to them, as a safe and tested route towards some known destination. Instead, in the beginning of my studies, Amanda did the opposite. She encouraged and fully supported me to pursue a research direction that I believed in, even if that was outside of both our comfort zones. I was truly privileged to benefit from her unfaltering and unconditional support at all times, and I hope this thesis serves as a motive for her pride, as much as being a founding member of her lab is an earnest reason for my own.

My privilege extends once further, as I had the guidance of Professor Hatice Gunes as my co-orientadora. Hatice gave me an important opportunity to present and explore my ideas as my first point of contact in the University. I was extremely fortunate to have been given a chance to work with her and learn a great deal about the human side of robotics, and the fascinating unanswered questions therein. That aspect formed a fundamental part of the argument constituting this thesis. Without the experience of working with Hatice and the AFAR lab, my work would lose an important component that adds much richness to it. I would like to express my unlimited gratitude to her.

Many hands supported me during my research endeavour, but some have contributed directly to its development and results. This work would have been much different without the contributions of my brilliant co-authors Francesco Sovrano, Guilherme Paulino-Passos, Luke Guerdan, and Matthew Malencia. Their insights and contributions have been momentous to the shape and quality of this thesis. It was a pleasure to collaborate with them, and I am forever honoured for being chosen as their co-author.

I would like to thank Richard Daltry and Dan Hook for being the first to believe in my research, and breaking new ground to support me in fulfilling this dream. Their support was decisive in setting this important chapter of my life in motion, and I only hope one day I can extend this generosity chain and do for someone else what they have done for me.

Winning the 2018 Industrial Fellowship Award with an original research project is yet another accomplishment that I will proudly remember for the rest of my life. I am delighted to be an Industrial Fellow, and I would like to thank the Royal Commission for the Exhibition of 1851
for the financial support and the amazing experiences and connections I will take with me throughout my career.

During this research, I was also sponsored by L3Harris ASV (formerly ASV Global). This research opens new avenues for interaction between humans and autonomous vehicles in regulated environments, and I hope it serves as inspiration for the future of their technology in making the oceans a safer place. I am very grateful for their support.

I was very fortunate to count on the pastoral and academic support of Professor Alan Blackwell in numerous occasions. It has been a privilege to collaborate with him and learn from his vast academic knowledge in numerous fields. I hope we can continue to collaborate in outstanding and future projects.

The Department of Computer Science and Technology at the University of Cambridge was my home for the earlier years of my PhD. The global pandemic transformed this reality in remote work for the second half of it, but I am grateful for the material, academic, and pastoral support provided throughout my degree. I extend a much special acknowledgement to Lise Gough, Joy Rook, and Marketa Green from the PhD Admin team. Their kindness made a measurable difference in my life as a student, and it is recomforting to know that they are there for us.

I stumbled across the field of Argumentation Frameworks during the very early days of my PhD. Rather serendipitously, the Computer Science department hosted Professor Francesca Toni for a talk on the very same subject during that period. Very generously so, Prof Toni agreed to spend some time after the talk personally explaining some of the concepts and philosophy of Argumentation Frameworks to me, then a puzzled fledgeling PhD student. Her explanations encouraged me to look more deeply into this fascinating field, and gave me the confidence to use this theoretical toolset as the underpinning mechanism behind this thesis. I am grateful for her time on that day, and it will be an honour to show her the positive consequences of that conversation 3 years later, as one of my thesis examiners.

Many friends supported me during this period. Choosing a selective roster of names to appear in this acknowledgements section is perhaps much harder than meeting strict page limits before paper submission deadlines. Throughout those years, my group of friends only grew larger and larger – I made dozens of new friends from Prorok Lab, from AFAR Lab, from Darwin College, from the Computer Science department, from conferences and other continents in the globe. I cherish the fact that I was well-looked after by all of them and would like to thank them for supporting me, consoling me during the hard days, and multiplying the joy of the bright ones.

I would like to thank my family in Brazil, France, and the UK for supporting me during this period. Sharing my successes and frustrations with them was always important to me, and I hope we will have further reasons to celebrate soon.

I am very fortunate to have crossed life paths with dearest Sara, who taught me so much in so little time. Her impact is still felt 16 years after her departure, and I hope this thesis serves as part of her living memory in our world.
To my parents, Nina Rosa and Alain, I owe them much of what I am today. My mother, with her fierce determination and protective spirit, made immeasurable sacrifices to ensure that I had a bright future in whichever direction I chose to follow. My father, with his staunch positivity and adventurous essence, was always closely interested in whatever I set out to do, and showed me how wide the world out there was. Her strength taught me to fight for this dream, and his optimism taught me to believe that this fight is worth it.

I reserve the last lines of this section to the most important person in my life. My wife Marina has been with me for every step of this journey. From coaching me before my Royal Commission interview in the middle of South Kensington to building our new life in Cambridge, she has inexhaustibly showered me with the purest form of love and companionship. She has given me strength during every day of this journey, as the oxygen I breathe for this marathon. Marina reassured me that everything would be fine in whatever way – after all, regardless of the outcome, we would still walk side by side. In kind, it would be pointless to need to dedicate this thesis to her. This thesis, much like the rest of my days, already belongs to her.

Thank you,

Alex Raymond
# Contents

Abstract v

Acknowledgements and Dedication vii

Contents xi

1 Introduction 1
   1.1 Research Questions 3
   1.2 Thesis Outline 4
   1.3 Contributions 7

2 Literature Review and Background 9
   2.1 Explanations 9
   2.2 Explainability in Autonomous Agents 10
      2.2.1 Validating Explainability 13
      2.2.2 Explainable Planning 15
      2.2.3 Explainable Agency 17
      2.2.4 Rule-aware Agency 17
   2.3 Computational Argumentation 19
      2.3.1 Abstract Argumentation 20
      2.3.2 Argumentation in Human-Agent Systems 22
   2.4 Multi-Agent Path Deconfliction 22
   2.5 Human-Agent Path Deconfliction 24
   2.6 Model-Free Reinforcement Learning 25
   2.7 Explanations for RL 28
   2.8 Prioritised Experience Replay 29
   Summary 33

3 Cultures: A Framework for Explainable Rule-Aware Agency 35
   3.1 Explainable Conflict Resolution 36
   3.2 Background 37
      3.2.1 Dialogue Games 37
      3.2.2 Explanations from Argumentation 39
   3.3 Cultures 40
      3.3.1 Propositional Dialogues 40
      3.3.2 Argument Verification 41
      3.3.3 Explanation Generation 43
3.4 Privacy-aware Cultures ........................................ 45
  3.4.1 Alteroceptive Cultures .................................. 46
  3.4.2 Privacy-Aware Dialogues ................................. 48
Summary .......................................................... 51

4 Explainable Human-Agent Deconfliction .................. 53
  4.1 Proof of Concept Study ..................................... 55
    4.1.1 Path Deconfliction Environment ....................... 56
    4.1.2 The Busy Barracks Game ............................... 57
    4.1.3 User Study ............................................. 60
  4.2 Results ....................................................... 62
    4.2.1 Score ................................................... 62
    4.2.2 Collisions .............................................. 63
    4.2.3 Times .................................................. 63
    4.2.4 User Experience ....................................... 64
  4.3 Discussion and Conclusion ................................. 66
Summary .......................................................... 67

5 Explanations and Subjective Fairness in Multi-Agent Systems 69
  5.1 Explanations, Dialogues, and Privacy ..................... 71
  5.2 Problem Definition ......................................... 73
    5.2.1 Information, Privacy and Fairness ................... 74
    5.2.2 Fairness Loss ......................................... 78
  5.3 Empirical Analysis ......................................... 81
    5.3.1 Setup .................................................. 81
    5.3.2 Evaluation Metrics .................................... 82
    5.3.3 Strategies ............................................. 83
    5.3.4 Simulations ............................................ 84
    5.3.5 Privacy Efficiency ..................................... 85
  5.4 Multi-agent Application .................................. 86
    5.4.1 Simulator .............................................. 87
    5.4.2 Boat Culture ........................................... 88
    5.4.3 Dialogues and Collision Avoidance .................. 88
    5.4.4 Privacy Efficiency Experiments ...................... 91
    5.4.5 Ride Quality .......................................... 91
    5.4.6 Subjective and Objective Trajectories ............... 93
  5.5 Discussion .................................................. 96
    5.5.1 Randomised Cultures .................................. 97
    5.5.2 Multi-Agent Experiment ............................... 97
  5.6 Conclusion .................................................. 98
Summary .......................................................... 101
## List of Figures

1.1 Explanations from machines to humans ........................................ 5  
1.2 Explanations from machines to other machines .......................... 5  
1.3 Explanations from humans for machines ................................. 6  

2.1 Opaque, interpretable, and comprehensible systems ................. 13  
2.2 Explainable systems ..................................................................... 13  
2.3 Argumentation framework ......................................................... 19  
2.4 Argumentation framework. ......................................................... 21  

3.1 Example diagram of a culture ...................................................... 43  
3.2 Example of alteroceptive culture ................................................. 47  

4.1 The Busy Barracks game UI ......................................................... 58  
4.2 Hints in UI ................................................................................. 61  
4.3 User study performance results ..................................................... 63  
4.4 Facial affect analysis ................................................................. 65  

5.1 Fairness losses across different strategies ................................... 85  
5.2 Privacy costs for unrestricted dialogues ....................................... 86  
5.3 Multi-agent collision avoidance simulation ................................ 87  
5.4 Collision avoidance mechanism .................................................. 90  
5.5 Privacy costs for unrestricted dialogues ....................................... 92  
5.6 Simulation data ......................................................................... 93  
5.7 Nominal, Subjective and Objective trajectories ......................... 94  

6.1 Overview of XAER ................................................................. 107  
6.2 Cluster size proportion ............................................................. 112  
6.3 GridDrive and GraphDrive ....................................................... 115
List of Tables

4.1 Pairwise MW for Score (S), Collisions (Col), and Challenge (Cha) .................. 64
5.1 Boat culture properties .................................................................................. 89
6.1 XAER experiment results ............................................................................ 118
Introduction

‘No, no! The adventures first, explanations take such a dreadful time.’

— Lewis Carroll, *Alice in Wonderland*

Imagination and ingenuity coexist in a feedback loop that fuels the development of robotics and artificial intelligence (AI). Art propels our imagination and poses questions about what is possible in futures that do not yet exist [1], whilst ingenuity is the ingredient for materialising such futures, rekindling our imagination, and broadening our horizons. As we once conjectured about ‘reasoning machines’ [2], this vision has been realised by current advancements in the field, which granted machines sophisticated decision-making capabilities. Modern AIs have attained super-human performance in an ever-increasing set of tasks [3], and the persistent increase in computing resources makes it all but inevitable that this trend is due to continue. Such systems will greatly overtake human intelligence in terms of raw processing speed.

Seeing that artificial decision-making agents emerge into ubiquity in modern everyday lives, increasingly more non-experts will come into contact with those applications. In many cases, the sophistication and complexity that accompany the technology yield good results with regards to sheer optimality, but humans are not any the wiser with regards to the reasoning supporting a given decision. Could they keep up with humans in terms of clarity? It can be argued that the subjective and emotional layers of human behaviour are not major contributors to clearness and perspicuity – after all, humans can regularly confuse other humans in numerous contexts. Nonetheless, the desire for clarity moves our societies and mechanisms, as it represents the degree of effectiveness in transferring information across individuals. Society requires cooperation, and maximising our ability to relay knowledge and to express clear intentions is fundamental for establishing trust and stability amongst its members.
Inasmuch as robots and intelligent artificial agents progress in sophistication, they obtain increasingly more autonomy and start taking part in the same systems and societies that humans do – no longer as tools, but rather as peers. For example, the much-anticipated autonomous vehicles of the future will share the public space with other human drivers, and should be subjected to similar levels of scrutiny and accountability to those of humans. The prospect of human-agent societies elicits questions about how the relationship between humans and autonomous agents will be shaped. Will it be a relationship of near-mystical belief in the wisdom of intelligent agents where ‘trust the machine, the machine is seldom wrong’ becomes an inevitable mantra? Or will it be one of a more rational and empirical nature, where humans and agents, as fellow intelligent beings, share a rational relationship based on observation and logical reasoning? Towards the latter, we can draw from our experience in dealing with another type of intelligent agents over millennia: other humans.

Theological debates aside, the lack of a centralised mechanism that governs the behaviour of all humans means that the ability to explain things proves necessary for us to resolve local conflicts with other humans and transfer information in an effective way. This very thesis serves as a self-referential example of an exposition of arguments that attempts to explain and justify new findings. In such wise, the scientific paradigm requires those explanations for two main reasons: to allow for the reasons behind decisions and results to be challenged and trusted by peers, and for the collective expansion of human knowledge. Trusting autonomy without explanations sets aside the need for justifying decisions, and confers a new authority to a different entity in the form of the AI system. If machines are able to learn and decide without the requirement of explaining their process, we as humans will no longer learn with them – and might ultimately be left with little beyond blindly trusting their decisions.

This tenet partly underpins the motivation for our present work. Yet, it is not a necessarily novel statement that intelligent agents should be explainable to humans. The field of eXplainable AI (XAI) attracted significant attention in the last years, and is ripe with flourishing research from diverse communities. However, this work also balances itself on another much less ex-
1.1 Research Questions

This thesis aims to establish foundational steps towards the scenario of collaborative human-agent societies described above. The actions taken throughout this research project revolved around the vision of a future where humans and autonomous agents partake in peer-to-peer interactions and collective decision-making. To this end, this work has led towards the following research questions:

**Question 1.** Which mechanisms could allow humans and autonomous agents to exchange explanations and reason collectively about their state and observations?

This question is motivated by the imminent proliferation of multi-agent and human-agent systems in our society. Can we provide a mechanism that allows heterogeneous agents to reason with compatible representations/symbols? This pertains more specifically to scenarios where human agents and machine agents share environments where their actions might need justification, and leads us to our second research question:

**Question 2.** How can explanations be useful to humans in human-agent environments?

When agents are tasked with achieving a goal in a shared environment, it is expected that their actions may occasionally be in conflict. If these agents are required to justify their actions, then how will they do so? And will explanations always be beneficial to humans in such environments? The purpose of this research question is to explore whether this would be beneficial for the agents, and to evaluate in which cases those explanations improve human agent performance. In addition, we would like to explore the consequences of agents withholding information about actions taken, and whether this would lead to suboptimal decisions being taken. This is defined by:

In guaranteeing that humans and machines can exchange explanations in both directions, the difference between human and machine agents should ideally be minimised for the purposes of our argument. On those grounds, our concept of ‘multi-agent systems’ might include human agents in its composition without any significant loss of generality.
Question 3. In explainable conflict resolution environments, what effects can be observed when agents have privacy restrictions?

This question addresses the trade-off between the need to provide explanations that justify actions and improve transparency, and the need to protect the privacy of agents and the information that they hold. A need for privacy might arise to protect agents from malicious agents that are not part of the same community, or to protect agents that wish to keep their information private. When information exchange is limited, are decisions suboptimal? And if so, what strategies can be placed to mitigate this effect?

Ultimately, we turn to the practical applications of these questions – specifically, how can we use the answers to these questions to improve the performance of agents using machine learning methods:

Question 4. How can we leverage explainable facilities of a system to improve the performance of self-learning machine agents?

Self-learning agents are agents that are capable of learning from data and making decisions based on past experience. Can the answers to these questions help us provide explanations that improve the performance of these agents?

1.2 Thesis Outline

In this work, we regard explanations as a tangible outcome of an explainable process or system. We depart from more fluid notions of ‘interpretability’, where systems benefit from some degree of increased transparency to assist human deduction into what may have contributed to a decision. For us, explanations are units of information that may transition between different agents in a system, whether these are human or machine agents. To facilitate this exchange, we build on argumentation frameworks [5] to devise cultures, a framework for generating symbolic explanations in dynamic rule-aware systems in a human-compatible way. Our choice is supported by the fact that classical logic does not provide an authentic representation of common sense reasoning, as under a scenario of incomplete information a human may draw conclusions that
can be withdrawn later, when new information is presented [6]. Argumentation-based approaches attempt to fill in this gap by providing a framework for defeasible reasoning [5], which grants systems clear decision-making mechanisms that provide not only resolutions, but also the reasons that support it [7].

Our first step is to demonstrate how humans can benefit from explanations generated by artificial agents in human-agent environments (Figure 1.1). We exemplify our architecture with a multi-agent resource contention application in the context of the problem of multi-agent path deconfliction. We show how humans and agents can deconflict trajectories whilst respecting externally-defined ‘rules of way.’ Towards this end, we design a computer game implementing the proposed architecture and conduct a user study to validate it. Our results show that explanations provide a significantly higher improvement in human performance when systems are more complex. Via quantitative and qualitative findings, we learn that when rules are more complex, explanations significantly reduce the perception of challenge for humans.

Ensuingly, we turn to observations on the phenomenon of machines explaining to other machines (Figure 1.2). The vision of decentralised systems of the future involves cooperation and conflict resolution not only in human-agent scenarios, but in agent-agent circumstances as well. Autonomous agents can be heterogeneous in terms of shape, form, and capability – but even identical agents in terms of hardware and software might serve different purposes, and preserve other interests at hand. A fictitious example would be a busy channel where autonomous vessels from different countries, civilian, commercial, and military, all must share information and mutually negotiate right of way in a constricted passageway. If there are concerns or restrictions of privacy (say, diplomatic details), agents may be limited in terms of the information chosen to compose explanations in the first place. Obfuscating the actual reasons that support an explanation might lead to other agents – unbeknownst to perfect information – to experience a perception of unjustifiability, or subjective unfairness.

Observing this aspect, we extend our culture formalism to support privacy-aware dialogues (called alteroceptive cultures), and we introduce a formalism of the relationship between privacy and fairness in multi-agent systems. Our proposed interaction

[6]: Amgoud et al. (2009)  
[5]: Dung (1995)  
[7]: Zeng et al. (2018)
approach is an architecture for privacy-aware explainable conflict resolution where agents engage in a dialogue of hypotheses and facts. This inquest upon the privacy-fairness relationship takes form as we define subjective and objective fairness on both the local and global scope and formalise the impact of partial observability due to privacy in these different notions of fairness. We study our proposed architecture and the privacy-fairness relationship in the abstract, testing different argumentation strategies on a large number of randomised cultures. We empirically demonstrate the trade-off between privacy, objective fairness, and subjective fairness and show that better strategies for building explanations can mitigate the effects of privacy in distributed systems. In addition to this analysis across a broad set of randomised abstract cultures, we analyse a case study for a specific scenario: we instantiate our architecture in a multi-agent simulation of prioritised rule-aware collision avoidance with limited information disclosure.

And finally, our central argument is hemmed in by the remaining unexplored direction: could machines benefit from explanations provided by humans (Figure 1.3)? We shift to the domain of learning agents for the last part of our investigation. As further integration and adaptation will require autonomous systems to partake in said societies’ implicit and explicit rules, Reinforcement Learning (RL) agents can encounter difficulties to master environments governed by complex sets of rules (and their corresponding exceptions). Intuitively, we refer back to our first step showing that human performance in rule-heavy environments is improved upon the presence of expert explanations provided by argumentative cultures, and extend this reasoning to machine agents.

With large numbers of corner cases arising as a consequence of dense rulesets, generating a sufficiently diverse set of experiences and exposing these exceptions to an RL agent can be challenging. The technique of Prioritised Experience Replay (PER) looks at over-sampling experiences that are most poorly captured by the agent’s learned model. Building on that, we propose a method for organising experience by means of partitioning the experience buffer into multiple clusters, each representing a distinct explanation associated with a collection of experiences that serve as examples. The notion of explanation engineering surfaces as a mechanism to allow humans to ori-
ent the learning agent through means of selecting which experiences (and explanations) are more important to the task at hand. We introduce new RL environments that are compatible with cultures and modular rulesets, and modify 3 seminal algorithms from literature with this technique. Performance is consistently superior with the explanation-aware versions of those algorithms, compared to traditional PER baselines, indicating that explanation engineering can be used in lieu of reward engineering for environments with explainable features.

1.3 Contributions

This thesis contributes to the fields of human-agent interaction, multi-agent systems, and artificial intelligence by showcasing the benefits that explanations can bring to machine agents as well as human agents. More specifically, we list our contributions below:

- The formalism of *cultures* as an application of argumentation frameworks to dynamic multi-agent systems in rule-aware environments. This theoretical foundation allows for agents in multi-agent systems to provide contextual and up-to-date explanations about their state with regards to a present ruleset. The concept is described in detail in Chapter 3, and applications can be seen on Chapters 4, 5, and 6.

- An extension to the taxonomy of the need for explanations in human-agent systems. We demonstrated with a user study that rule complexity is also a determining factor as to whether explanations are necessary in human-agent systems.

- We formalise the concept of *subjective fairness* for machine agents as an outcome of environments where explanations are limited in privacy and demonstrate practical effects in simulation.

- We create a mechanism for *explanation-aware experience replay* and show how distinct types and instances of explanations can be used to partition replay buffers and improve the rule coverage of sampled experiences in 3 seminal RL algorithms.
The work developed for this thesis resulted in 1 conference and 2 journal publications, listed below:


Some claims presented in this thesis are also supported by the following study:


The rest of this thesis is organised in 6 more chapters. Chapter 2 introduces a panorama of related work and preliminaries required for this thesis. Chapter 3 presents our proposed framework of cultures and alteroceptive cultures. Chapter 4 demonstrates our human-agent explainable deconfliction study. Chapter 5 contains our proposed formalism of the privacy-fairness trade-off and the effect of explanations in multi-agent scenarios. Chapter 6 stages our model for explanation-aware experience replay for RL agents. We summarise our results and present remarks towards future work in Chapter 7.
In this chapter, we introduce an overview of recent literature on topics that surround this thesis. We look into explainability, argumentation, explainable planning, multi-agent path deconfliction, human-robot path deconfliction, as well as insights on Reinforcement Learning (RL). As we lay out the literature, we construct a line of justification that builds on the topics presented in Chapter 1.

2.1 Explanations

The notion of explanations is traditionally linked to the concept of understanding, in the sense that a particular entity is able to explain why a particular event (or set of events) is happening. This is the case, for example, of causal explanations [10], which are linked to the notion of causal understanding [11].

Causality [12] is a notion with a long history in the philosophy of

[10]: Gorovitz (1965)
[11]: Corrigan et al. (1996)
[12]: Pearl (2009)
of science, and it is often associated with the notion of causal relationships [13] [14].

In particular, there are other types of explanations that can be useful for autonomous agents, such as teleological explanations [15], mechanical explanations [16], and explanations by means of rules [17].

Teleological explanations are closely linked to the notion of purpose [18]. To explain something teleologically means to explain something by means of the purpose for which it was created (or for which it is being used). Mechanical explanations are closely linked to the notion of causal mechanism [16]. To explain something mechanically means to explain it by means of the causal mechanism that is responsible for making it work. Explanations by means of rules are closely linked to the notion of prescription. To explain something by means of rules means to explain it by means of the rules that are expected to apply to it.

Explanations by means of rules are of particular interest in this thesis, as they are closely related to the notion of rule-based explanations [19]. Rule-based explanations are directly linked to the notion of rulesets (also called cultures in the context of this work), which is a collection of rules that should be followed in a particular context by a population of agents.

Case-based explanations [20] are also of particular interest. They are closely linked to the notion of examples that contain information about past events, and that can be used to explain current events by means of case-based reasoning.

More importantly, in this work explanations are regarded as a medium for communicating important and specific information to relevant parties, be it from agents to humans, agents to agents, or from humans to agents.

2.2 Explainability in Autonomous Agents

If we consider environments where heterogeneous human and machine participants have independent agency (collaboratively or not), sharing a common protocol is only the first step towards multi-agent engagement. For example, in the context of path planning, participants with individual goals and incomplete knowledge might need to justify and argue why they should
have right of way over a conflicting third party, e.g., an aeroplane with a seriously ill passenger requests taxiing priority over others, or a ship informs another oncoming boat that the channel ahead is closed. Likewise, the same reasoning applies to single-agent systems. Autonomous vehicles would benefit from a human-interpretable decision-making process that provides justification in a way that is understandable by non-experts. If passengers in an autonomous vehicle fail to understand a decision reached by the vehicle’s planning algorithms, it is in their interest to probe the system for clarification, and perhaps even contest that decision, if a plausible argument can be made.

The onset of machine learning systems and the popularity of methods such as support vector machines and artificial neural networks have led to AI solutions that are efficient, but indecipherable with regards to their behaviour or rationale to draw a conclusion [21]. For that reason, the field of explainable AI (XAI) [22] is interested in systems that are not only clear regarding ‘how’, but also as to ‘why’ certain decisions were made. In more traditional data-driven domains, the focus is on explaining black-box algorithms [23, 24] and their decisions regarding the data presented as input. Three main reasons are presented as the motivation for explainable AI systems: the need for trust, the need for interaction, and the need for transparency [25]. In order to satisfy those conditions, it is decisive that robotic agents must be capable of deliberating and explaining their decision-making process [26], particularly when humans are involved [27].

The community have not yet agreed on a standard definition for explainable AI [28, 29], but have acknowledged the urgency of a well-defined nomenclature [30]. This disagreement leads to a high number of taxonomies being proposed by the community. Speith [29] introduces a meta-analysis of explainable AI taxonomies published in recent years. His contribution includes the proposal a larger taxonomy, joining aspects of chosen works into a more complete classification.

Commonly, terms such as explainability, interpretability, understandability, comprehensibility, and intelligibility are oftentimes used in literature interchangeably to convey the same idea. Interpretability is favoured in machine learning contexts [31]. For simplicity, we adhere to the taxonomy proposed by Doran et al. [32] (see Figures 2.1 and 2.2), namely:

[21]: Pedreschi et al. (2019)
[22]: Holzinger et al. (2022)
[23]: Guidotti et al. (2018)
[24]: Zarlenga et al. (2021)
[25]: Fox et al. (2017)
[26]: Langley et al. (2017)
[27]: Wang et al. (2016)
[28]: Adadi et al. (2018)
[29]: Speith (2022)
[30]: Lipton (2018)
[31]: Molnar et al. (2022)
[32]: Doran et al. (2018)
Opaque systems: systems where the user cannot see the mechanisms mapping inputs to outputs; systems that are closed-source or whose working mechanisms are inaccessible or undisclosed to the interested stakeholder. Example: proprietary AI systems with limited access.

Interpretable systems: systems where the user can study and understand the mapping of inputs to outputs with sufficient technical or expert knowledge. Modern interpretability techniques \cite{22, 31} attempt to bridge the gap between opaque and interpretable systems, and can provide some insight on the relevance of certain features. Examples: linear classifiers, SVMs, and deep neural networks.

Comprehensible systems: systems that emit symbols (words, visualisations, etc) along with its output to help the user to associate properties of the inputs to the output. Some implicit knowledge is required from the user to create semantic connections between the input, symbols, and output. It is worth noting that comprehensible systems need not be transparent such as interpretable systems. Example: a collection of networks that list identified objects/features that compose a scene and outputs its classification of the scene.

Explainable systems: systems that can specify the line of reasoning involved in the decision-making process using human-understandable features of the input data. Interpretable and comprehensible models may enable explanations of decisions, but the explanations are not directly given by themselves.

The work in Adadi et al. \cite{28} performs a survey on techniques and methods used to provide or guarantee explainability in systems. Most studies available concern machine learning systems. They propose three dimensions to this problem: the intrinsic interpretability vs the opaqueness of methods, global vs local interpretability, and model-specific vs model-agnostic interpretability. For our approach, we are mostly interested in justifying the explainability of our approach via designing a model that is transparent by nature.
2.2 Explainability in Autonomous Agents

2.2.1 Validating Explainability

To whom is explainability beneficial? Despite all the formal and computational devices to achieve a meaningful comprehension of AI systems, the benefit of explainability concerns *humans* most of all. Thus, the HCI community is organising itself to accommodate and support the development of systems that are explainable by design [33, 34].

---

[33]: Abdul et al. (2018)  
[34]: Gilpin et al. (2018)
Attention has been drawn towards XAI in very recent times, and agreement is scarce on taxonomies and terminology. Moreover, deciding on how to evaluate or measure explainability is still up for debate. One of the first observations into this problem was presented by Huysmans et al. [35], who performed an empirical study to evaluate interpretability in decision trees, decision tables, and textual descriptions.

In the context of machine learning systems, Doshi-Velez et al. [36] propose a taxonomy for evaluating interpretability:

▷ **Application-grounded evaluation:** conducting human experiments in a real application. Example: compare it to domain experts deriving explanations, and the efficacy of such explanations in assisting others.

▷ **Human-grounded metrics:** simpler human-subject experiments, often simplifying or abstracting away the task at hand. Example: humans are presented with an explanation and an input, and must simulate the model’s output.

▷ **Functionally-grounded evaluation:** no human experiments, focus on a formal definition of interpretability.

Complementarily, Poursabzi-Sangdeh et al. [37] postulate three questions that evaluations should answer:

▷ How well can people estimate a model’s prediction?

▷ How much do people trust those predictions?

▷ How well can people detect mistakes in the model?

Bau et al. [38] presents the question: is interpretability achieved through a specific alignment of the feature space? They define interpretability as an alignment with a set of human-interpretable concepts, and quantify it by first identifying a set of visual concepts where humans agree with the ground-truth interpretations of the dataset, then observing how hidden variables affect such concepts, and lastly, generating a pairwise hidden variable-concept alignment. The evaluation then takes place by showing the most highly-activating regions for each unit and ask humans if a given phrase describes the highlighted region. They conclude that although user accuracy in simulating...
2.2 Explainability in Autonomous Agents

those models increases with transparent models, interpretability is not an inevitable result of the discriminative power of a representation.

Rosenfeld et al. [39] postulate that the necessity for explainability in human-agent systems follows a taxonomy of three types of explanations: not helpful, beneficial, and critical. They posit that if humans will not accept a system without an explanation, then the need for explainability is critical. Likewise, explanations can range in significance depending on their ability to engender trust in human users. We aim to introduce another dimension to this analysis, by empirically observing that the complexity of the rules governing a system may also affect the usefulness of explanations where human performance is concerned.

In a similar vein, Fan et al. [40] classify explanations by their level of abstraction, motivated by the necessity or interest of the explainee. This taxonomy proposes lean explanations as representations that associate outputs of an AI model directly to its input space, without further probing the details that compose the input. Conversely, argumentative explanations rely on exposing the reasoning that supports the decision-making mechanism beyond selecting within the input, by diving into the workings of the model and uncovering fine-grained representations of its components.

### 2.2.2 Explainable Planning

As artificial intelligence is a vast and heterogeneous domain, it is expected that the adoption of explainability will encounter challenges that are particular to each sub-discipline of AI. For our work, we are specifically interested in XAIP, or explainable planning [25, 41].

The field of XAIP intersects with other disciplines from classical planning. Almost homonymously, the concept of plan explanation [42, 43] functions in a different way: those approaches are concerned in presenting the outcome of a planner in human-understandable symbols or protocols, such as using dialogue systems or using visual descriptions of the final plan. Following the trend, the also surprisingly homonymous term plan explicability [44, 45] also appears, this time using learning to represent
a human-aware classification problem where interpretability is defined as a mapping of symbols to actions by human agents. The work of Kulkarni et al. [45] attacks this problem by minimising the distance between the plan devised by the agent and the human’s expected plan. Those terms might appear interchangeably in some works, as there is no established taxonomy in the community.

Explicitly, explainable planning concerns justification, or why a planner decided on a specific plan and not any other. Langley et al. [26] elects the desiderata for explainable agents: they must be able to explain decisions made during plan generation, report on actions executed, explain divergences in events and its response, and communicate in ways that humans may understand. Some approaches resort to visualisation as a process of explanation. Chakraborti et al. [46] utilises a visual mechanism to highlight information that contributed to the chosen plan, and outlining why a certain action outperform other candidate actions.

In the same manner with explainable AI in general, the community have recently been working on using argumentation-based approaches towards explainable planning. Those approaches benefit from the intuitive representation of causality present in argumentation frameworks and extend them to model decisions in a planner. Once those decision points are semantically grounded, the explanation falls out naturally. The challenge lies on adequately modelling the planning problem as an argumentation framework. Two of those works are described below.

Collins et al. [47] propose an algorithm to extract causal relationships between actions into a knowledge base, and derive an argumentation framework afterwards, using extension semantics to generate plans. They demonstrate that semantically grounded decision points can generate traceability due to the nature of the system.

In similar fashion, an important result from Čyras et al. [48] shows how argumentation frameworks provide an accessible knowledge representation tool for modelling optimisation problems. Their work consist of building three different argumentation frameworks modelling sub-properties of the problem, mapping decisions as arguments and capturing feasibility and optimality conditions as attacks. They also propose a definition of
2.2 Explainability in Autonomous Agents

a good explanation: it needs to be efficiently attainable, with few causal relationships, and admitting simple natural language interpretations. The authors posit that in order to build trust, explanations must be associated with formal representations that provide interpretable certificates.

2.2.3 Explainable Agency

The concept of explainable agency is proposed by Anjomshoae et al. [49] to refer to autonomous agents that explain their actions and reasons supporting specific decisions. The phases of explanation are threefold [49, 50]:

1. Explanation generation: this phase consists of providing justification regarding an action taken or result obtained by an agent.

2. Explanation communication: from the generated explanations, decide what should be shown to the user, and how to show it.

3. Explanation reception: an evaluation on how the explainee understands/processes the explanation. Can be done with user studies, subjective evaluation, or similar metrics.

This is underpinned by the concept of theory of mind [51], which attributes and projects a mental state to others in order to understand intentions and beliefs that are different to one’s own. With explanations, users would be able to understand the behaviour of an explainable agent by using those explanations to build a coherent theory of mind. As humans create mental models of intelligent agents via anthropomorphisation [52], we look for a formalism that allows for the bilateral exchange of explanations between humans and agents in a way that is computationally feasible, but also amenable to human models of reasoning.

2.2.4 Rule-aware Agency

We are interested in ensuring that autonomous agents consider their liability [53] and can express justification for their agency with regards to the present ruleset to be followed — preferably in a human-understandable way.
Rizaldi et al. [54] tackle the liability and accountability problem in the autonomous vehicle domain with a manual formalisation of specific traffic rules, using automated theorem proving techniques. Their approach is formally rigorous but is specialised for a specific set of traffic rules only and does not generalise beyond. Cranefield et al. [55] propose that ideal accountable agents must:

- **i)** understand what is expected from them (from rules/obligations);
- **ii)** answer queries about their decision-making (being explainable);
- **iii)** carry out argumentative dialogues in which beliefs and plans are challenged and justified;
- **iv)** adapt their reasoning apparatuses or update their plans as a result of accountability dialogues;
- **v)** take human values into account when reasoning.

In order to achieve more realistic explainability for humans, we spur the necessity for more realistic models of reasoning. Expressly, classical logic does not provide an authentic representation of common sense reasoning, as under a scenario of incomplete information a human may draw conclusions that can be withdrawn later, when new information is presented [6]. This phenomenon conflicts with the assumption of monotonicity of entailment present in classical logic, i.e., the set of conclusions does not grow with the addition of new information. Argumentation-based approaches attempt to fill in this gap by providing a framework for defeasible reasoning [5], which grants systems clear decision-making mechanisms that provide not only resolutions, but also the reasons that may support it [7].

Argumentation approaches walk hand in hand with the desiderata proposed by Cranefield et al. [55], as they allow us to:

- **i)** enable norm-aware reasoning [56];
- **ii)** generate explanations [57, 58];
- **iii)** carry argumentative dialogues to support their positions [6, 59];
- **iv)** perform meta-reasoning [60];
- **v)** consider human values [61].
2.3 Computational Argumentation

As a strong mechanism for providing accountable and explainable agency, we choose argumentation frameworks [62] as the underlying theoretical tool for our investigation. The next section introduces important literature and preliminaries on computational argumentation.

2.3 Computational Argumentation

The field of non-monotonic logic [63] attempts to capture inferences in which reasoners draw tentative conclusions that can be retracted in the presence of further conflicting evidence. This type of reasoning is called defeasible reasoning.

An alternative formalism to defeasible reasoning comes in the form of computational argumentation [5, 64–66]. Argumentation frameworks, or argumentation systems, are reasoning models where information is established as a set of semantically meaningful arguments alongside an attack relationship between items of that set, composing a directed graph (see Figure 2.3). Dung’s seminal work [5] demonstrates that logic programming, defeasible reasoning, and most formalisms of non-monotonic logic are argumentation systems. Argumentative reasoning allows for persuasive or deliberative multi-agent strategic interaction by performing inferences on the argumentation graph or by playing dialogue games between participants [67].

Dung’s work [5] has spurred a community of researchers to work on variations and extensions of his original Abstract Argumentation (AA) model. Some models look deeper into the logical structure underpinning arguments, exemplified by the Assumption-based Argumentation (ABA) framework [64]. ABA uses a deductive structure with inference rules, which are combined with arguments that correspond to a set of assumptions in order to prove claims. Another popular model is ASPIC+ [68], which focuses on resolving arguments via explicit preferences and inference rules.

Authors have identified the potential for using argumentation as an instrument to constitute explainable systems [57, 58, 69]. Fan et al. [58] pose the semantics of related admissibility, composed of arguments that defend the argument in question. They argue that this model goes beyond just checking the acceptability of arguments, instead using other arguments to justify and

Figure 2.3: Example of an argumentation framework \( \mathcal{AF} = (\mathcal{A}, \mathcal{R}) \), where \( \mathcal{A} = \{a, b, c, d\} \) and \( \mathcal{R} = \{(b,a),(c,b),(d,b)\} \). Every node in this argumentation framework represents an argument, and the arcs denote attacks between arguments. In this example, since \( c \) and \( d \) attack \( b \), which in turn attacks \( a \), it can be said that \( c \) and \( d \) defend \( a \).
explain why a specific argument was admitted. Contrariwise, Zeng et al. [69] suggest modelling explainability by using argumentation to identify points of failure, i.e., identifying the attackers from which the arguments cannot be defended.

For this thesis, we chose Dung’s model as our reference for argumentation frameworks due to its simplicity and expressiveness.

### 2.3.1 Abstract Argumentation

A seminal paper from Dung [5] introduces the concept of an argumentation framework, also called abstract argumentation (AA). His framework considers arguments as purely abstract entities, with no special attention paid to their internal structure. Modelling occurs at the level of relationships between those abstract entities.

We will use and adapt some definitions from Dung’s work and other authors [70–72], as follows.

**Definition 2.3.1** An argumentation framework is a directed graph $AF = (\mathcal{A}, \mathcal{R})$, where $\mathcal{A}$ is a set of arguments (vertices) and $\mathcal{R}$ is a set of directed, binary attacks between arguments (arcs), i.e., $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$. We also say attacks $(a, b)$ holds iff $(a, b) \in \mathcal{R}$. Likewise, a set $S$ of arguments attacks another set of arguments $T$ (or $T$ is attacked by $S$) if any argument in $S$ attacks an argument in $T$.

**Definition 2.3.2** An argument $a \in \mathcal{A}$ is acceptable with respect to a set $S$ of arguments iff for each argument $b \in \mathcal{A}$ that attacks $a$ there is a $c \in S$ that attacks $b$. In that case, $c$ is said to defend $a$.

**Definition 2.3.3** A set of arguments $S$ is said to be conflict-free if there is no attack within its arguments, i.e. there are no arguments $a, b \in S$ s.t. $a$ attacks $b$. Likewise, a set $S \subseteq \mathcal{A}$ of arguments is admissible iff it is conflict-free and each argument in $S$ is acceptable with respect to $S$.

The notion of extensions [5, 73] presents itself as semantics of acceptance for sets of arguments.
Definition 2.3.4  $S$ is a complete extension of $AF$ iff it is an admissible set and every acceptable argument with respect to $S$ belongs to $S$, i.e., it coincides with the set of acceptable arguments with regard to itself.

Definition 2.3.5  $S$ is a preferred extension of $AF$ iff it is a maximal element (with regard to set inclusion) among the set of admissible subsets of $AF$.

Every preferred extension is a complete extension.

![Diagram](image_url)

Figure 2.4: An example argumentation framework $AF$ with $\mathcal{A} = \{a, b, c, d, e\}$ and $\mathcal{R} = \{(b, c), (b, d), (c, b), (c, d), (c, e), (d, c), (d, b), (e, c)\}$, as seen in Coste-Marquis et al.[70].

The following examples are based on $AF$ (see Figure 2.4). Let $E_1 = \{a\}$, $E_2 = \{a, b, e\}$, $E_3 = \{a, c\}$, $E_4 = \{a, d, e\}$.

$E_1$, $E_2$, $E_3$, and $E_4$ are the complete extensions of $AF$.

$E_2$, $E_3$, and $E_4$ are the preferred extensions of $AF$. Note that $E_1$ cannot be a preferred extension since it is not maximal with regard to set inclusion. In this example, the other extensions are counter-examples: they are admissible sets and are supersets of $E_1$.

Definition 2.3.6  For any type of extension, we can label individual arguments based on their acceptance across multiple extensions of the same kind. If an argument $a \in \mathcal{A}$ is present in all extensions of $AF$ of a given type, we say $a$ is sceptically accepted in those extensions. Conversely, if an argument $a \in \mathcal{A}$ is present in at least one extension of $AF$ of a given type, we say $a$ is credulously accepted in those extensions.

In the example above, we can say $a$ is sceptically accepted across the preferred semantics, since it appears in $E_2$, $E_3$, and $E_4$. Oppositely, $d$ can only exhibit credulous acceptance across the preferred semantics, as it only exists in $E_4$. 
2.3.2 Argumentation in Human-Agent Systems

In terms of human-agent systems, Polberg et al. [74] challenge the assumption that parties always correctly identify the relationships present in an argumentation framework. Their evaluation consists of presenting a dialogical framework (see Section 3.2.1) to participants, where arguments are laid out chronologically, but their attack/support relationships are not displayed. The participants must then define how they believe the statements are related (i.e. building their own argumentation frameworks). Those frameworks are then compared to the original ones, by checking whether their declared graphs contain the ground-truth graphs at a given step.

Cerutti et al. [75] point out that very few studies approach human evaluation of argumentation frameworks from a perspective other than natural language interfaces. They are interested in checking whether humans agree with the reasoning achieved by argumentation systems. They compare extensions (see Definition 2.3.4) obtained by the logical definitions with those found by humans, obtained with a questionnaire regarding the human’s agreement and perceived relevance of each argument in a framework and their overall conclusion regarding which agent is victorious in the dialogue game. Cramer et al. [76] recently extended this experiment to identify specific semantics that are closer to the choices that humans make when presented with a collection of conflicting arguments. They show that grounded and CF2 semantics were the best predictors of human argument evaluation.

2.4 Multi-Agent Path Deconfliction

The problem of multi-agent path deconfliction can be rudimentarily summarised as the following question: who has right of way? In less abecedarian terms, when multiple agents share a limited space and have local incentives (e.g., arriving sooner at their destination), can we find a globally optimal way to resolve path conflicts in such a system?

Many strategies for deconflicting paths have been conceived in the multi-robot/multi-agent systems literature. Centralised
approaches assume the existence of a global entity that coor-
dinates all agents, either coupling all agents optimally (but in-
tractably) in the same configuration space [77], or decoupling
configuration spaces to generate paths independently [78], but
suboptimally and often without a completeness guarantee. Re-
cent developments introduced algorithms with guaranteed com-
pleteness [79, 80].

Decentralised path deconfliction occurs in the absence of such a
global entity, requiring agents to solve conflicts independently.
Their disadvantages are akin to those of decoupled systems,
with the addition of incomplete information in most environ-
ments, which leads to the possibility of deadlocks [81]. On the
other hand, decentralised systems tend to scale well as the com-
putation can be distributed between agents. Another advan-
tage is that a large number of real-world applications operate
in a decentralised way due to communication limitations (e.g.,
autonomous vehicles), and are thus more adequately modelled
by those systems. Since there is no centralised coordinating
agent, agents need to solve local sub-problems in an orderly
manner, often by generating a prioritisation when a conflict is
found. A common way to achieve coordination in decentralised
robot-robot environments is to imbue agents with communi-
cational capabilities. Such mobile agents can then explicitly
convey their intent to other agents [82–85] to solve the con-
lict, with varied degrees of bandwidth availability. Still, there
is no universal protocol for message-passing, and each appli-
cation in literature implements their bespoke convention for
inter-agent data communication, usually as mathematical/nu-
merical abstractions that represent concepts such as vehicle
states/observations [86] [87], intents [88], solutions [81, 89],
or even task re-assignments [90]. All the aforementioned ap-
proaches require all participants in the system to partake in
the same data protocol, which renders most ‘dialogues’ unint-
telligible between agents of different applications (or human
agents).

As mobile agents, humans solve particular instances of this prob-
lem frequently in their everyday lives. Whether it is negoti-
ating junctions with other drivers, stopping for pedestrians at
crossings [91], or even granting passage to others in narrow
corridors, there are implicit heuristics, communication methods,
and cultural norms [92] that come into play at the resolution of

[77]: Erdmann et al. (1986)
[78]: Yu et al. (2017)
[79]: Alotaibi et al. (2018)
[80]: Wiktor et al. (2014)
[81]: Velagapudi et al. (2010)
[82]: Scerri et al. (2007)
[83]: Best et al. (2018)
[84]: Pechoucek et al. (2006)
[85]: Hafner et al. (2013)
[86]: Hoffmann et al. (2008)
[87]: Ayanian et al. (2010)
[88]: Kuwata et al. (2011)
[89]: Velagapudi et al. (2010)
[90]: Dutta et al. (2017)
[91]: Evans et al. (1998)
[92]: Wilde (1980)
In this thesis, we look at how multi-agent path deconfliction can be resolved in rule-aware scenarios via the introduction of culture-based dialogue games. More specifically, we show how explanations are used to improve human performance in those scenarios, especially in cases where the rulesets may vary in complexity.

2.5 Human-Agent Path Deconfliction

Alternatively, other decentralised systems remove the need for explicit communication, instead relying on the ability of inferring intent from other participants and reacting to it [94–96]. Non-communicating systems then require mechanisms to detect and deal with contentious objectives. Perception and estimation can be used to derive models of expected behaviour, which is useful in the presence of human agents [97–100]. Since no explicit interface is required, all sorts of agents may participate in deconfliction scenarios. Thomas et al. [99] recently addressed the problem of path deconfliction at doorways using only motion detection and estimation for human-human, human-robot, and robot-robot environments. However, those systems are only as good as their abilities of accurately inferring and guessing the intention of other agents.

Some real-life applications, such as sea and air traffic control, require explicit statements of intention from agents, as their motion is often hard to be inferred over long distances and visibility constraints. In order to enable explicit communication to happen between heterogeneous agents, a common protocol must be established to allow for meaningful interaction. Natural language is the universal interface for human interaction, thus systems that aspire to achieve seamless explicit human-robot communication must possess some capability of understanding and using natural language for its own benefit. The problem of binding natural language constituents to a robot’s symbolic world model representation is called symbol grounding. Many approaches in symbol grounding are either unidirectional (human to robot) or with limited dialogue capabilities [101]. To our knowledge, there is a dearth of work in the symbol grounding literature that specifically address the problem of potential conflicts [93].
collision avoidance (where goals are not immediately relevant) or path deconfliction (conflicting agents with destinations that must be met).

Ergo, when considering the case for explicit path deconfliction strategies in hybrid environments, it is clear why such solutions would fail to work with human agents. For instance, Dutta et al. [89] show a decentralised deconfliction algorithm that communicates graph matchings between agents to find routing alternatives in the event of conflicting intent. Although computationally practical and efficient, this solution would not work between human agents as-is, for it would either: a) require extensive mental calculation by experts on both sides to process the graph matchings and their underlying consequences for the deconfliction strategy; or b) find an efficient and clear way to describe graph matchings in natural language to non-experts such that it is semantically meaningful towards deconfliction. As humans cannot initially fathom the seemingly abstract and complex numerical/symbolic message-passing representations that arise in commonplace robot-to-robot communications, we are left with the necessity for semantic representations that are compatible with human reasoning [102]. We introduce an example application of human-agent path deconfliction in Chapter 4, where explanations are extracted from a semantic representation of the ruleset.

2.6 Model-Free Reinforcement Learning

The field of Reinforcement Learning (RL) has attracted increased attention in recent years with the success of deep learning in a variety of cognitive tasks. In the simplest case, an agent interacts with a dynamic environment, trying to achieve a reward signal as it learns a policy that maps states and actions to outcomes. During simulation, the agent observes a sequence of states, selects an action, and receives a reward signal at each time step, represented as a scalar. The agent subsequently chooses an action for the next time step based on its policy. The main objective of RL is to design an agent that learns from experience to choose actions that maximise the reward signal received over the course of its lifetime. Influential milestones achieved by DeepMind [3] boosted the AI community’s interest in Reinforcement Learning (RL) algorithms.

[89]: Dutta et al. (2017)

[102]: Chandarana et al. (2017)

[3]: Mnih et al. (2015)
A Reinforcement Learning problem is typically formalised as a Markov Decision Process (MDP). In this setting, an agent interacts at discrete time steps with an external environment. At each time step $t$, the agent observes a state $s_t$ and chooses an action $a_t$ according to some policy $\pi$, that is a mapping (a probability distribution) from states to actions. As a result of its action, the agent obtains a reward $r_t$, and the environment passes to a new state $s' = s_{t+1}$. The process is then iterated until a terminal state is reached.

The agent’s immediate reward $r_t$ is called the reward signal. In addition, the agent receives a discounted sum of these immediate reward signals over the course of its lifetime $R_t$ called the cumulative reward signal. The reward signal is discounted to account for the fact that an action taken at time step $t$ impacts long-term rewards. The main objective of the agent is to maximize its cumulative reward signal. This can be formalised as the minimisation of a cost function $\mathcal{C}_t$ that computes the difference between the cumulative reward signal and the current value of the cumulative reward signal:

$$\mathcal{C}_t(r_t) = R_t - \sum_{i=0}^{t-1} \gamma^i r_i$$

$$= \sum_{i=0}^{t-1} \gamma^i \sum_{s_i} v_{a_i}(s_i, s_{i+1})$$

$$= \sum_{s_i} \gamma^t(s_i, s_{i+1})$$

The cost function is a function of the immediate reward signal $r_t$. The cost function is then learned by the agent to yield an optimal action policy $\pi^*$ that maximizes the cumulative reward signal. The discount rate $\gamma > 0$ plays a key role in determining the reward signal. In the simplest case, the discount rate is a constant and is set to $1/\gamma$. The objective function is then equal to the difference between the discounted cumulative reward signal, denoted by $R_{t+1}$, and the current value of the cumulative reward signal, $R_t$. 

Choosing a value of $\gamma$ typically depends on the problem studied. A lower value of $\gamma$ results in a higher discount factor and thus in a higher value of the reward signal. The objective function is then maximized by choosing actions that lead to a large reward later in time.

The goal of the agent is to maximise the expected cumulative return starting from an initial state $s = s_t$. The action value $Q^\pi(s, a) = E_{\pi}[R_t | s = s_t, a = a_t]$ is the expected return for selecting action $a$ in state $s_t$ and following with strategy $\pi$. Given a state $s$ and an action $a$, the optimal action value function $Q^*(s, a) = \max_\pi Q^\pi(s, a)$ is the best possible action value achievable by any policy. Similarly, the value of state $s$ given a policy $\pi$ is $V^\pi(s) = E_{\pi}[R_t | s = s_t]$ and the optimal value function is $V^*(s) = \max_\pi V^\pi(s)$.

Two fundamental dynamic programming algorithms support the methodology of RL: value iteration and policy iteration. In the value-based approach, one defines the parameters of a value function that quantifies the maximum cumulative reward obtainable from a state belonging to the observation space. Value-based algorithms, such as Deep Q-Networks (DQN) [3], use temporal difference learning, where policy extraction is done after an optimal value function is found. Alternatively, in the policy-based approach, the policy parameters are tuned in a direction of improvement. Value iteration and policy iteration approaches can also be combined. The algorithms using a combination of them are usually mentioned with the umbrella term Actor-Critic methods. Actor-critic methods, such as Twin-Delayed DDPG (TD3) [103] and Soft Actor-Critic (SAC) [104], rely on evaluating and improving a policy (via gradient descent) together with a state-value function.

In the case of deep RL, the agent’s policy is represented by a deep neural network, mapping input states to actions. DQN is one of the very first value-based deep RL algorithms, designed to work on discrete action-spaces only. It learns to approximate the function that associates values to actions, so that from a
state $s$ the action with the best value can be selected through the operator max. Adaptations for continuous action-spaces, as DDPG and then TD3, propose to address value overestimation problems by means of clipped double Q-learning, delayed update of target and policy networks, and target policy smoothing. However, one of the main limitations with TD3 is that it randomly samples actions using a pre-defined distribution.

To overcome the issue of being limited by a fixed distribution, Soft Actor Critic (SAC) empowers the agent with the ability to also learn the distribution with which to sample actions, empowering the agent to explore more different strategies through entropy maximisation. SAC is an algorithm based on the maximum entropy reinforcement learning framework. In it, randomness is encouraged by establishing the entropy of a policy into the objective, combining off-policy updates with a stable stochastic actor-critic formulation. The actor aims to maximise expected reward while also maximising entropy, i.e., to succeed at the task while acting as randomly as possible.

We consider the notion of explanations for RL in this thesis as a method to enhance the performance of algorithms that benefit from experience replay, by means of organising the experience space in an explanation-based manner. The next section covers relevant work on that area.

### 2.7 Explanations for RL

Recent efforts in Explainable AI (XAI) \cite{BarredoArrieta2020} set out grounds towards a newfound development of AI systems capable of generating explainable information regarding their decisions. Two important types of explanations are rule-based explanations \cite{Branting1991} and case-based (or example-based) explanations \cite{Aamodt1994}. In rule-based explanatory models, the system has explicit knowledge of the underlying rules’ mechanism and will specify which rule acted as a decision boundary to promote the given outcome. Contrarily, case-based models do not share this assumption, and instead elicit historically-similar past events to justify the present decision \cite{Waa2021}. Within this domain, explanations are commonly conveyed via textual/visual descriptive representations of the decision criteria (i.e. rule-based), or with similar examples (i.e. case-based). An example of rule-based explanation...
is ‘you will get a penalty for reaching 75, which is above the speed limit of 50’, based on the rule ‘if speed is above 50, you will get a penalty’. While an example of case-based explanation is ‘you get a penalty because you are in a situation similar to this other vehicle that reached speed 74 and was previously penalised’.

Dietterich et al. [107] frame explanation-based RL as a case-based explanatory process where prototypical trajectories of state-transitions are used to tackle similar but unseen situations, while Chow et al. [108] implement a rule-based method, constraining the Markov Decision Process by means of Lyapunov functions. Many rule-based methods for explaining to RL agents usually fall under the umbrella of a sub-discipline called Safe RL [109]. Safe RL includes techniques for both encoding rules in the optimality criterion [108, 110] and incorporating such external knowledge into the action/state space [111]. Although not generating explicit explanations, those methods engineer safety rules into the learning process, implicitly explaining to the agent what not to do. Alternatively, an important example of case-based methods for explaining to RL agents is Imitation Learning [112], where demonstrations (as trajectories of state-transitions generated by a human or expert algorithms) are used to train the RL agent. These can be seen as high-quality cases/examples provided by an expert human or algorithm. However, access to human expert data may not scale well to every domain, and not all problems dispose of accessible expert algorithms.

We are interested in sampling the most useful experiences to cover a particular agent’s gap in knowledge. An agent-centred explanatory process is an iterative process that follows the agent through the process of learning and selects the most useful explanations for it, at every time-step. The notion of ‘usefulness’ in this case is usually defined by how much the experiences defined by that explanation can contribute to the agent’s learning. Below, we look at how experience replay techniques tackle this issue in off-policy RL.

### 2.8 Prioritised Experience Replay

We begin by explaining the difference between on-policy and off-policy algorithms. On-policy algorithms are algorithms that
follow the actual policy of the agent. They sample from the state-action space of the agent’s current policy and are evaluated by comparing their performance against the optimal performance of the policy. The main advantage of on-policy RL is that it allows the agent to modify its policy when it is off-policy (i.e. during training), leading to a more robust policy. On-policy methods are thus the current state-of-the-art in reinforcement learning.

Off-policy algorithms, on the other hand, are algorithms that follow a different policy that is not the one produced by the agent. Off-policy methods are not affected by policy changes and are thus not affected by gradual changes in policy that are common in on-policy methods. They are evaluated by comparing their performance against that of the agent with the optimal policy. Off-policy methods are usually used for exploratory RL, where it is acceptable to follow a sub-optimal policy.

In the case of on-policy methods, the space of experiences for each state-action pair is usually sampled uniformly because on-policy methods are designed to exploit the policy gradient. This means that the algorithm is constantly adjusting its actions in order to improve its expected reward. This adjustment is based on the experience of the agent, so it is important that the space of experiences be sampled uniformly in order to ensure that the algorithm is learning without bias, i.e., all experiences are sampled with equal probability. However, the fact that some experiences are more useful than others is what makes experience replay a concern. In off-policy methods, we are interested in selecting the most useful experiences (cases) to replay, in order to achieve the highest performance.

Off-policy algorithms such as DQN, TD3, and SAC aim to find a policy that maximises the cumulative return, by keeping and learning from a set of expected returns estimates for some past policy $\pi$. This set of expected returns is kept in an experience buffer, enabling experience replay. Experience replay \cite{9} consists in re-utilising information from the space of sampled experiences. The agent’s experiences at each time-step $t$ are stored as transitions $e_t = (s_t, a_t, r_t, s_{t+1})$, where $s_t$, $a_t$, $r_t$ represent the state, action, and reward at time $t$, followed by the next state $s_{t+1}$. These transitions are pooled over many episodes into a replay memory, which is usually randomly sampled for a mini-batch of experiences.
Experience sampling can be improved by differentiating important transitions from unimportant ones. In Prioritised Experience Replay (PER) \cite{9}, the importance of transitions with high expected learning value is measured by the magnitude of their temporal-difference (TD) error. Experiences with larger TD are sampled more frequently, as TD quantifies the unexpectedness of a given transition \cite{113}. This prioritisation can lead to a loss of diversity and introduce biases. Bias in prioritised experience replay occurs when the distribution is changed without control. This effect therefore changes the solution that the estimates will converge to. This bias can be corrected through importance-sampling (IS) weights.

Many approaches to Prioritised Experience Replay (PER) in RL \cite{9} can be re-framed as mechanisms for achieving some form of agent-centrality, re-ordering experience by relevance in the attempt of explaining to the agent and selecting the most useful experience, as indirectly suggested by Li et al. \cite{113}. Over the years, many human-inspired intuitions behind PER drove researchers towards improved, more sophisticated and agent-centred mechanisms to RL \cite{114–116}. Among these works, the closest to a fully agent-centred explanatory process is Experience Replay Optimisation \cite{116}, which moves towards agent-centrality by providing an external black-box mechanism (or experience sampler) for extracting arbitrary sequences of information out of a flat (no abstraction involved) experience buffer. The experience sampler is trained to select the most ‘useful’ ones for the learning agent. However, due to its non-explainable nature, it is not clear whether the benefits given by Experience Replay Optimisation are due to the overhead given by the experience sampler increasing the number of neurons in the agent’s network.

Another work trying to achieve agent-centrality in this sense is Attentive Experience Replay \cite{114}, suggesting to prioritise uncommon experience that is also on-distribution (related to the agent’s current task). However this work, as the previous one, also falls short of explicitly organising experience in an abstract-enough way by conveying human-readable explanations to the agent. Hierarchical Experience Replay \cite{115} has attempted to address the abstraction issue in an attempt to simplify the task to the agent, decomposing it into sub-tasks. However, they do not do so in an agent-centred and goal-oriented
way, given that its sub-task selection is uniform and not curricular. Yin et al. [115] tries to organise experience hierarchically by implementing an abstraction heuristic. It does it in the attempt to simplify the task to the agent, decomposing it into sub-tasks, even if not in an agent-centred and goal-oriented way, considering that sub-task selection is uniform and not curricular.

On the other hand, a curricular approach for training RL agents was proposed by Ren et al. [117], exploiting PER and the intuition that simplicity is inversely proportional to TD-errors, but not exploiting any abstract and hierarchical representation of tasks. Similarly to ours, Sovrano [118] aims to organise experience abstractly, based on its explanatory content — framed as the ability to answer how good or bad a sequence of state-transitions is with respect to average experience. This work only considers explanations about the immediate performance of the agent (i.e. HOW explanations), and lacks any consideration of other and richer types (i.e. WHY), as well as curricular prioritisation facilities.

The work in Chapter 6 investigates how the application of an explainer mechanism such as cultures can enable existing state-of-the-art RL algorithms to learn more efficiently, especially in the presence of hard rulesets that contain many exceptions. We utilise HOW and WHY explanations as above to explore performance-based and domain-based ways for organising the replay space, respectively.
Summary

In this chapter, we visit the literature incorporating relevant research work to this thesis, navigating the fields of Explainability, Computational Argumentation, Multi-Agent Deconfliction, and Reinforcement Learning. The main conclusions of this chapter are:

- Explanations are the product of explainable systems, and these can be used as a medium for exchanging valid information between humans and agents. The necessity of explanations in human-agent systems is grounded in acceptance and trust. We argue that system complexity is another valid dimension to be observed.

- Argumentation Frameworks are a powerful mechanism to underpin explainable systems meant to be compatible between humans and agents, as it maps well to human reasoning and can be used by machine agents to generate explanations.

- Reinforcement Learning agents learn by amassing experiences associated with rewards. Relevant episodes can be replayed to improve the performance of those agents. Choosing relevant episodes is analogous to providing case-based explanations. We indicate that explainable systems can enhance this technique by organising the replay space in terms of explanations.
A culture is a set of shared beliefs, values, and norms that shape the way agents interact with their environment. For human and machine agents alike, the ability to justify their decision based on such beliefs, values and norms accounts for the explainable part of their agency. The ability to adhere to shared beliefs, values and norms is a rule-aware capability.

We argue that the use of cultures, or rule-aware agency, can be beneficial in situations where agents need to justify their actions to humans through explanations. We present an example where agents can generate explanations for their actions by providing evidence for the rules that they used to make their decisions. We believe that this capability will allow agents to build trust with humans by providing them with the ability to understand, verify, and critique the rules that agents use to make their decisions.

Our first research question stated: ‘can humans and autonomous agents exchange explanations and reason collectively about their state and observations?’ In this chapter, we argue that establishing a common culture is the ground through which agents can exchange explanations and justify actions during conflict resolution. The use of cultures, or rule-aware agency, can be beneficial in situations where agents need to justify their actions to humans through explanations. We present an example where agents can generate explanations for their actions by providing evidence for the rules that they used to make their decisions. Here, we introduce definitions and concepts that are used in the construction of our architecture. Throughout this section,
we will refer to concepts in Section 2.3.1, which introduces essential definitions for Abstract Argumentation frameworks, the principal deliberation tool in our framework.

### 3.1 Explainable Conflict Resolution

We introduce an architecture for explainable conflict resolution (cultures) as a mechanism that provides explainable deliberation capabilities for dialectic interactions between agents. Below, we elaborate on definitions and concepts that were created for the purpose of this application.

**Example 3.1.1** Suppose the following situation: vehicle $A$ crosses a green light in a junction and is about to collide with vehicle $B$, who ran a red light. In most highway codes, the rule ‘*a vehicle shall not cross the stop line on a red light*’ can be ignored if rule ‘*a vehicle may cross the stop line on a red light if it is an emergency vehicle*’ also applies to that situation. Therefore, in this specific situation, the right of way can be determined by $B$’s status: if it is an emergency vehicle, then it could refer to the aforementioned rule and argue in favour of its right of way. Likewise, if $B$ is not an emergency vehicle, then it would not be able to defeat $A$’s claim of the first rule and $B$ would find itself at fault.

Consequently, what would happen in Example 3.1.1 if either $A$ or $B$ is an artificial autonomous agent? Would the human counterpart benefit more from being explicitly told which rules are being used, or is having an implicit knowledge sufficient? We can rather evidently demonstrate that there is no challenge in having autonomous agents follow rule-based systems if the rules are encoded in the agent’s behaviour. Instead, our architecture aims to create a direct mapping between rulesets in human-readable form and corresponding argumentation frameworks.

For our problem domain, we assume a setting where agents perform localised decision-making. When acting, we are interested in ensuring that each agent’s behaviour is compliant with an overall culture (represented by an argumentation framework $AF$) shared amongst all participants in the system. In order to check which rules apply in a specific event of a conflict, we
introduce a mechanism of argument verification (Section 3.3.2). Finally, after agents and humans share a common model and can provide evidence for their rule-compliant justification, we demonstrate how to build explanations from this framework in Section 3.3.3.

3.2 Background

In this section, we will introduce related work that is drawn upon for the building of our theoretical framework of cultures. This builds on the definitions of Argumentation Frameworks seen in Section 2.3.1.

3.2.1 Dialogue Games

The extension semantics introduced by Dung [5] are powerful in asserting global properties of the argumentation framework, but their output is static and monological in nature.

In pursuance of a more dialogical approach [119], one must consider the dynamics of dialogue and the assumptions therewithin. Using Jakobovits and Vermeir’s position framework formalism [120], 'the combination of a set of rules that govern the game, and the determination of winning criteria, constitute a dialectic semantics for the “theory” that underlies the player’s arguments.’ We will adapt some of the definitions from Jakobovits et al. [120], as follows.

**Definition 3.2.1** A position framework paired with an argumentation framework \( AF = (\mathcal{A}, \mathcal{R}) \) is an argumentation framework \( PF = (\mathcal{P}, \mathcal{R}^\ast) \), where \( \mathcal{P} \) consists of conflict-free subsets of \( \mathcal{A} \), and \( \mathcal{R}^\ast \) denotes the set of finite sequences of elements from \( \mathcal{R} \). Elements of \( \mathcal{P} \) are called positions.

**Definition 3.2.2** A player \( c \) can be categorised as the proponent \( (p) \) or opponent \( (o) \). The adversary of \( p \) is denoted \( \overline{p} = o \). Conversely, \( \overline{o} = p \).
**Definition 3.2.3** Let a player \( c \in \{p, o\} \) and a position \( X \in \mathcal{P} \). A **move** in \( \mathcal{P} \) is a pair \((c, X)\). For a move \( m = (c, X) \), we use \( \text{player}(m) \) to denote \( c \) and \( \text{pos}(m) \) to denote \( X \).

**Definition 3.2.4** A dialogue type is a tuple \((\mathcal{P}, \mathcal{R}^*, \phi)\), where \((\mathcal{P}, \mathcal{R}^*)\) is a position framework and \( \phi : \mathcal{P}^* \to 2^\mathcal{P} \) is a legal-move function, where \( \mathcal{P}^* \) denotes the set of finite sequences of elements from \( \mathcal{P} \). A dialogue \( D \) in \((\mathcal{P}, \mathcal{R}^*, \phi)\) is any finite sequence \( d_0, d_1, \ldots, d_n \) of moves in \( \mathcal{P} \) that satisfies, for \( 0 \leq i < n \):

1. \( \text{player}(d_{i+1}) = \text{player}(d_i) \), i.e. the players take turns.
2. \( \text{pos}(d_{i+1}) \in \phi(\text{pos}(d_0) \ldots \text{pos}(d_i)) \), i.e. the next move is legal.
3. \( d_{i+1} \notin \{d_0, d_1, \ldots, d_i\} \), i.e. a move cannot be repeated twice.
4. \( \text{attacks}(\text{pos}(d_{i+1}), \text{pos}(d_i)) \), the player’s move attacks the adversary’s previous move
5. \( \text{player}(d_0) = p \), i.e. the proponent makes the first move.

The dialogue \( D \) is said to be **about** the position \( \text{pos}(d_0) \).

**Definition 3.2.5** Let \( X^- \) and \( X^+ \) denote the sets of positions that attack and are attacked by \( X \in \mathcal{P} \), respectively. A player \( c \) is said to **win** \( D \) if \( D \) is finite and ends with a move \((c, X)\) s.t. \( X^- \cap \phi(D) = \emptyset \), i.e., the dialogue cannot be continued.

**Definition 3.2.6** Let \((\mathcal{P}, \mathcal{R}^*)\) be a position framework. The legal-move function \( \psi(\mathcal{P}, \mathcal{R}^*) : \mathcal{P}^* \to 2^\mathcal{P} \) which disallows self-defeating and useless moves in \((\mathcal{P}, \mathcal{R}^*)\) is defined as follows: for all \( Y_0, \ldots, Y_i \in \mathcal{P}^* \):

\[
\psi(\mathcal{P}, \mathcal{R}^*)(Y_0, \ldots, Y_i) = \mathcal{P} \setminus (\{X | \text{self-defeating} \} \cup \bigcup_{j=0}^{i} Y_j^+) \]

We now have sufficient tools to formalise types of dialogues that encompass the previously chosen rules, with single or multiple arguments per move:
3.2 Background

Definition 3.2.7 Let $AF = (A, R)$ be an argumentation framework. A useful-single-argument dialogue in $AF$ is a dialogue in the dialogue type $(A', R, \psi(A', R))$, where $A' = \{a \mid a \in A\}$ and $\psi(A', R)$ designates moves that are not self-defeating nor useless. The dialogue type $(P, R, \psi(P, R))$ is called the useful-multiple-argument dialogue type, where $P$ is the set of conflict-free subsets of $A$.

3.2.2 Explanations from Argumentation

Fan et al. [58] propose an argumentation semantics aimed at generating explanations. This formalism promotes the notion of explanations as sets of arguments, taking into consideration which arguments contribute to the justification (or $r$-defence) of a specific premise (argument). We utilise their definitions for our framework, as follows.

Definition 3.2.8 Given an AA framework $AF = (A, R)$, let $a, b \in A$. $a$ $r$-defends $b$ iff:
1. $a = b$; or
2. $\exists z \in A$, s.t. $a$ attacks $z$ and $z$ attacks $b$; or
3. $\exists z \in A$, s.t. $a$ $r$-defends $z$ and $z$ $r$-defends $b$.

$S \subseteq A$ $r$-defends $a \in A$ iff $\forall b \in S$: $b$ $r$-defends $a$.

Definition 3.2.9 A set of arguments $S \subseteq A$ is related admissible iff $\exists a \in S$ s.t. $S$ $r$-defends $a$ and $S$ is admissible. $a$ is said to be a topic of $S$. For any argument $a \in A$, an explanation of $a$ is $S \subseteq A$ s.t. $S$ is a related admissible set and $a$ is a topic of $S$.

Their definition of explanations is further characterised by a classification with regards to cardinality and set inclusion:

Definition 3.2.10 Let $a \in A$ and $E_a$ be the set of all possible explanations of $a$. For every $S \in E_a$, we say $S$ is a minimal or maximal explanation iff $S$ is the smallest or largest subset of $E_a$ with regards to cardinality, respectively. Similarly, $S$ is a compact or a verbose explanation iff $S$ is the smallest or largest subset of $E_a$ with regards to set inclusion, respectively.
3.3 Cultures

Orderly behaviour can only happen if all agents share common guidelines and understand the same rules. We define the notion of culture as a collective agreement of beliefs, values, norms, and priorities, represented by an argumentation framework.

**Definition 3.3.1** Let any two players $p$, $o$ be the proponent and opponent in a dialogue game $D$. We say a proposition is any argument $a \in A$ that may be used by proponent $p$ to request a contended resource from opponent $o$. The proposition is the argument that settles the dispute. If the proposition is successfully validated by $p$ against $o$, $p$ wins the dialogue game and the contended resource/dispute. Conversely, if $p$ fails to defend the proposition, $o$ will win the dispute.

**Definition 3.3.2** Let $A$ be a set of arguments and $R$ be the set of attack relationships in $A$. Let $K \subseteq A$ be the set of all propositions in $A$. We say a system has a culture $C = (A, R, K)$ iff $|K| > 0$ and it is used by all agents in the system.

**Example 3.3.1** A simple example would be: suppose a culture that contains three arguments, $A = \{\mu, \alpha, \beta\}$, where $\mu$ represents the proposition ‘I have right of way’, $\alpha$ represents ‘I am an ambulance’ and $\beta$ represents ‘I am a fire rescue truck’. Defining $R_a = \{(\alpha, \mu), (\beta, \mu), (\alpha, \beta)\}$ is akin to defining that, in this application, ambulances have priority over fire trucks. Conversely, defining $R_b = \{(\alpha, \mu), (\beta, \mu), (\beta, \alpha)\}$ would mean the opposite. Despite having the same argument set $A$, $C_a = (A, R_a, K)$ is a different culture from $C_b = (A, R_b, K)$.

3.3.1 Propositional Dialogues

When agents are presented with conflicts that require a compliant resolution (with regards to the culture), a dialogue game starts from the proponent $p$. Each agent then takes turns in choosing arguments that are potentially able to defeat the previous, as shown in Definition 3.2.4. Agents can use one or multiple arguments at each turn depending on whether it is a useful-single or useful-multiple-argument (Definition 3.2.7) dialogue type. The game ends when one agent provides an argu-
3.3 Cultures

We extend the set of requisites for a dialogue seen in Definition 3.2.4 and introduce the idea of a propositional dialogue:

**Definition 3.3.3** Let \( D = \{d_0, \ldots, d_i\} \) be a dialogue in a position framework \( \text{PF} = (\mathcal{P}, \mathcal{R}^*, \phi) \) paired with a culture \( C = (\mathcal{A}, \mathcal{R}, \mathcal{K}) \). We say \( D \) is a propositional dialogue iff \( D \) is about a position \( \text{pos}(d_0) \) where \( \exists a \in \text{pos}(d_0) \) s.t. \( a \) is a proposition.

**Definition 3.3.4** Let \( D = \{d_0, \ldots, d_i\} \) be a propositional dialogue. We denote \( d_0 \) as the motion of the dialogue.

The player who wins \( D \) then takes priority or ownership with regards to the contended resource disputed in the proposition. However, as traditional argumentation frameworks may not react to dynamic changes in the environment, cultures need to cater to most changes in circumstances. In Example 3.3.1, the ability of an agent-player \( c \) using \( \alpha \) as an argument to defeat \( \mu \) depends exclusively on the fact of \( c \) being, in fact, an ambulance. We introduce the concept of argument verification to deal with this matter.

3.3.2 Argument Verification

Most applications of argumentation frameworks consider frameworks as static, i.e., the combination of all the arguments may generate an extension or labelling that represents an insight about which arguments should or should not be admitted. In our case, the culture denotes a ruleset that does not account for a specific happenstance (for example, a unique and specific conflict resolution instance), but rather a more comprehensive model that encompasses different future scenarios. For that purpose, every agent has to verify which arguments are valid at a particular moment. We propose the architecture of argument verification to address the issue of factual correctness and validity of a specific argument given the context of its proponent.

The verification of arguments can be modelled as decision problems:
Definition 3.3.5 Let $\alpha \in \mathcal{A}$ and $c \in \{p, o\}$ be an argument and a player, respectively. We denote $\alpha$ as demonstrable by agent-player $c$ iff checking the correctness of that argument admits a finite and computable\(^1\) decision procedure.

This decision procedure can be represented by a predicate function that evaluates, in the current context, whether a specific argument may be used or not.

Definition 3.3.6 Let $\zeta$ denote the set of all possible contexts in the environment. $\forall \alpha \in \mathcal{A}$, $\alpha$ admits a predicate function $f_\alpha : c, z \rightarrow \{\text{True}, \text{False}\}$, where $c$ represents the player and $z \in \zeta$ is a context. We say $f_\alpha$ is the verifier function of argument $\alpha$. A special case applies for propositions, as they are hypothetical and cannot be checked for correctness.

Corollary 3.3.1 The verifier function of every proposition always returns True.

Definition 3.3.7 Let $\alpha \in \mathcal{A}$ be an argument. Let $c \in \{p, o\}$ be a player and $z \in \zeta$ a context. We say argument $\alpha$ is demonstrably true by player $c$ iff $f_\alpha(c, z) = \text{True}$.

Definition 3.3.8 Let $D = \{d_0, \ldots, d_i\}$ be a dialogue. Let $c \in \{p, o\}$ be a player, $z \in \zeta$ a context, and $X \in \mathcal{P}$ be a position. We say $d_{i+1} = (c, X)$ is a verified move iff $\forall a \in \text{pos}(d_{i+1}), f_a(c, z) = \text{True}$.

Note that demonstrably true arguments do not mean that they are universally true – not even that they are true at all. All it means is that an agent will be able to compute a procedure to check if that statement stands against its own knowledge in the current context. The notion of demonstrably true is, in fact, a local definition of truth, as it only requires the perception of a single agent, even if the agent is mistaken/uneducated about the world (such as in systems with imperfect/incomplete information.)

All deliberation in this system is delegated into the specifics of the predicate functions that accompany each argument in the system. Hence, the design of a system using cultures is divided in two phases (see Figure 3.1):
3.3 Cultures

- Mapping rules into pairs of arguments and verifier predicate functions;
- Establishing attack relationships between generated arguments.

As every rule is represented by an argument, tracing the history of exchanged arguments in this manner provides an insight on each agent’s attempt to justify their prioritisation based on the ruleset provided. We facilitate the interaction with humans by providing explanations to justify the results of the dialogue game.

### 3.3.3 Explanation Generation

Cultures do not bind agents into a specific strategy for choosing moves in a dialogue game. The justification is that humans cannot be bound to a unique way of thinking, or be programmable as an artificial agent can. Therefore, we allow agents to freely choose their (verified) moves and focus on generating post-hoc explanations derived from the history of a dialogue $D$. For that purpose, we propose some mechanisms for generating explanations using cultures, based on the definitions seen in Fan et al. [58].

**Definition 3.3.9** Let $D = \{d_0, ..., d_i\}$ be a completed dialogue and $d_i$ be the winner move. We say a set $W \subseteq D$ is the set of winning moves where $W = \{d_k \in D \mid \text{player}(d_k) = \text{player}(d_i)\}$. The set of losing moves is denoted by $L = D \setminus W$.

**Definition 3.3.10** Let $d_i$ be the winner move in $D$. An explanation $E_D$ of $D$ is defined as $E_D \subseteq D$ s.t. $\exists X = d_i$ and $X \in E_D$, i.e.,
it always contains the winner move. We denote \( \mathcal{E}_D \) as the set of all possible explanations of a dialogue \( D \).

We can create a notion of contrastive explanations to include losing moves. The idea behind contrastive explanations is to provide extra justification as to why a specific argument was not accepted. We denote explanations without losing moves as plain explanations.

**Definition 3.3.11** Let \( d_i \) be the winning move in \( D \). A contrastive explanation \( CE_D \) of \( D \) is defined as \( CE_D \in \mathcal{E}_D \) s.t. \( \exists X = d_i, \exists Y \in L, \) and \( X, Y \in CE_D \). A plain explanation \( PE_D \) of \( D \) is defined as \( PE_D \in \mathcal{E}_D \) s.t. \( \forall X \in PE_D, X \in W \).

**Definition 3.3.12** Adapted from Definition 3.2.10. Let \( \mathcal{E}_D \) be the set of all possible explanations of \( D \). We therefore say that, for any \( S \in \mathcal{E}_D \), \( S \) is a: minimal or maximal explanation iff \( S \) is a smallest or largest subset of \( \mathcal{E}_D \) with regards to cardinality, respectively. \( S \) is a compact or a verbose explanation iff \( S \) is a smallest or largest subset of \( \mathcal{E}_D \) with regards to set inclusion, respectively. In the case of single-argument dialogue types, the compact explanation is always the winner move, and the verbose explanation is the entire dialogue.

One could observe the entire footprint of uttered arguments and generate an explanation by writing all their natural language representations, but this approach is too verbose and unwieldy in most cases (especially if agents operate under useful single-argument dialogue rules). We can specify a bound on the number of positions chosen to support an explanation (see Example 3.3.2).

**Definition 3.3.13** We say \( E' \in \mathcal{E}_D \) is an \( n \)-reason explanation iff \( |E'| = n \).

**Example 3.3.2** Let \( p \) and \( o \) be the proponent and the opponent, respectively. Let \( D = \{ p_0, o_1, p_2, o_3, p_4 \} \) be a complete dialogue and \( p_4 \) be the winner move. A 3-reason contrastive explanation could be \( E_D = \{ p_2, o_3, p_4 \} \). A 2-reason plain explanation could be \( E'_D = \{ p_2, p_4 \} \).
3.4 Privacy-aware Cultures

With the definitions laid out above, we set the scenario where cultures enable explainable conflict resolution in multi-agent norm-aware environments. Notwithstanding its abstract nature, this architecture relies on two important assumptions: i) that agents have complete information about themselves and other agents; and ii) that dialogues extend indefinitely until an agent is cornered out of arguments – thus being convinced to concede. In most real-life applications, however, those assumptions occur rather infrequently. Fully decentralised agents often rely on local observations and communication to compose partial representations of the world state, and indefinite dialogues are both practically and computationally restrictive. We build on it by extending this architecture to support gradual disclosure of information and privacy restrictions.

Example 3.4.1 Suppose 2 agents, Belle and Cadence, are disputing a priority seat on public transport. They share a culture $C = (\mathcal{A}, \mathcal{R}, \mathcal{K})$ where $\mathcal{A} = \{\gamma, a, b\}$ contains the arguments

- $(\text{proposition}) \gamma \equiv 'I\ think\ I\ should\ have\ this\ seat.'$
- $a \equiv 'I\ am\ older\ than\ you.'$
- $b \equiv 'I\ have\ a\ more\ serious\ health\ condition.'$

The relations $\mathcal{R} = \{(a, \gamma), (b, \gamma), (b, a)\}$ determine the priority of those concepts in the agents’ shared culture. If Belle and Cadence do not know each other’s age and health condition, they will have no information to verify arguments $a$ and $b$, i.e., they cannot argue that they are demonstrably older or more infirm than their adversary. At this stage, any of those arguments could only be raised as a hypothesis. Additionally, both Belle and Cadence would need to break privacy and mutually reveal some personal information to reach a decision.

To allow agents to engage in dialogues under partial information and privacy constraints, we present a novel explainable conflict resolution architecture. We begin by looking at the aspect of alteroceptive cultures, our proposed mechanism organising the space of interactions for privacy-aware conflict resolution.

2: ‘Alteroceptive’ is our coinage from a portmanteau of the Latin word alter (other) + the word reception. It shall connote a ‘sense of other.’
3.4.1 Alteroceptive Cultures

As shown in Example 3.4.1, when no information is available and arguments are comparative, agents can only raise hypotheses. Those serve as media for sharing information between agents, by enforcing that every agent shares their corresponding piece of information when eliciting a hypothesis, in order to avoid infinite exchanges arising from cycles of empty suppositions.

Once an agent raises a hypothesis and shares their pertaining information regarding an argument, the adversary should have enough information to effectively verify the argument into a fact, since it has full knowledge of its own description and was given the other agent’s partial description. This can be done without necessarily sharing further information.

Additionally, the relations between arguments need to be woven in such a way as to guarantee that hypotheses and facts only defeat the arguments pertaining to the adversary. In Example 3.4.1, if the same agent wants (and is able) to utter both arguments \( a \) and \( b \), the attack \((b, a)\) in the culture should not yield an actual attack in an instantiation if both arguments are articulated by the same agent. After all, having a trumping condition would be yet another reason in favour of – not against – the agent’s claim. Hence, assuring that attacks only occur between adversaries renders a bipartition in the culture and its arguments. The conjunction of those elements culminates in an alteroceptive culture.

Definition 3.4.1 (Alteroceptive Culture) Let \( C = (\mathcal{A}, \mathcal{R}, \mathcal{K}) \) be a culture. We define the expansion function \( \kappa \), that maps an argument \( a \in \mathcal{A} \) to a set of new arguments (not in \( \mathcal{A} \)). For every \( a \in \mathcal{A} \setminus \mathcal{K} \), \( \kappa(a) = \{a_{pr}^H, o_p^H, a_{pr}^F, o_p^F\} \) amounts to four new arguments that represent, respectively:

- \( a_{pr}^H \): hypothesis (\( H \)) of \( a \) where proponent \( pr \) wins,
- \( a_{op}^H \): hypothesis (\( H \)) of \( a \) where opponent \( op \) wins,
- \( a_{pr}^F \): verified-fact (\( F \)) of \( a \) where proponent \( pr \) wins,
- \( a_{op}^F \): verified-fact (\( F \)) of \( a \) where opponent \( op \) wins.
For propositions in $\gamma \in \mathcal{X}$, $\kappa(\gamma) = \{\gamma_H^p, \gamma_H^o\}$, since propositions do not admit verification. We define the expanded set of arguments (in an alteroceptive culture) $\mathcal{A}_x = \bigcup_{a \in \mathcal{A}} \kappa(a)$. The expanded set of attacks $\mathcal{R}_x$ is constructed from $\mathcal{R}$ and $\mathcal{A}$, and satisfies the following rules (see Figure 3.2). For each element of $\kappa(a)$, where $a, b \in \mathcal{A}$ and $(a, b) \in \mathcal{R}$. For all $w \in \{pr, op\}$:

1. $(a_H^w, a_H^{\overline{w}}) \in \mathcal{R}_x$ if $a \notin \mathcal{X}$, i.e., non-proposition hypotheses mutually attack each other;

2. $(a_F^w, a_F^{\overline{w}}) \in \mathcal{R}_x$, i.e., verified-facts mutually attack each other;

3. $(a_F^w, a_H^{\overline{w}}) \in \mathcal{R}_x$ i.e., each verified-fact attacks their adversary’s hypothesis;

4. $(a_H^w, b_H^{\overline{w}}) \in \mathcal{R}_x$ and $(a_F^w, b_F^{\overline{w}}) \in \mathcal{R}_x$, i.e., hypotheses reproduce their original attacks to both hypotheses and verified-facts;

5. there are no more elements in $\mathcal{R}_x$.

We say $C_x(C) = (\mathcal{A}_x, \mathcal{R}_x)$ is an alteroceptive expansion of $C$.

The separation between hypotheses and facts is important to introduce a mechanism for gradual disclosure of information.

Example 3.4.2 (cont. 3.4.1) Given the original culture seen in Example 3.4.1, we can generate an extended set of arguments $\mathcal{A}_x = \bigcup_{\gamma \in \mathcal{X}} \kappa(\gamma), \kappa(a), \kappa(b)$ and the corresponding $\mathcal{R}_x$ as described in Def. 3.4.1. Suppose, then, that Cadence raises the motion $y_{\text{Cad}}^{\text{H}} \equiv \text{‘I think I should have this seat,’}$ to which Belle promptly replies with $a_{\text{Belle}}^{\text{H}} \equiv \text{‘I may be older than you, I am}$.

Figure 3.2: Alteroceptive cultures transform the arguments of a given culture (grey) into hypotheses (dashed nodes) and verified-facts (solid nodes). Shown here from Example 3.4.1, argument $a$, "Is older", is transformed into the respective nodes for the orange opponent, $a_F^p$ and $a_F^o$, and the blue proponent, $a_H^p$ and $a_H^o$. $A$ and $B$ represent the agents in the dispute. The inbound attack (left) from "Has worse health" attacks all four nodes and the outbound attacks (top) attack "Deserves seat" from the hypotheses.
60 years old.’ Figure 3.2 illustrates the expanded argument set in question. Note that Belle had to break privacy and reveal their age in order to use this argument. Cadence now has the necessary information (since they know their own age) to verify argument $a_{CAD}^F$. If they succeed in verifying this argument (i.e., they are older than Belle), they have two options for rebuttal, depending on what information they intend to share:

1. Use $a_{CAD}^F$, reveal their age and refute Belle’s claim of being older;
2. or use $b_{CAD}^H$, ignore the age dispute and move on to the health argument, revealing their health condition and hypothesising that they might be more ill than Belle.

The example above shows that, for her next move, Cadence can choose to either reveal her age or health condition to progress the debate. This decision is affected by which aspect of its description an agent would like to keep private.³ Laying out arguments within this structure still allows for the same dialogue game rules as before, but without the original assumption of complete information. A dialogue under this framework would carry out normally and extend indefinitely as agents exchange moves until one of the agents loses by running out of arguments. Information can, therefore, still be communicated unreservedly. In the interest of quantifying relinquished privacy, we associate arguments to privacy costs.

### 3.4.2 Privacy-Aware Dialogues

Every argument in an alteroceptive culture corresponds to a concession of privacy through communication of one’s partial description. Fortunately, those changes preserve the structure of an argumentation framework, which makes alteroceptive cultures compatible with prior mechanics of dialogue, as well as extensions. Analogously to how humans are comfortable with sharing some types of personal information but not others, some features in agents’ descriptions might be considered more sensitive, and thus having a higher privacy cost to reveal.⁴ We combine the rules of dialogue (Def. 3.2.4) and the notion of privacy cost to provide an instantiation for our architecture.
Definition 3.4.2 (Privacy Cost and Budget) Let $\mathcal{A}_x$ be the extended set of arguments of an alteroceptive culture. We say $	au : \mathcal{A}_x \rightarrow \mathbb{Z}^+$ is a privacy cost function. Let $A$ be the set of all agents. We define $\beta : A \rightarrow \mathbb{Z}^+$ as the privacy budget function of an agent.

For example, in Example 3.4.1, the motion ‘I think I should have this seat’ could have a privacy cost of 0 while the argument ‘I may be older than you, I am 60 years old’ has cost 1, since there is some personal information being revealed within that argument.

The privacy cost of an argument is, however, only the potential cost of revealing partial information. Agents, who are not compelled to share information, decide what information to share and when. If agents are unwilling to share information about their descriptions, they can choose to withhold the information and not raise related arguments. In this way, agents can decide when to reveal information about their partial descriptions. If agents choose to withhold information, their partial description remains non-revelatory, and thus the information is not conveyed and the agent’s privacy budget remains unchanged.

Privacy costs and privacy budgets can be used to model privacy-aware dialogues. In privacy-aware dialogues, agents decide to share information regarding their partial descriptions only when they are willing to sacrifice some privacy. A privacy-aware dialogue, in addition to the properties mentioned in Def. 3.2.4, also has to satisfy a privacy budget constraint.

Definition 3.4.3 (Privacy-Aware Dialogue) Let $w \in \{pr, op\}$ be any two players and $D = m_0, m_1, \ldots, m_n$ be a dialogue. Let $\text{moves}(w,n) = \bigcup_{m_i \in D} \{m_i \mid \text{player}(m_i) = w \text{ and } i \leq n\}$ denote all the moves of a player $w$ up to round $n$. We say a dialogue $D$ is privacy-aware iff, in addition to the criteria in Def. 3.2.4, it also satisfies, for all $n$:

6. $\beta(\text{player}(m_{i+1})) \geq \sum_{m_i \in \text{moves}(w,n)} \tau(\text{arg}(m_i))$, i.e., the player cannot cumulatively spend more than its privacy budget.
In order to keep track of the privacy budget of agents and their corresponding privacy costs, the budget function $\beta$ is initialised with a certain amount, which is determined by the agent. If agents have unlimited privacy budgets, then agents would have no incentive to be judicious with their arguments. With a limited budget, agents can only reveal information to a certain extent. When an agent reaches the end of their privacy budget, they cannot raise any more arguments with privacy costs. If agents want to continue raising arguments, they are compelled to reveal information about their descriptions. In other words, agents cannot continue to raise arguments without revealing partial information about their descriptions. In the interest of protecting agents’ privacy, we must prevent agents from revealing their description.

Our definition imposes a hard limit privacy: during a dialogue, agents cannot use an argument that would aggregately exceed their pre-defined privacy budgets. With finite privacy budgets, the assumption of dialogues extending indefinitely until reaching a unanimous conclusion no longer holds. Disputes end with an agent losing in one of two different ways. Either:  

i) the agent runs out of arguments and is convinced out of the dialogue, or  

ii) the agent still has valid hypotheses and/or verified-facts that could attack the last move, but cannot afford to use them and is forced to concede. In the former case, both agents agree on the correct outcome as the losing agent did not have any valid reasons to challenge the winner. In the latter case, however, the agent concedes and interrupts the dialogue before having all arguments successfully refuted - as it prefers to preserve their privacy. The agents tolerate the result but do not agree with it: they agree to disagree.

Notably, the choice of arguments can have a decisive impact on the dialogue game. If we consider scenario (ii) above to be less desirable than (i) for both agents, both agents must maximise the effectiveness of their moves so that dialogues end on the basis of argumentation instead of privacy budgets. Yet, optimal strategies are not readily available, as partial observability qualifies our privacy-aware dialogue game as a game of incomplete information.\footnote{Typically such games are analysed by a completion approach using Bayesian reasoning, and thus turned into imperfect information games [121]. The alternative allows agents to adopt mixed strategies, using information about an underlying probability distribution. We consider only pure strategies due to fewer needed assumptions, where agents do not use any probabilistic information. However, for pure strategies the optimal solution can be intractable [122, 123], and we thus limit ourselves to simple ones in Chapter 5.}

The remaining chapters will discuss novel applications of cultures and their subsequent results. Chapter 4 demonstrates examples of cultures for resolving conflicts in a human-agent
scenario without privacy restrictions. Chapter 5 applies alte-
roceptive cultures to the fairness-privacy trade-off in a multi-
agent environment. Lastly, in Chapter 6 we explore the use of
cultures for autonomous agents in single-agent environments,
with a direct application for machine learning.

Summary

This chapter introduces an architecture for explainable con-
lict resolution, or cultures, as a mechanism to provide ex-
plainable deliberation capabilities for dialectic interactions
between agents.

We introduced the concept of cultures as a collective
agreement of beliefs, values, and norms, represented
by an argumentation framework.

We then introduced the concept of propositional dia-
logues as a way of modelling conflicts and argued that
argument verification is necessary to check which rules
apply in a specific event.

For environments with privacy restrictions, alterocep-
tive cultures allow for gradual disclosure of informa-
tion by introducing a separation between facts and
hypotheses. This structure allows agents to decide
when to reveal information about their partial descrip-
tions, as every argument has an associated privacy
cost.

We introduced two mechanisms to the architecture:
i) an extended argument set and ii) a privacy bud-
get constraint, which are used to control the level of
privacy agents are willing to divulge.
We are not discussing the evolution of the legal system since then, but rather the concept of rules governing a system.

The Code of Ur-Nammu is the oldest known written law code, inscribed around 2100 BC in ancient Mesopotamia [124]. Its structure is a set of rules (carved in a stone tablet) designed to aid denizens to settle potential conflicts. Conceptually, little has changed since then,¹ as humans historically and currently rely on sets of rules to specify their own systems’ behaviours, expecting peers to abide by those regulations when conflicts arise. Different regimens are defined for several environments, be it traffic, competitive sports, business, civil society, etc.

The increasing apposition of agents and humans in several domains, from embodied agents in human settings, to humans in mixed-reality universes, has created a need to design environments that facilitate the integration of human and machine roles. These environments require mechanisms designed to support and facilitate human-agent interactions. They will be more and more common in the coming years and need to be designed to allow humans to interact with agents and to explain the agents’ behaviours.

We address situations where humans and machines act with independent agency and are subjected to the same rules and conditions.² In such environments, agents and humans are expected to follow the rules defined by the system and to interpret these rules correctly when resolving conflicts. We define these as norm-aware environments, where agents and humans are expected to follow the rules defined by the system.

¹: We are not discussing the evolution of the legal system since then, but rather the concept of rules governing a system.

²: The environment may be a human-agent society, an installation, or any other shared environment where agents and humans are expected to follow the rules defined by the system.
are expected to follow the rules defined by the system. We define a \textit{norm} as a set of rules that govern agents’ behaviours in a specific environment.

\begin{example}
\textbf{Example 4.0.1} Consider this particular example of conflict resolution in human-agent scenarios by argumentation and explaining rules. In a hypothetical disaster relief operation, let us imagine autonomous robots that are assisting in rescue missions. In this example, these robots are navigating a disaster area in order to find survivors. At the same time, humans are also present in this disaster area. Both humans and robots follow a set of rules that define the \textit{right of way} in the environment.

A robot requests passage through the apartment of a human. The human explains that she is the rightful owner of the apartment and that the robot should respect her rights. The robot could either accept the argument and move around, or provide a different explanation that would be valid in the environment. Likewise, the human could either accept the explanation and grant passage, or provide an argument that refutes the robot’s explanation.

The process of providing arguments continues until one party is convinced to let the other party pass. In this example, the robot needs to explain why it should have right of way. The robot explains that it is engaged in a rescue mission, and that it needs to access that apartment immediately to perform a rescue, and that it cannot do so if the human does not allow it to pass. The human then understands the robot’s behaviour and allows it to pass.

If the robot already has information about the human, it could speed up the dialogue process by locally instantiating the verifier functions of its culture and providing a set of arguments that yield a definitive explanation, and communicating that to the human.
\end{example}

Most explainable approaches in human-agent systems are classified with regards to their human-centric or agent-centric [39] approaches, but relatively few are interested in emulating human-agent societies [125]. Can agents and humans with individual goals coexist as peers in a norm-aware environment where resources are limited? Can such peers resolve conflicts and pro-
provide accountability to their decisions, to both humans and other agents alike? Namely, given a multi-agent environment with resource contention, can we define a mechanism that allows us to facilitate human-agent integration by providing: i) an equivalence between human-readable rulesets and agent policies and ii) in a way that is explainable and allows humans to interact successfully with agents to resolve conflicts?

In this chapter, we run an empirical study to investigate the effect of explanations with cultures in varying levels of complexity. We instantiate cultures with a multi-agent resource contention application in the context of the problem of multi-agent path deconfliction. We show how humans and agents can de-conflict trajectories whilst respecting externally-defined ‘rules of way.’

Towards this end, we design a computer game implementing the proposed architecture and conduct a user study to evaluate it. Humans are given path deconfliction rulesets with different amounts of rules each and are asked to navigate in a multi-agent environment and avoid collisions with agents. In our setting, we define complexity as the number of rules that govern the deconfliction of resources. We observe how humans perform in terms of ruleset complexity and the presence/absence of explanations. Our results show that the benefit of explanations is correlated with the complexity of the underlying system. Qualitative results show that human experience in systems with explanations is superior when such systems are sufficiently complex.

4.1 Proof of Concept Study

In order to investigate the usefulness and efficiency of explanations in human-agent deconfliction settings, we designed a user study by instantiating a multi-agent resource contention environment. The problem of multi-agent path deconfliction lends itself naturally to our objectives: it is a sufficiently intuitive problem, requires minimal prior knowledge, and disputed resources are obvious (space).3

Our hypotheses are:
**H$_1$**: Explanations provide a higher improvement for human performance in more complex systems than in simpler systems.

**H$_2$**: Explanations provide a higher decrease in time spent by a human in a task in more complex systems than in simpler systems.

**H$_3$**: User perception of explainable systems is more positive in more complex systems than in simpler systems.

Next, we introduce our definition of a path deconfliction environment and our application: the Busy Barracks game.

### 4.1.1 Path Deconfliction Environment

We define the path deconfliction environment in the form of a 2D discrete time and discrete space grid, represented by a directed acyclic graph (DAG).

Let $L = (V, E)$ be a DAG whose vertices are contained within the points in $\mathbb{Z}^3$, representing a bi-dimensional discrete space as $x, y$ coordinates and time as $t$. Let $u = (x_1, y_1, t_1)$ and $v = (x_2, y_2, t_2)$ be any two points in this space.

We denote $(u, v) \in E \iff (d(u, v) \leq d_{\text{max}} \text{ and } t_2 - t_1 = 1)$, where $d_{\text{max}}$ is the maximum distance achievable by any agent on a single time step, expressed as the Manhattan distance $d_M(u, v) = |x_1 - x_2| + |y_1 - y_2|$ between two points in a $G_{x\times y}$ grid graph.

A set of $K$ obstacles $\Upsilon = \{\Upsilon_1, \ldots, \Upsilon_K\}$ is given as input, where $\Upsilon \subset V$. The resulting traversable graph $G$ is defined as $G = L \setminus \Upsilon$.

A set of $N$ agents $\mathcal{Q} = \{q_1, \ldots, q_N\}$ is placed over $V(G)$, where $N \geq 2$. Each agent can traverse one edge for every time step. This edge may traverse longer distances in $(x, y)$ space, depending on the value of $d_{\text{max}}$ given as input. A goal $g_i \in V(G)$ is defined for every agent $q_i \in \mathcal{Q}$.

Plans to reach goal vertices are represented in the form of path subgraphs of $G$. Given an agent $q_i$ and its corresponding goal $g_i$, the agent’s plan is represented in the form $P(q_i) = \{v_0, \ldots, v_l\}$, where $v_0$ is $q_i$’s current position and $v_l = g_i$. The length of plan $P(a_i)$ is equal to $|P(a_i)|$. 
4.1 Proof of Concept Study

Definition 4.1.1 Two paths \( P(a_i) \) and \( P(a_j) \) are said to be conflicting if \( P(a_i) \cap P(a_j) \neq \emptyset \) (they attempt to visit the same vertex at the same time step) or if \( \forall u = (x, y, t), \forall v = (x', y', t + 1) \) s.t. \( u, v \in P(a_i) \) and \( \exists u' = (x', y', t), \exists v' = (x, y, t + 1) \) s.t. \( u', v' \in P(a_j) \), e.g., agents swap positions.

4.1.2 The Busy Barracks Game

We present the previously-defined Path Deconfliction Environment to human participants as a computer game called Busy Barracks. In the Busy Barracks (BB) game (see Figure 4.1), the human controls a military official represented by an agent \( q_h \in Q \). The human can choose one of two actions: move towards a direction (north, south, west, east), or choose to wait in place for a round. Agents move in lockstep, i.e., once the human makes a decision, all the agents make their planned move at the same time. The human is given 50 arbitrary units of fuel and told to navigate towards a goal destination under the following constraints:

- For every move or wait action, the human will lose 1 unit of fuel.
- If a collision occurs (see Definition 4.1.1), the human loses 5 units of fuel.
- Every 10 seconds past the first move in the game (in clock time), the human loses 1 unit of fuel.

The human is encouraged to reach their destination whilst maximising their remaining fuel. In practice, in order to achieve good scores, players will need to make short, collision-free trajectories with quick reaction times. The environment is composed of several other autonomous agents, who have individual goals and will also move towards their destinations concurrently with the player.

Humans are informed that agents will follow the rules and automatically reroute to clear the way if they understand that the human has priority in a specific conflict setting. Agents will remain in their original trajectory and expect the human to clear the way if they understand that they have priority according to the rules. It is down to the human to make the decision to either...
remain in their original trajectory (assuming that the agent will clear the way) or make way (assuming that the agent will keep their trajectory and will potentially collide if evasive action is not taken).

In order to explore the traits of a system with explicit rule sets, the human is provided with a deconfliction ruleset, presented in textual form on a sheet of paper. This document introduces arbitrary and game-specific properties that each agent has, and how those properties play out in generating a prioritisation when a spatial conflict arises. In other terms, by observing the properties and the rules correctly, every agent should unequivocally understand if they have the right of way or if they should concede and grant passage to the opponent.

**Example 4.1.1** Suppose the following ruleset:

1. You should have right of way if:
   a) Your rank is higher than the other agent’s rank.
b) You are tasked and the other agent is not tasked, regardless of their rank.

This ruleset implies the existence of two properties: rank and tasked status; and two rules: (a) and (b), as seen above. Thus, if we have agents \( q_1 : \{ \text{rank}(q_1) : 2, \text{tasked}(q_1) : \text{yes} \} \) and \( q_2 : \{ \text{rank}(q_2) : 4, \text{tasked}(q_2) : \text{no} \} \), even though \( q_2 \) might be able to argue that it has a higher rank (rule (a)), it will be defeated when \( q_1 \) invokes rule (b).

Following this textual ruleset, we devise an example culture \( C_{\text{easy}} = (\mathcal{A}, \mathcal{R}, \mathcal{K}) \). We instantiate the set of arguments \( \mathcal{A} = \{ \mu, a, b \} \), where \( \mu \) represents the proposition (1) ‘you should have right of way’ and \( a, b \) represent rules (a) and (b), respectively. Let \( c \) be a player and \( \overline{c} \) their immediate opponent. The verifier functions are defined as follows:

\[
\begin{align*}
    f_a(c, \overline{c}) &= \begin{cases} 
        \text{True} & \text{if rank}(c) > \text{rank}(\overline{c}), \\
        \text{False} & \text{otherwise}.
    \end{cases} \\
    f_b(c, \overline{c}) &= \begin{cases} 
        \text{True} & \text{if tasked}(c) = \text{yes} \text{ and tasked}(\overline{c}) = \text{no}, \\
        \text{False} & \text{otherwise}.
    \end{cases}
\end{align*}
\]

Since we know that rule (b) supersedes rule (a), we define \( \mathcal{R} = \{ (a, \mu), (b, \mu), (b, a) \} \) to complete the specification of \( C_{\text{easy}} \).

**Culture:** For the BB game, we created three different rulesets, ranging in different levels of complexity. We posit that cultures become more complex as they grow in number of rules, hence our nomenclature. We refer back to the taxonomy seen in Rosenfeld et al. [39] (*not useful, beneficial, and critical*) to create three cultures with different sizes: easy, medium, and hard. Each culture was created from a textual ruleset that was handed over to human players. These rulesets are shown in full in Appendix A.

- **C_{\text{easy}}:** 2 properties and 2 rules (described in Example 4.1.1.)
- **C_{\text{medium}}:** 4 properties and 4 rules.
- **C_{\text{hard}}:** 6 properties and 9 rules.

**Dialogue Game:** All players can publicly see the destination and intended trajectory of their opponents. When any two
agents find themselves in conflict, they initiate a dialogue game and try to persuade the other to give way to them based on the culture that is being used in that specific instance of the game ($C_{easy}$, $C_{medium}$, $C_{hard}$). In the BB game, all exchanges are *useful-single-argument* dialogues. Moves are chosen randomly among the subset of demonstrably true arguments. The argumentative exchange happens in the background and is not visible to the human.

The decision reached by this dialogue game decides the next action taken by the autonomous agent (to concede via rerouting or to continue in their original trajectory). The human must observe the rules and take action based on their belief of what the agent will do next. Agents always play optimally and do not make mistakes. A wrong decision from the human leads to two possible outcomes: either a collision or an unnecessary diversion from both human and agent, who both try to give way to each other (as the agent assumes the human will also play optimally.) There are 8 agents plus the human in every round, where exactly four of them will have right of way against the human, regardless of difficulty level.

It is worth pointing out that the difficulty level does not affect the map layout, agent behaviour or any other factors that might influence scores or time other than the rules involved in deciding who gives way. If a human played with the same speed and the same success rate in every difficulty level, their scores would always be identical. Differences in score are uniquely defined by human performance.

### 4.1.3 User Study

We recruited 35 participants (21 male, 14 female, ages 20-39) within the university (students and staff). Participants were invited to play the BB game in a quiet room. Every new participant would be allocated to play one out of three versions of the game: either $C_{easy}$, $C_{medium}$, or $C_{hard}$. Participants did not know that any other versions of the game were available.

Our study is organised in a within-subject design in order to measure how each individual participant’s performance is altered in the presence or absence of explanations. Each participant played two rounds of the game (within the same al-
located culture): one version containing the only properties and rules, and another version containing properties, rules, and additional explanations generated in the form of hints in the game UI (see Figure 4.2). Those explanations were generated live by the software based on the outcome of the background dialogue game between the human-controlled agent and the autonomous opponent. In Figure 4.2, the explainable round (right) contains a hint that indicate arguments in the culture that are relevant to the decision at hand, in textual form. For brevity, we shall henceforth denote the non-explainable round as N and the explainable round as X. We alternated the starting order of the rounds (X or N) to minimise familiarity bias.

An experience questionnaire (extracted and modified from the Game Experience Questionnaire [126]) was given to each participant at the end of each round. We clustered questions in three main groups (GEQ indices in brackets): Competence (questions 10, 15, 17, 21); Affect (questions 9, 22, 24); and Challenge (questions 23, 26, 33). We included four custom questions to evaluate game-specific criteria, such as how often they consulted the text rules and if they anticipated/agreed with agents’ actions. Answers were collected in a 5-point Likert scale. We collected game performance data, such as: score (represented by fuel units remaining at the end of the game), number of collisions, and time taken until completion.

The non-explainable version allows the human to visualise their opponent’s trajectory and their properties. Based on this available information (and the rules’ knowledge present in the rule-set), the human must then evaluate which rules apply and decide a course of action. In the explainable case, we decide to provide a succinct, or even partial explanation in the form of a hint.7

---

7: For that reason, in every dialogue game present in the game, our culture generates hints by selecting a 2-reason contrastive explanation $CE'$ (a minimal and compact contrastive explanation) and presenting in a textual form as seen in Figure 4.2. Our objective is not to compare two versions with different information available – but instead to evaluate the impact of having all the information required to make a decision (N) versus having all the information with the addition of an explanation (X) to assist the human.
We are interested in the differences in human performance between playing $N$ and $X$, namely, how much human performance improves or worsens between $N$ and $X$ in each difficulty level.

4.2 Results

Due to the limited number of participants, we choose to not make assumptions of parametrisation in the data. Every sample is grouped into a difficulty level ($E$, $M$, and $H$ representing $C_{easy}$, $C_{medium}$, and $C_{hard}$, respectively). Users play two rounds ($N$ and $X$, in alternated order). Thus, a player allocated to $M$ would play both $MN$ and $MX$ rounds, respectively. Number of participants on each level are: Easy ($n = 11$), Medium ($n = 12$), and Hard ($n = 12$).

We define our measures as:

- **Score ($S$):** normalised score $(S_X - S_N)/(S_{\text{max}} - S_{\text{min}})$. Positive values of $S$ mean score improvement in $X$.
- **Collisions ($Col$):** normalised number of collisions $(Col_X - Col_N)/(Col_{\text{max}} - Col_{\text{min}})$. Negative values mean reduced number of collisions in $X$.
- **Time ($T$):** normalised time elapsed $(T_X - T_N)/(T_{\text{max}} - T_{\text{min}})$. Negative values of $T$ mean reduction in time elapsed in $X$.

4.2.1 Score

Given 3 sample sets: $S_E$ ($E$ scores), $S_M$ ($M$ scores), and $S_H$ ($H$ scores) (see Figure 4.3), we run a Kruskal-Wallis H-Test (KW) under the alternative hypothesis that at least one of the distributions come from a different population and confirm significant differences ($H = 11.63$, $p = 0.003^{**}$). We then perform a pairwise one-sided Mann-Whitney U-Test (MW) under the alternative hypothesis that easier categories have significantly smaller $S$ than harder ones, meaning that the improvement in $X$ is less pronounced in easier rounds.

The results in Table 4.1 show that score improvement in $M$ is significantly smaller than $H$, whilst score improvement in $E$ is very significantly smaller than $H$, although not significantly smaller than $M$. 

---

8: (*) $p < 0.05$; 
(**) $p < 0.01$; 
(***) $p < 0.001$. 

---
4.2 Results

4.2.2 Collisions

Like the previous sets, we consider Col to isolate the number of wrong decisions that specifically led to collisions (see Definition 4.1.1), and how did that differ within subjects between N and X. We run a KW under the alternative hypothesis that at least one of the distributions come from a different population, confirming the hypothesis (H = 9.83, p = 0.007**).

Since the distributions are different, we perform a one-sided MW, this time with the alternative hypothesis that easier categories have significantly higher number of collisions, i.e., they do not improve (and reduce) their number of collisions as well as harder levels. The results in Table 4.1 show that collision improvement in M is significantly smaller than H, whilst collision improvement in E is also significantly smaller than H, although not significantly smaller than M. Both S and Col results support H₁.

4.2.3 Times

Similarly to Col, T represents the change in time elapsed to complete each round from N to X. We run a KW in order to isolate the distributions but did not find significant differences (H = 3.61, p = 0.16). However, a pairwise one-sided MW reveals a
Table 4.1: Pairwise MW for Score ($S$), Collisions ($Col$), and Challenge ($Cha$). Note how the populations between Easy ($E$) and Hard ($H$) are distinguished with high statistical significance. Significant p-values also appear between Medium ($M$) and Hard ($H$). This indicates that participants on the Hard level had significantly larger improvements in score, number of collisions and perception of challenge when presented with explanations compared to the non-explainable round.

<table>
<thead>
<tr>
<th></th>
<th>$S_E$</th>
<th>$S_M$</th>
<th>$S_H$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$U = 48.5$</td>
<td>$U = 11.5$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$p = 0.1472$</td>
<td>$p = 0.0004^{***}$</td>
</tr>
<tr>
<td>$S_M$</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$U = 34.0$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$p = 0.015^{*}$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$Col_E$</th>
<th>$Col_M$</th>
<th>$Col_H$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$U = 72.5$</td>
<td>$U = 111.5$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 0.346$</td>
<td>$p = 0.002^{**}$</td>
<td></td>
</tr>
<tr>
<td>$Col_M$</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$U = 112.5$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 0.008^{**}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$Cha_E$</th>
<th>$Cha_M$</th>
<th>$Cha_H$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$U = 92.0$</td>
<td></td>
<td>$U = 116.5$</td>
</tr>
<tr>
<td></td>
<td>$p = 0.056$</td>
<td></td>
<td>$p = 0.001^{***}$</td>
</tr>
<tr>
<td>$Cha_M$</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$U = 101.0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 0.047^{*}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

significant improvement in $T$ between $E$ and $H$ ($U = 94.0$, $p = 0.04^{*}$), showing that participants in $H$ have a superior reduction in time in $X$ compared to those in $E$. This result supports $H_2$.

4.2.4 User Experience

In order to evaluate the effect of ordering (whether users who played their first round as $N$ or $X$ had a significantly different perception of the game), we ran KWs for each cluster of questions, separating populations by their starting mode ($N$ or $X$). The alternative hypothesis for all cases was that there was a significant difference in populations, which was not confirmed for any: Challenge ($H = 0.24$, $p = 0.61$); Competence ($H = 0.48$, $p = 0.48$); Affect ($H = 0.13$, $p = 0.71$); and Game-Specific ($H = 1.73$, $p = 0.18$). This was expected and indicates that there was no ordering bias influencing the results.

We then evaluate the populations based on the difficulty level. To evaluate the differences in populations, we ran KWs under the alternative hypothesis that the populations differ significantly depending on difficulty level. We manage to validate this hypothesis for Challenge ($Cha$) ($H = 10.28$, $p = 0.005^{**}$), but not for Affect ($At$) ($H = 3.91$, $p = 0.14$), Competence ($Com$) ($H = 4.98$, $p = 0.08$) and Game-Specific ($Gam$) ($H = 5.52$, $p = 0.06$). We perform pairwise one-sided MWs under the alternative hypothesis
that easier categories have a significantly smaller improvement in the perception of challenge from \( N \) to \( X \).

The results in Table 4.1 show that the improvement of perception of \( Cha \) in \( M \) is significantly smaller than \( H \). The improvement in \( E \) is very significantly smaller than \( H \), although not significantly smaller than \( M \). Despite populations being not isolated in the previous KW for \( Gam \), similar pairwise MW results are found: \( Gam_L \) vs. \( Gam_H \) (\( U = 56.5, p = 0.01^* \)) and \( Gam_M \) vs. \( Gam_H \) (\( U = 42.5, p = 0.04^* \)).

Additionally, user experience results in \( At \) and \( Com \) clusters also revealed significant improvements (\( At: U = 96.0, p = 0.03^*; Com: U = 32.0, p = 0.02^* \)) from \( N \) to \( X \) between between \( E \) and \( H \) levels. These results support \( H_3 \).

**Figure 4.4:** Extracted from Guerdan et al \[127\]. A user plays the game in the \( E \) difficulty with and without explanations. The presence of explanations in an environment with rules deemed too simple causes high activation of components (derived from PCA) associated with negative affect.

**Facial Affect Analysis.** This study was postliminarily analysed from an angle of facial affect. We collected data from users who consented to having their facial expressions recorded during playthrough. While not the immediate focus of this thesis, the study by Guerdan et al. \[127\] reinforces the findings of our user experience evaluation by demonstrating how objective facial affect predictions are correlated with subjective user perceptions of competence, task difficulty, agency, and explanation utility. In this study, we recorded the facial expressions of the participants of the Busy Barracks experiment as they played the game in both \( N \) and \( X \) rounds, in different levels of difficulty. We connect facial affect features with participants’ use of explanations and examine how explanations and

\[127\]: Guerdan et al. (2021)
task difficulty interact to influence facial affect signals (see Figure 4.4). The study examines whether predictions differ during key game events (e.g., post-collision) and finds that several affect predictions vary as a function of task difficulty, and vary during difficult game events.

4.3 Discussion and Conclusion

We achieved significant results in demonstrating how the benefit of explanations in human-agent deconfliction correlates to the complexity of the underlying system. Our results demonstrated clear differences between within-subject improvement when comparing their performance in $N$ against the performance in $X$, which demonstrates that humans benefit from explanations – but mostly when the system is sufficiently complex to warrant such explanations.

In fact, when the complexity is small, humans might actually perform better without any explanations. We probe this claim by running a one-sided MW considering the alternative hypothesis that global (between-subjects) $EN$ scores were higher than $EX$ scores ($U = 88.0, p = 0.03^*$), which was significant. Contrariwise, a similar test under the alternative hypothesis that global $HN$ scores are lower than $HX$ scores ($U = 36.0, p = 0.019^*$) also proved significant. In many cases, $M$ populations were harder to distinguish between $E$ and $H$ in nondirectional tests, such as in $T$, $At$, and $Com$ analyses. Still, hypotheses $H_1$, $H_2$, and $H_3$ are validated for all $E$ and $H$ within-subject results. We believe that a larger scale study and further refinement of $M$ in terms of complexity might consolidate all populations more clearly.

Post-experiment interviews were conducted to discuss the user experience. Participants were asked to self-report on how they felt about the hints. Six out of 11 participants who played the $E$ version reported finding the hints not useful. At the $M$ level ($n = 12$), 4 participants found the hints not useful, and 5 expressed using hints as a useful confirmation mechanism to check their mental computation. Last, at $H$ ($n = 12$), 9 players reported that hints were very useful and primarily relied on the hints to act. These findings map well to the taxonomy of Rosenfeld et al. [39] (not useful, beneficial, and critical) and suggest that the
taxonomy of the need for explanations can be considered under a new dimension: that of system complexity.

**Summary**

This chapter presented an empirical study to investigate the effect of explanations with cultures in varying levels of complexity. In our setting, we define complexity as the number of rules that govern the deconfliction of resources. We implemented a computer game called Busy Barracks and conducted a user study to evaluate it. Our results showed that the benefit of explanations is correlated with the complexity of the underlying system. Qualitative results indicate that human experience in systems with explanations is superior when such systems are sufficiently complex.

- We addressed the need for a mechanism to facilitate human-agent integration by providing equivalence between human-readable rule sets and agent policies, in a way that is explainable and allows humans to interact successfully with agents to resolve conflicts.

- We proposed a proof of concept study, instantiating a multi-agent resource contention environment as a computer game called Busy Barracks.

- We ran a user study to investigate the effect of explanations with cultures in varying levels of complexity. We observed how humans perform in terms of rule-set complexity and the presence/absence of explanations.

- Our results showed that the benefit of explanations is correlated with the complexity of the underlying system. Qualitative results show that human experience in systems with explanations is superior when such systems are sufficiently complex.
Cognition and autonomy allow intelligent agents in nature to capture information about their immediate surroundings and independently choose a course of action in consonance with their unique and individual decision-making process. Societies thus emerged as collectives of individuals with (mostly) collaborative intent, aided by implicit and explicit systems of norms and rules. The particular complexity of some of these collectives lead some observers to personify or anthropomorphise these groups, attributing a misleading notion of centralised intent and agency to a coherent, but fully decentralised system. However, even the most harmonious and conformable populations in reality exhibit differences in perspective and disagreements across their members.

In view of the above, we understand that humans do not make decisions with truly global information by virtue of the implausible assumption of omniscience. Rather, each individual acts on their own subjective perspectives of local and global status. This subjectivity is not a flaw but instead a fundamental truth of decentralised systems where information is ultimately incomplete or imperfect. In such systems, agents judge outcomes of conflicts based on their partial knowledge of the world, which can lead to perceptions of unfairness when outcomes differ from
what other peers perceive as correct. Moreover, individual perspectives can differ drastically when agents choose to retain information under concerns of privacy. As such, it is germane to transpose such considerations to human-agent and multi-agent systems of the future.

Example 5.0.1 Consider two individuals Alice and Bob who are vying for a resource, such as a parking spot. In this example, assume that the agents know that there is a ‘priority rule’ for those who require a disabled parking spot. Alice is 3 months pregnant and Bob has a serious hidden disability. Alice is willing to provide an explanation to Bob which reveals that she is pregnant and therefore entitled to the spot according to the priority rule. However, Bob does not want to reveal that he has a disability, and thus has no explanation to provide to Alice. In this situation, Alice will likely win the dispute, but Bob will perceive this result as unfair.

Now, suppose that, besides being pregnant, Alice also had a serious hidden disability. If Bob revealed his condition on the first dialogue, Alice could present her case as well and Bob would be convinced that yielding to Alice is fair. However, since Bob did not, the dialogue still ended with an objectively fair decision (Alice got the parking spot) but one of the participants is expressing subjective unfairness (Bob is unaware of Alice’s condition and believes the spot should be his).

The concepts of fairness and justice are prevalent issues in human history along millenia and have been a central topic in areas such as economics [128], philosophy [129], medicine [130], and more recently, computer science [131]. However, fairness is predominantly regarded as a global property of a system [132–134]. In allocation problems such as nationwide organ transplant allocation, fairness concerns the outcome, i.e., whether patients are prioritised to receive donor organs accurately and without discrimination [135]. Assumptions of global scope and complete knowledge come naturally, as we can only guarantee that a system is fair to all if the information regarding all subjects involved is known. Those assumptions form an objective concept of fairness.

In this chapter, we elicit a provocation to the habitual definition of fairness: in decentralised applications, where the as-
sumptions of complete global knowledge are withdrawn, can we also discuss fairness with regards to the individual perception of each agent in a system, i.e. representing a subjective notion of fairness? If knowledge is incomplete, can we enable agents to understand that apparently (subjectively) negative decisions are globally and objectively fair, i.e., decisions actually made for the greater good?

## 5.1 Explanations, Dialogues, and Privacy

As the gap between objective and subjective fairness resides on the need for reasoning and justification, explanations [39, 136, 137] artlessly lend themselves as desirable tools to address this issue. This is the focus of much of prior literature concerning explanations to human users. However, also a concern is about explanations between artificial agents themselves, or from a human to an artificial agent. Thus, in a fairness-aware decentralised system with disputed resources, agents with explainable capabilities can expound the reasons for deserving resources in a dialogue that is informed by their respective observations and grounded to a mutually agreed-upon definition of what is and is not fair, or a fairness culture.

Since the concerns for fairness also gravitate around the integration of humans and agents together via reasoning, we lean on our architecture for cultures. Ideally, participants in a multi-agent system should engage in dialogues [65, 120] and provide explanations for claiming specific resources. However, not every scenario or application allows for unrestricted exchange of information between agents [138]. Sensitivity and confidentiality aggravate the poignancy of privacy concerns, rendering the subjective perception of fairness issue, at best, non-trivial when privacy is concerned. We defend subjective fairness as a perennial dimension of the fairness argument whenever a system has non-global knowledge or privacy is concerned.

Studies of privacy in multi-agent systems have gained recent popularity [139–141]. More closely, [142] also propose the use of argumentation techniques in privacy-constrained environments, although applied to distributed constraint satisfaction problems. Their approach, however, treats privacy in an abso-
lute way, while in our notion is softer, with information having costs, and with vary- ing degrees of privacy restrictions.

Contemporaneously, the burgeoning research efforts on fairness in multi-agent systems focus on objective global fairness, assuming complete knowledge about all agents [143]. Some works break the global assumption by applying fairness definitions to a neighbourhood rather than an entire population [144] or by assuming that fairness solely depends on an individual [145]. The former studies objective fairness of a neighbourhood, assuming full information of a subset of the population subset, whilst the latter assumes agents have no information outside of their own to make judgements about fairness. These works do not address fairness under partial observability, wherein agents have partial information on a subset of the population, which we call subjective local fairness.

Whilst the dimensions of privacy, subjective fairness, and objective fairness are desirable in all systems, the simultaneous maximisation of all those aspects is often irreconcilable. If the agents lack full information, there cannot be any guarantee of an objectively fair outcome. Additionally, if there are privacy considerations, subjective fairness cannot be guaranteed as the agent requesting a resource does not receive justification for denial. Therefore, with privacy, neither objective nor subjective fairness is achieved.1

This perspective can be observed in society. We illustrate this point as a situation where two passengers are disputing a priority seat in public transport. Even if both individuals are truthful and subscribe to the same culture of fairness (e.g. both agree that a pregnant person should have priority over someone who just feels tired), they must still share information regarding their own personal condition (and therefore, their reasons to justify their belief) to determine who should sit down. How much is each person willing to divulge in order to obtain that resource? Assuming that these individuals have a finite and reasonably realistic threshold on how much privacy they are willing to abdicate for a seat, certain situations will inevitably engender the dilemma of either: a) forfet ing the dispute and conceding the resource to observe their privacy limit, or b) exceeding their reservations in privacy and revealing more information than they should in order to remain in contention.

1: The following discussion assumes that agents can be fully honest, and that their cultures are not in conflict. We do not introduce any notion of trust or deceit.
With adamantine restraints on privacy, contenders will always concede and end the debate if presenting a superior reason would exceed their limits of divulged information. Note that this does not mean that the conceding party agrees with the outcome. Objectively, the fair decision can only be guaranteed if all relevant information is made available by all parties for a transparent and reasonable judgement. In the subjective perception of the bested agent, they still believe they have a superior reason at that point in time, and are left with no choice but to civilly agree to disagree (see Section 3.4.2).

This example highlights that, in the absence of full information, there can be no guarantee of fairness, even if all agents subscribe to the same fairness culture. For example, an agent with a disability that is not perceivable to other agents, who are vying for a disability priority seat and are not willing to reveal information about their disability, may be unable to justify their need for the seat, and will consequently lose the dispute. This, however, does not mean that the agent is being treated unfairly, but rather that the agent is unable to prove that they deserve the seat. In some situations, agents may elude such a dilemma by openly revealing their reasons for requiring a resource, but this is often not possible due to privacy considerations. This leads to the conclusion that, by definition, the concept of fairness cannot be guaranteed when privacy restrictions are in place.

Contributions. In this chapter:

- We introduce a formalisation of the subjective fairness problem and the corresponding privacy trade-off.
- We simulate interactions in random environments and compare how different argumentation and explanation strategies perform with regards to our fairness metrics.
- We instantiate a multi-agent prioritised collision avoidance scenario and demonstrate practical instances of subjective fairness and privacy limitations.

5.2 Problem Definition

While the new argumentation architecture presented above enables dialogues under the partial information of privacy constraints, such partial observability may cause discrepancies be-
between the outcome of a conflict resolution and the agents’ individual perception of what the correct outcome should be, which we define as subjective unfairness. This new slant on the perspective of fairness introduces another dimension for conflict resolution whenever a system has privacy reservations or non-global knowledge. This section thus defines a foundational mathematical formalism to study fairness and its relationship to privacy in decentralised conflict resolution.³

Consider a situation in which a population of agents in a multi-agent system interacts and disputes advantages or resources. Agents interact in a pairwise manner via disputes, and a final system state is achieved through a sequence of such interactions. We observe whether this final state is fair given complete knowledge: this we call objective fairness. A second aspect is whether the outcome is fair from the perspective of each agent. This is subjective fairness. Scope further augments our setting: the term global relates to population-wide observations, whilst local concerns pairwise interactions.

We remove the assumptions of unrestricted information-sharing, instead assuming that privacy restricts an agent’s willingness to divulge information. Particularly, we analyse how privacy between agents is an impediment towards a guarantee of both objectively and subjectively fair outcomes. We formalise such concepts below.

### 5.2.1 Information, Privacy and Fairness

We postulate that agents possess idiosyncratic features with varied values across a population, and that the set of all features that describe an agent is, by reason, called a description. Assume a distributed multi-agent setting where agents have perfect knowledge about their own features but none-to-partial knowledge about other agents’ descriptions. We first define the (full) description of an agent as an n-tuple of features.

**Definition 5.2.1 (Description)** For the rest of this chapter, we consider a fixed set A of existing agents in an environment and an underlying set F of all possible values of a feature. The set of features is \( I = \{1, \ldots, n\} \). We define \( \mathcal{F} = F^n \) as the set of feature descriptions, that is, the set of n-tuples, each entry representing
the value of a feature. The description function \( d : A \rightarrow \mathcal{F} \) is a function that maps each agent to its full description in terms of features.

Since we are considering privacy-sensitive applications, we attribute a privacy cost to each feature, which is then aggregated to produce a cost for the full description. For brevity, all further mentions of cost shall refer exclusively to privacy cost. The constant unknown represents an undisclosed value of a feature.

**Definition 5.2.2 (Privacy Cost)** The set of private features is \( \mathcal{F}_p = \mathcal{F} \cup \{ \text{unknown} \} \). A partial description is an element of the set \( \mathcal{F}_p^n \), that is, an n-tuple with each entry corresponding either to a feature value or to unknown. Each feature \( i \in I \) has an associated cost, \( k_i \). A cost function is defined as \( \tau : \mathcal{F}_p \times I \rightarrow \mathbb{Z}^+ \), such that \( \tau(f_p, i) = 0 \) if \( f_p = \{ \text{unknown} \} \), and \( \tau(f_p, i) = k_i \in \mathbb{Z}^+ \), otherwise. Let \( x = (x_1, \ldots, x_n) \in \mathcal{F}_p \). The description cost function \( \mathcal{T} : \mathcal{F}_p \rightarrow \mathbb{Z}^+ \) is given by \( \mathcal{T}(x) = \sum_{i \in I} \tau(x_i) \).

We can now consider how agents interact: we denote every one-to-one interaction a conflict resolution dispute, or dispute, for brevity. Agents will act with antagonistic intent: one of them pushes for a change in the status quo, and is thus called proponent, while the other is called opponent.

In such pairwise disputes between agents, we need a mechanism for defining what determines an objectively fair outcome of a contest, given two descriptions of agents. Following the idea behind cultures, we assume there is a shared definition of fairness across agents.\(^4\) An objectively fair outcome disregards privacy, since it is the outcome which should be achieved under complete information.

As opposed to this perfect outcome, we also consider what an agent involved in a dispute considers a fair outcome. This we define as the subjectively fair outcome, according to one of the agents in the dispute, which is dependent on the information available to it. Any agent always has full information of oneself, but may only have partial information about the other party.

**Definition 5.2.3 (Fair Outcomes)** We define the set of roles in any dispute to be \( R = \{ \text{pr}, \text{op} \} \), two constant symbols denoting respectively “proponent” and “opponent”, for the rest of the sec-
An objectively fair outcome function \( \omega : \mathcal{F} \times \mathcal{F} \rightarrow \mathbb{R} \) receives the descriptions of the proponent and opponent agents, respectively, and decides which should win a dispute by returning the role of the winning agent in the dispute. The subjectively fair outcome function \( \omega_p : \mathbb{R} \times \mathcal{F} \times \mathcal{F}_p \rightarrow \mathbb{R} \) receives as input whether the agent under consideration (the subject) is the proponent or opponent, the description of the agent, and the private description of the adversary, and outputs the role of the agent which deserves to win the dispute.

The choice of information disclosure to the adversary is determined by a disclosed information function \( \alpha \), which takes an agent (who is deciding which information to disclose), as well as the adversary, and produces a partial description \( \mathcal{F}_p \) of the agent that fits under a specific privacy budget \( g \in \mathbb{Z}^+ \).

**Definition 5.2.4 (Disclosed Information)** We say \( \alpha : A \times A \times \mathbb{Z}^+ \rightarrow \mathcal{F}_p \) is a disclosed information function if and only if, for any agents \( a, b \in A \) and privacy budget \( g \in \mathbb{Z}^+ \), \( \alpha \) satisfies \( \mathcal{T}(\alpha(a, b, g)) \leq g \).

Once both agents made their decisions on how to generate their partial descriptions, the dispute resolution function returns the winner of a dispute with partial information on both sides.

**Definition 5.2.5 (Dispute Resolution)** A dispute resolution function \( \phi : \mathcal{F}_p \times \mathcal{F}_p \rightarrow \mathbb{R} \) decides whether the proponent or the opponent wins the contest given their partial feature descriptions, respectively, by returning the role of winning agent in the dispute.

The difference between functions \( \omega, \omega_p, \) and \( \phi \) goes as follows: the first, \( \omega \), returns the objectively fair outcome if all information is public and mutually known between both agents (complete). The function \( \omega_p \) represents the perspective of the subject upon receiving disclosed information from the adversary, after all, an agent always knows its full description but has partial information about other agents (depending on what is disclosed). Lastly, \( \phi \) represents the final outcome given both agents’ partial descriptions, after the dispute is resolved in some way.

While those definitions cover the essential concepts for pairwise (i.e. local) interactions, we still lack a relation of such pairwise interactions to a global state of the system. We assume
the existence of a set $S$ of possible (global) states in which the system can be, with regards to the population of agents and a global notion of fairness. From an initial state, a set of local interactions transition the system into a final state. This is said to be the transition of the system.

**Definition 5.2.6 (System Transition Function)** Let $A$ be the set of agents, $\alpha$ be a disclosed information function, $\phi$ the dispute resolution function, $g$ a privacy budget, and $S_0 \in S$ a state (referred as the initial state). A system transition function $\sigma$ is a function such that $\sigma(A, \alpha, \phi, g, S_0) \in S$, and we call $\sigma(A, \alpha, \phi, g, S_0) = S_f$ the final state.

We have introduced two aspects of fairness: perspective and scope. Perspective differentiates objective from subjective fairness, while scope differentiates global from local fairness. In terms of perspective and scope, we can think of combined concepts such as *subjective local fairness*, where it pertains to the perception of the individuals in a local conflict. Conversely, *objective global fairness* would observe the actual correctness/fairness of the global state of a multi-agent system. See Example 5.2.1 for some informal intuition. We formalise divergences in conflict resolution outcomes through the proposed notion of a *fairness loss*.

**Example 5.2.1** Some examples can be devised to illustrate the dimensions of perspective and scope and provide some intuition:

- **Objective local fairness**: whether the rightful participant has the disputed resource. Jackie parks a vehicle on a disabled spot and they have a valid reason to do so, compared to the other drivers.

- **Subjective local fairness**: whether the participants believe or perceive a fair decision has been made. If the disabled sticker on Jackie’s car falls off by accident, another driver might arrive later and subjectively see this as an unfair allocation of resources - but they lack information to validate that information objectively.

- **Objective global fairness**: whether the global state of the system is fair. Are the people using the disabled spots
the ones that should have it, or would a different state be fairer to the overall population?

- **Subjective global fairness**: whether the population in a system perceives the system as fair. Is this parking lot rated poorly by the community because of a perception that the disabled spots are not used fairly?

### 5.2.2 Fairness Loss

The outcomes of privacy-restricted fairness disputes may disagree with our previously-defined notions of fair outcomes. We introduce **fairness loss functions** as a means for comparing these.

The first definition for loss evaluates whether the dispute resolution outcome matches the objectively fair one, for a given privacy budget.

**Definition 5.2.7** *(Objective Local Fairness Loss)* The objective local fairness loss function \( l_{OL} : A \times A \times \mathbb{Z}^+ \rightarrow \{0, 1\} \) is defined as

\[
l_{OL}(a, b, g) = \begin{cases} 
0, & \text{if } \phi(\alpha(a, b, g), \alpha(b, a, g)) = \omega(d(a), d(b)) \\
1, & \text{otherwise}
\end{cases}
\]

where \( a, b \in A \) are agents, with \( a \) the proponent and \( b \) the opponent, and \( g \in \mathbb{Z}^+ \) denotes a privacy budget.

Our second definition formalises whether the dispute resolution outcome is the same as the subjectively fair outcome for the agent playing the role \( r \) in the dispute.

**Definition 5.2.8** *(Subjective Local Fairness Loss)* The subjective local fairness loss function \( l_{SL} : A \times A \times R \times \mathbb{Z}^+ \rightarrow \{0, 1\} \) is defined as

\[
l_{SL}(a, b, r, g) = \begin{cases} 
0, & \text{if } \phi(\alpha(a, b, g), \alpha(b, a, g)) = \omega_p(r, d(a), \alpha(b, a, g)) \\
1, & \text{otherwise}
\end{cases}
\]

where \( a, b \in A, r \in R, \) and \( g \in \mathbb{Z}^+ \) denotes a privacy budget.
Finally, for objective and subjective global fairness, we characterise its requirements, but leave it to be specified application-wise. We present an applied experiment of objective and subjective global fairness in Section 5.4.

**Definition 5.2.9** (Global Fairness Loss) *The objective global fairness loss function* \( \Omega : S \rightarrow \mathbb{R}^+ \), *maps a state of a system to an unfairness value according to an objective notion of fairness. Analogously, a subjective global fairness loss function* \( \Omega_p : S \rightarrow \mathbb{R}^+ \) *maps the same state to an unfairness value that stems from the subjective perception of the population. A higher value in either means that the state is less desirable, that is, less fair.*

We now introduce definitions pertaining to ‘orderly’ cases of the previous definitions. Intuitively, a dispute resolution function is one such that when all information is available, an objective local fair outcome is achieved. When that is the case, we call it *publicly sound*, formally as follows:

**Definition 5.2.10** A dispute resolution function \( \phi \) is publicly sound iff for all \( f_a, f_b \in \mathcal{F} \), \( \phi(f_a, f_b) = \omega(f_a, f_b) \).

We also define an agent being reasonable when, given perfect information about the other agent, \( a \) always considers the objectively fair outcome a subjectively fair outcome, in any contest with another agent \( b \). Being reasonable thus implies that the only deterrent towards an objectively fair outcome is partial information, as a reasonable agent will always agree with the objectively fair outcome when given enough information.

**Definition 5.2.11** An agent \( a \in A \) is reasonable iff for any other agent \( b \in A \), and any role \( r \in R \), \( \omega_p(r, d(a), d(b)) = \omega(d(a), d(b)) \).

Global outcomes can now connect with local outcomes by stipulating that, whenever complete information is available and the dispute resolution function is publicly sound, then the objective global fairness loss is 0. This corresponds to system transitions in which, whenever all disputes are resolved with objectively fair outcomes, then the final state is always objectively globally fair.
Definition 5.2.12 Let \( \alpha \) be a disclosed information function, \( \phi \) be a publicly sound dispute resolution function, and \( g \in \mathbb{Z}^+ \) be such that, for every pair of agents \( a, b \in A \), \( \phi(\alpha(a,b,g), \alpha(b,a,g)) = \omega(d(a), d(b)) \). A system transition function \( \sigma \) is publicly unbiased iff for any state \( S_0 \in S \), we have that \( \Omega(\sigma(A, \alpha, \phi, g, S_0)) = \Omega_p(\sigma(A, \alpha, \phi, g, S_0)) = 0 \).

Finally, we can state a result, thus proving that if dispute conditions are sound and reasonable, and the privacy budget is high enough, then subjective and objective fairness losses are equivalent:

Theorem 5.2.1 If \( \phi \) is a publicly sound dispute resolution function, every agent in \( A \) is reasonable, \( g \in \mathbb{Z}^+ \) is such that for all \( a \in A \), \( \alpha(a,g) = d(a) \), and \( \sigma \) is publicly unbiased, then, for all \( a, b \in A \) and state \( S_0 \in S \), \( l_{OL}(a,b,g) = l_{SL}(a,b,g) = \Omega(\sigma(A, \alpha, \phi, g, S_0)) = \Omega_p(\sigma(A, \alpha, \phi, g, S_0)) = 0 \).

Proof. Considering Def. 5.2.7, we know that for a privacy budget \( g \) high enough such that \( \alpha(a,g) = d(a) \) and \( \alpha(b,g) = d(b) \), we achieve a global loss of 0. We can then replace the private descriptions \( \alpha(a,g) \) and \( \alpha(b,g) \) with their public equivalents \( d(a) \) and \( d(b) \) respectively, obtaining \( \phi(d(a), d(b)) = \omega(d(a), d(b)) \). Since \( \phi \) is publicly sound, this holds.

A similar reasoning applies to \( l_{SL} \) in Def. 5.2.8. Suppose a \( g \) high enough such that \( \alpha(a,g) = d(a) \) and \( \alpha(b,g) = d(b) \), for every \( a, b \in A \). Applying this to \( l_{SL} \), for any \( o \in \{a,b\} \), the value is 0 if \( \phi(d(o), d(b)) = \omega_p(o, d(a), d(b)) \). This holds, since \( a \) and \( b \) are reasonable agents.

Finally, for \( \Omega \), it is a direct consequence of Def. 5.2.12. Due to the constraint on \( g \), \( \alpha(a,b,g) = d(a) \), and similarly for \( b \). Thus \( \phi(\alpha(a,b,g), \alpha(b,a,g)) = \omega(d(a), d(b)) \) is satisfied since \( \phi \) is publicly sound.

With the extant formalisms providing a frame of reference and context, we introduce a problem statement that predicates the motivation and sets the direction for our resulting contributions in this chapter.
Problem Statement

Let \( A \) be a set of agents. We assume all agents in \( A \) are self-interested and will experience mutual conflicts of interest regarding contended resources. We assume an initial state \( S_0 \), which transitions depending on outcomes of agent conflicts. For every conflict between any agents \( a, b \in A \) that arises over a contended resource, agents will engage in a dispute with mutually-exchanged partial information. This will be governed by means of: an equal privacy budget \( g \) for both agents, a disclosed information function \( \alpha \) representing the strategy of both agents for releasing information, and a dispute resolution function \( \phi \). For any given \( g \), find an \( \alpha \) that minimises \( \Omega \) and \( \Omega_p \).

5.3 Empirical Analysis

In this section, we propose an experiment to measure the effectiveness of distinct dialogue strategies with respect to global and local fairness under varying degrees of privacy. As shown by Theorem 5.2.1, both objective and subjective fairness loss definitionally converge to zero if the conflict resolution mechanism is publicly sound and agents act reasonably. However, this is a loose bound – as those definitions do not account for the specific mechanics of dialogues, conflict resolution, or fairness.

Consequently, we use the architecture proposed in Section 3.4 to instantiate scenarios with randomly-generated agents and cultures. Our interest lies in empirically observing the impact on global and local fairness under different privacy budgets using four different argumentation strategies. Below, we introduce the details for our experimental setup and evaluation mechanisms.

5.3.1 Setup

Let \( A = \{q_1, q_2, ..., q_{|A|}\} \) be a finite set of \( |A| > 2 \) agents. Every agent \( q \in A \) possesses an \( m \)-tuple \( d(q) = (i_1, i_2, ..., i_m) \) for all \( i \in I \), where \( |I| = m \), representing that agent’s feature description, i.e., its internal traits and characteristics. The value function
\( \mu : A \times I \rightarrow \mathbb{Z}^+ \) returns the numerical value of a feature \( i \in I \) for an agent \( q \in A \).

Let \( C = \{ A, R, K \} \) be a culture, where \( A = \{ y, a_1, a_2, \ldots, a_m \} \) is composed of \( m + 1 \) arguments (with a single motion \( K = \{ \gamma \} \)). Let \( index(a_j) = j, a_j \in A \) denote the index of an argument. We generate \( R \) randomly, satisfying the conditions: i) the underlying \( AF = (A, R) \) has exactly one connected component; ii) for every \((a, b) \in R\), \( index(a) > index(b) \).

Non-motion arguments in \( C \) will represent a feature comparison between two agents. We consider the alteroceptive expansion of \( C \), \( C_x = (A_x, R_x) \) (Def. 3.4.1). Every verified-fact argument \( a^g_{Fj} \in A_x \) is associated to a verifier function \( v_a(q_{jF}, q_{k}) = True \) if \( \mu(q_{jF}, index(a^g_{Fj})) > \mu(q_{k}, index(a^g_{Fj})) \); otherwise \( False \), for \( q_{j}, q_{k} \in A \). All hypotheses are trivially associated to verifier functions that always return \( True \).

Informally, this means that every feature \( i \in I \) is represented in a dialogue between two agents by the respective hypotheses and verified-facts regarding which agent has a superior value in feature \( i \). This abstractly represents any potential feature in a system.

### 5.3.2 Evaluation Metrics

To empirically study the impact of privacy on local and global fairness using different argumentation strategies, we define two evaluation metrics. First, we consider the aggregated subjective local unfairness as the summation of the subjective local fairness loss \( l_{SL} \) over all pairs of distinct agents. In our case, this is modelled by dialogues that end due to a lack of privacy budgets from one of the agents, as noted by dispute result (ii) in Section 3.4.

To measure objective local fairness loss, we calculate a ground truth of all pairwise interactions of agents. This ground truth is defined as an \(|A| \times |A|\) matrix \( GT \), with entries being elements of \( R = \{ pr, op \} \). The entries of \( GT \) are the objectively fair outcome of the dispute in which agents \( a_j \) and \( a_k \) assume the roles of \( pr \) and \( op \), respectively, that is: \( GT_{j,k} = \omega(d(q_j), d(q_k)) \).

Our instantiation of the objectively fair outcome function \( \omega \) is defined the following procedure:
5.3 Empirical Analysis

1. Given the complete descriptions $d(pr)$ and $d(op)$, run all verifier functions for all arguments and remove all verified-fact arguments that return False.

2. Check for sceptical acceptance\(^5\) of the proponent’s motion $y^p_H$. If yes, then return $pr$. Return $op$ otherwise.

The resulting ground truth can also be represented as a digraph $G_{GT} = (A, E)$, where for every two distinct agents $q_j, q_k \in A$, we say an arc $(q_j, q_k) \in E$ iff $GT_{j,k} = pr$ and $GT_{k,j} = op$. We generalise this definition for any strategy $\alpha$ and denote the digraph $G_\alpha$ as a precedence graph of agents.

Objective global fairness loss compares the ground truth precedence graph, $G_{GT}$, to the precedence graph resulting from a strategy, $G_{RE(g,\alpha)}$. To compare these two precedence graphs, we use the DAG dissimilarity metric seen in Malmi et al.\(^{147}\), defined below.

Let $G_1 = (A, E_1)$ and $G_2 = (A, E_2)$ be two precedence graphs. Let $e$ denote an arc $(q_j, q_k)$. Let $c_1$ denote the number of occurrences where $e \in E_1$ and its reverse $(q_k, q_j) \in E_2$, $c_2$ as the number of occurrences where $e$ exists in either $E_1$ or $E_2$ but not the other, and $c_3$ as the number of occurrences where neither $E_1$ or $E_2$ contain $e$. The DAG dissimilarity between two graphs $G_1 = (A, E_1)$ and $G_2 = (A, E_2)$ is $K(G_1, G_2, y_1, y_2) = c_1 + c_2y_1 + c_3y_2$, where $0 \leq y_2 < y_1 \leq 1$ are constants. We choose $y_1 = 2/3$ and $y_2 = 1/3$, via integration.\(^6\)

5.3.3 Strategies

Let $\mathcal{D}$ be the set of all possible privacy-aware dialogues. Let $D = m_0, ..., m_n$ be a privacy-aware dialogue, with $m_n = (w, a)$ the last movement used. We define $\eta : \mathcal{D} \rightarrow 2^{\mathcal{A}_X}$ as the function that returns the set of all arguments $r \in \mathcal{A}_X$ such that $r$ attacks $a$ and $m_{n+1} = (w, r)$ can be used as the next move in the dialogue, that is, $m_0, ..., m_n, m_{n+1}$ is also a privacy-aware dialogue. We enumerate 4 strategies for choosing a rebuttal argument $r \in \eta(D)$.

1. Random: $r$ is sampled randomly with equal probability for all $r \in \eta(D)$.

2. Minimum Cost: $r$ is the argument in $\eta(D)$ with lowest cost.

5: Sceptical acceptance means that the argument is present in all extensions of a certain type. In this case, we generated preferred extensions (see Definition 2.3.5). We used the $\mu$-tols1a SAT-based solver for the simulations.

[146]: Malmi et al. (2015)

[147]: Malmi et al. (2015)
3. **Offensive**: $r$ is the argument in $\eta(D)$ that attacks most other arguments in $A_x$.

4. **Defensive**: $r$ is the argument in $\eta(D)$ that suffers the least attacks in $A_x$.

Let $\alpha \in \mathcal{S}$ be a strategy, where

$$\mathcal{S} = \{\text{random, min\_cost, offensive, defensive}\}.$$ 

A *result* is defined as an $n \times n$ matrix $\text{RE} : \mathbb{Z}^+ \times \mathcal{S} \rightarrow R^{n \times n}$, where $R = \{\text{pr, op}\}$. Every item $\text{RE}_{jk} = \phi(\alpha(q_j, q_k, g), \alpha(q_k, q_j, g))$, for $\text{RE}_{jk}$ represents the *dispute resolution outcome* of a dispute between agents $q_j$ and $q_k$, assuming the roles of $\text{pr}$ and $\text{op}$, respectively. We instantiate $\phi$ by carrying out a dialogue between $q_j$ and $q_k$ until a winner is found, returning $\text{pr}$ or $\text{op}$ accordingly. Analogously to the ground truth model, we can generate a precedence graph $G_{\text{RE}(g,\alpha)} = (A, E)$, where for every two distinct agents $q_j, q_k \in A$, we say an arc $(q_j, q_k) \in E$ iff $\text{RE}_{jk} = \text{pr}$ and $\text{RE}_{kj} = \text{op}$.

### 5.3.4 Simulations

Each experiment consists of a set of $|A| = 50$ agents with fixed feature values initialised randomly. For each experiment, we create a random acyclic culture $C = (A, \mathcal{R}, \mathcal{K})$ with $|A| = 50$, $|\mathcal{R}| \approx 400$, and $\mathcal{K} = \{\gamma\}$. We generate an *alteroceptive expansion* $C_x$ of $C$ where the privacy cost $\tau(a_x)$ for all $a_x \in A_x$ is randomly initialised as $1 \leq \tau(a_x) \leq 20$. Using privacy-aware dialogue games as dispute resolution mechanisms, we generate a precedence graph $G_{\text{RE}(g,\alpha)}$ for each strategy $\alpha \in \mathcal{S}$.

We repeat each experiment for integer value privacy budgets $0 \leq g < 60$. We aggregate measures of global and local fairness over 1900 trials, where each trial randomly initialises all agents’ feature descriptions and the culture.

The plots in Figure 5.1 show the average subjective local fairness loss (left) and the average objective global fairness loss (right) for all four strategies over a range of privacy budget values. All strategies converge to zero subjective local fairness loss given high enough privacy budgets. However, at a given privacy budget value, the defensive strategy dominates the other strategies. Similarly, the defensive strategy is dominant with
5.3 Empirical Analysis

Figure 5.1: The defensive strategy outperforms the other strategies with respect to both local and global fairness under privacy constraints. We show the average subjective local fairness loss (left) and the average objective global fairness loss (right) across a range of privacy budgets for random, min_cost, offensive, and defensive. The shaded regions show 99% confidence intervals on the mean values. Note that the small change in behaviour for $g \geq 20$ is due to the fact that $\max_{\tau(\alpha_x)} = 20$ (see Section 5.3.4).

respect to objective global fairness. In addition to achieving a lower objective global fairness loss at a given privacy budget value, the convergence value of the defensive strategy is lower than that of the other strategies. This indicates the defensive strategy can resolve dialogues with lower fairness loss, even at lower privacy budgets compared to other strategies.

5.3.5 Privacy Efficiency

We run approximately 10 million dialogues between 95,000 different agents, using the same 1900 random cultures from the previous experiment. However, this time we observe how much privacy cost was used by each strategy to finish the dialogues. Agents were free to extend their dialogues for as long as required. Figure 5.2 shows an empirical cumulative distribution function of the proportion of dialogues with cost $z$, where $z \geq z'$, s.t. $k_L(q_i, q_j, r, z') = 1$, for any $\{q_i, q_j\} \subset A$, $\alpha \in \mathcal{S}$, and any $r \in \{pr, op\}$ (i.e. dialogues that need a privacy budget higher than $z$ to not be cut short by it). This illustrates the effect of different strategies in minimising privacy cost, even in unre-
stricted dialogues. Results are consistent with the findings in Section 5.3.4.

Figure 5.2: An empirical cumulative distribution function of privacy costs for unrestricted dialogues in over 1900 different random cultures.

5.4 Multi-agent Application

Manifestly, the previous empirical analysis explores important, but essentially abstract aspects of the problem. Observing properties under a spectrum of multiple different privacy budgets and with large randomly-generated cultures allows us to observe a space of multiple possible systems. However, realistic applications are not likely to explore an assortment of privacy budgets, as the limits on privacy are likely to be fixed or seldom vary. In like manner, cultures can be explainable representations of real-world rules and their exceptions - and much like in the privacy budgets’ case, these are also unlikely to fluctuate in applications predicated in reality.

On these grounds, we build on this conceptual foundation to demonstrate how said aspects may also materialise in an applied setting, even under assumptions of: i) a single fixed privacy budget, and ii) a single fixed alteroceptive culture. Drawing inspiration from the ‘Busy Barracks’ game seen in Chapter 4, we apply our architecture to a multi-agent simulation of speedboats (see Figure 5.3).
5.4 Multi-agent Application

Figure 5.3: Screenshot of our multi-agent collision avoidance simulation. Speedboats are required to engage in privacy-aware dialogues to resolve who has right-of-way before getting into the risk of collision. Boat and ripple art from Felvégi [148]. The dotted lines show the trajectory the boats took to avoid collisions with other agents.

5.4.1 Simulator

Our two-dimensional environment consists of a 20km long × 2km wide water surface, without any obstacles. In every trial, 16 agents are instantiated in a `head-on parade` scenario, as follows:

- 8 out of the 16 agents start from the west, and 8 others start from the east.
- Every agent starts with at least 1km of longitudinal separation from any other agent.
- Their vertical start and goal coordinates are randomised within ±200m from the vertical midpoint.
- Every agent’s destination is at the opposite side of the map, and they are initialised with an exact heading towards it. If undisturbed, any agent should be able to execute a perfect straight line towards their destination without any lateral movement or rotation.

Speedboats are simulated with a physics model adapted from Monster [149] and Linkovich [150], where elements such as water resistance, drag, and mass are taken into consideration when calculating frames. We simulate and control each boat at 20Hz refresh rate. Each trial spans around 10-20 minutes of simulation at 20Hz for 16 agents, and we collect around 400,000

[149]: Monster (2003)
[150]: Linkovich (2016)
telemetry data points per trial, including velocity, lateral acceleration, yaw rate, and lateral jerk. We run 900 trials and generate over 40GB of trajectory and telemetry data for our analyses.

The properties of the vehicle are coarsely modelled after the real-life high performance passenger boat Hawk 38 [151], with capacity for 7 seated passengers, dual 400hp motors, and top speed of 30m/s (approx. 60 knots). It is assumed that vehicles will drive at maximum speed whenever possible.

5.4.2 Boat Culture

Akin to Section 5.3.1, each agent is bestowed with properties extracted from a culture, initialised according to a given distribution. Table 5.1 illustrates the culture designed for the experiment. To preserve a certain degree of realism, the initialisation of each agent is still random, but respecting certain reasonable restrictions (e.g. civilians cannot have high military ranks or engage in combat, spies are always civilian or corporate, etc).

5.4.3 Dialogues and Collision Avoidance

In our simulation, agents have no prior knowledge of other agents’ descriptions. Properties with \( \tau = 0 \) can be communicated instantaneously. It is assumed that agents act lawfully and will not lie, but they are nonetheless self-interested and will fight for the right of way in every conflict that arises. In our scenario, we establish that more sensitive information requires further time to obtain clearance before communicating to another agent. Therefore, in this experiment, privacy costs are associated with time (and consequently, distance). We model their avoidance mechanism using a modified artificial potential field algorithm [152].

Let \( q_i, q_j \in A \) be any two agents, and \( \text{dist} : A \times A \times \mathbb{Z} \rightarrow \mathbb{Z}^+ \) be a function that determines the euclidean distance between two agents at a specific time \( t \). We define two constants \( r_{\text{max}} = 1000 \) and \( r_{\text{crit}} = 100 \) as the maximum effect and the critical minimum radii of a potential field. In our method, when \( \text{dist}(q_i, q_j, t) > r_{\text{max}} \) and \( \text{dist}(q_i, q_j, t + 1) \leq r_{\text{max}} \) (i.e., the agents just got within
Table 5.1: Properties present in our culture, along with their possible values for each agent. The ‘attacks’ column indicates which other properties can be defeated by the respective row.

<table>
<thead>
<tr>
<th>i</th>
<th>Property</th>
<th>( \tau(i) )</th>
<th>Possible values of ( \mu ) (ascending order of importance)</th>
<th>Attacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>motion</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>VehicleAge</td>
<td>4</td>
<td>{new, used, worn, old, vintage}</td>
<td>[0]</td>
</tr>
<tr>
<td>2</td>
<td>VehicleCost</td>
<td>10</td>
<td>{cheap, ok, expensive, very_expensive, millions}</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>3</td>
<td>HigherCategory</td>
<td>0</td>
<td>{civilian, corporate, police, coast_guard, military}</td>
<td>[0, 1, 2]</td>
</tr>
<tr>
<td>4</td>
<td>TaskedStatus</td>
<td>3</td>
<td>{at_ease, returning, tasked}</td>
<td>[0, 1, 2, 3]</td>
</tr>
<tr>
<td>5</td>
<td>PayloadType</td>
<td>5</td>
<td>{empty, food, medical_supplies}</td>
<td>[0, 1, 2]</td>
</tr>
<tr>
<td>6</td>
<td>TaskNature</td>
<td>7</td>
<td>{leisure, sport, trade, training, patrol, pursuit, combat}</td>
<td>[4, 5]</td>
</tr>
<tr>
<td>7</td>
<td>VIPOnBoard</td>
<td>13</td>
<td>{ordinary_person, business_person, celebrity, politician}</td>
<td>[0, 1, 2, 4]</td>
</tr>
<tr>
<td>8</td>
<td>MilitaryRank</td>
<td>8</td>
<td>{no_rank, officer, lieutenant, commander, captain, major, colonel, general, admiral}</td>
<td>[3, 5, 6, 7]</td>
</tr>
<tr>
<td>9</td>
<td>DiplomaticCredentials</td>
<td>12</td>
<td>{no_credentials, diplomat, united_nations}</td>
<td>[0.8]</td>
</tr>
<tr>
<td>10</td>
<td>SensitivePayload</td>
<td>15</td>
<td>{no_sensitive_payload, weapons, wanted_prisoner}</td>
<td>[0.9]</td>
</tr>
<tr>
<td>11</td>
<td>UndercoverOps</td>
<td>20</td>
<td>{no_spy, spy}</td>
<td>[3, 4, 6, 7, 8, 10]</td>
</tr>
<tr>
<td>12</td>
<td>EmergencyNature</td>
<td>10</td>
<td>{no_emergency, mechanical, sick_passenger, fire}</td>
<td>[0.11]</td>
</tr>
<tr>
<td>13</td>
<td>SuperVIPOnBoard</td>
<td>16</td>
<td>{no_super_vip, prime_minister, head_of_state}</td>
<td>[0.12]</td>
</tr>
</tbody>
</table>

the maximum effect radius for the first time), \( q_i \) and \( q_j \) will initiate a dialogue game \( \phi(q_i, q_j, g) \) to decide who has right of way, where \( g \) is the maximum privacy budget for each agent.

**Definition 5.4.1 (Activation Radius)**  Starting from the centre of each agent, we divide the space between \( r_{\text{max}} \) and \( r_{\text{crit}} \) into \( 2g \) concentric and uniformly-expanding rings. Let \( \mathcal{F}(q) \) be the description cost of an agent, that is, how much of its privacy budget was spent to reveal information. Without loss of generality, suppose \( q_i \in \{q_i, q_j\} \) is the winner of the dialogue game and \( z = \mathcal{F}(q_i) + \mathcal{F}(q_j), z \leq 2g \) is the total combined privacy cost for that dialogue, considering both agents. The activation radius

\[
r_{\text{act}} = r_{\text{max}} - \frac{z}{2g}(r_{\text{max}} - r_{\text{crit}})
\]
When the $r_{\text{max}}$ boundary is crossed, agents start a dispute.

The distance $(r_{\text{max}}, r_{\text{crit}})$ spans the privacy budget $(0, 2g)$.

The dispute is resolved with cost $z \leq 2g$. The loser recognises the winner’s potential field and is forced to divert and stay away.

**Figure 5.4**: Illustration of the collision avoidance mechanism from Definition 5.4.1.

*is the distance where the potential field (with full radius $r_{\text{max}}$) will be enabled for the first time for the losing agent, and will remain in effect until the end of the simulation.*

In plain language, agents will start the dialogue as soon as they cross the $r_{\text{max}} = 1000$ boundary, and their cumulative privacy cost $z$ (e.g. the time the dialogue takes) will determine how early or how late they ultimately decide which agent should have right of way (see Figure 5.4). Higher values of $z$ will lead to late, aggressive evasive manoeuvres. In case $\text{dist}(q_i, q_j) < r_{\text{crit}}$, the *winner* is also forced to divert to avoid a collision. In an ideal scenario, most agents will complete their dialogues in full with very low privacy costs and execute early, smooth manoeuvres to stay clear of other agents. Conversely, if dialogues are wasteful or activation radii are closer to $r_{\text{crit}}$ than $r_{\text{max}}$, then one of the following may happen in the local scope, for any $q_i, q_j \in A$ and $r \in \{pr, op\}$:

- the evader makes a very late and aggressive evasion turn *(low privacy efficiency: high $\mathcal{T}(a(q_i, q_j, g))$)*;

- the late evader accepts defeat but breaks into the critical radius of the winner and forces them out of their righteful trajectory *(objective unfairness: $l_{OL}(q_i, q_j, g) = 1$ and $l_{SL}(q_i, q_j, r, g) = 0$)*;
the right agent wins the dialogue game, but the losing agent is not fully convinced all their reasons are covered (subjective unfairness: \( l_{OL}(q_i, q_j, g) = 0 \) and \( l_{SL}(q_i, q_j, r, g) = 1 \));

- the wrong agent wins the dialogue game and forces the right one out of their way since they had no budget remaining to get to the correct decision (subjective and objective unfairness: \( l_{OL}(q_i, q_j, g) = 1 \) and \( l_{SL}(q_i, q_j, r, g) = 1 \)).

As per our previous abstract experiment, we demonstrated that different strategies for choosing arguments (explanations) during the dialogue game leads to varied levels of performance with regards to privacy efficiency, subjective fairness, and objective fairness. We will perform simulations using the same strategies seen in Section 5.3.3 in the sections below.

### 5.4.4 Privacy Efficiency Experiments

Since this experiment involves a real simulation of a multi-agent system, its magnitude is significantly smaller in terms of number of dialogues and agents. We repeat the experiment in Section 5.3.5, this time with only one culture. We observe the privacy costs of 48,000 dialogues between the previous set of 1600 randomly-initialised agents for each strategy without any restrictions in privacy.

Figure 5.5 shows the effect of different strategies in minimising privacy cost for unrestricted dialogues in the Boat Culture. Two phenomena can be observed: the greedy strategy for minimising cost (\( \text{min\_cost} \)) turns out to be the worse at that. This is due to the fact that, as privacy costs are not randomly determined this time, privacy costs are correlated with argument relevance/power, making stronger arguments more expensive. Additionally, both offensive and defensive exhibit similar performance. In a similar nature, in this culture, stronger arguments simultaneously attack many others, whilst being relatively unattacked.

### 5.4.5 Ride Quality

A more applied perspective on the same matter is to observe the status of vehicles across their trajectories. In our simulation,
most disturbances and evasions take place when both sides of the parade meet in the middle. These conflicts generate peaks in lateral acceleration and other measurements. When those conflicts are resolved, agents usually follow trouble-free paths to their destinations. We are interested in quantifying a general notion of ‘ride comfort’ (supposing an acceptable limit of ‘comfort’ for a powerboat at maximum speed) and minimising the ‘ride roughness’ for the fictitious average passenger on board. We discard the initial acceleration and final braking data points, as we are only interested in changes in acceleration caused by other agents.

To do so, we integrate the area under each of those metrics per agent, and compute the mean of each agent’s integrals per trial. We run 100 trials per strategy with a combined privacy budget value of $2\gamma = 60$ and aggregate these values. We associate higher values of lateral acceleration and jerk as strong components in a passenger’s experience of discomfort [153].

The first row in Figure 5.6 shows the distribution of lateral acceleration, yaw rate, and lateral jerk measurements over 1600 simulations (16 agents x 100 trials) per strategy. We observe significantly lower values from offensive and defensive in all metrics. The random strategy also exhibited significantly lower values of lateral jerk compared to min_cost.
5.4 Multi-agent Application

Figure 5.6: Aggregated data of 100 full simulations with a fixed combined privacy budget $2\gamma = 60$. Each strategy is shown in a different colour. We perform pairwise statistical two-sided and one-sided Student’s t-tests between distributions to check if they differ significantly, and if so, which one presents lower values ($^\ast p < 0.05$; $^{**} p < 0.01$; $^{***} p < 0.001$). The top row shows an aggregation of all agents’ simulation metrics ($n = 1600$ per strategy), where lower is indicative of a more comfortable ride. Each data point in the top plots is the area under the curve for that metric (obtained via integration). The bottom plots show values of objective unfairness, subjective unfairness and the subjectivity gap, measured as the Fréchet distance of the corresponding trajectories ($n = 32000$ per strategy). Lower values are indicative of more agreement between trajectories.

5.4.6 Subjective and Objective Trajectories

After observing the impact of privacy on the quality of the ride, we now measure agents’ perspectives in a global scope. Instead of counting local instances of objective and subjective unfairness, we generate entire trajectories that cater to specific requirements. From this point, we will apply the following nomenclature:

- **Nominal trajectory** ($\mathcal{J}_N$): this is the actual trajectory executed by the agent in a regular trial. Agents perform their dialogues normally and give way in case of defeat (as shown in Figure 5.4), according to the current strategy mutually adopted by all agents.

- **(Hypothetical) Subjective trajectory** ($\mathcal{J}_h^S$): same as Nominal, but whenever an agent loses a dialogue that it deems unfair, it hypothesises what would happen if the agent
Figure 5.7: An example of a Nominal trajectory and its hypothetical Subjective and Objective scenarios. The full line (Nominal) represents the actual trajectory performed by the agent in their trial. The Objective hypothesis (dotted line) can perform smoother manoeuvres and evade much sooner due to perfect information being immediately available at $r_{\text{max}}$. The Subjective hypothesis (dashed line) behaves like the Nominal trajectory, but will not concede upon subjective unfairness cases.

ignored the potential field of the opponent and forcefully drove straight through. In this case, agents will never ‘agree to disagree’ and will not give way if the dialogue ends due to privacy limitations.

- (Hypothetical) Objective trajectory ($\mathcal{F}^h_0$): this is a trajectory where all conflict resolution is optimal and derives from the ground truth extensions. Dialogues do not exist and agents have perfect information at cost 0. It hypothesises what would happen if all agents had full mutual knowledge and always made the right decisions.

These trajectories are advantageously intuitive and serve as an example of application-based representations of policies under different perspectives. An illustration of these trajectories can be seen in Figure 5.7. If agents have efficient strategies for selecting relevant information/explanations, their Nominal trajectories will converge to the objectively correct answer more frequently (and more quickly), thus becoming increasingly similar to the Objective hypothesis. Likewise, if strategies are efficient, subjective unfairness will be reduced as agents will be
able to get to the end of their dialogues – and thus the decisions made in the Nominal trajectory and its corresponding Subjective hypothesis will agree.

Let $\mathcal{J}_1, \mathcal{J}_2$ denote any 2 trajectories. We calculate dissimilarity measures across comparable trajectories using the discrete Fréchet distance\textsuperscript{8} $F_{\text{rec}}(\mathcal{J}_1, \mathcal{J}_2)$. We repeated the tests with two other methods (Partial Curve Mapping and Dynamic Time Warp) and achieved similar results. We chose the aforementioned metric for its intuitive value and simplicity, and will omit the other results as they are redundant. For every strategy, we will compare 3 pairs of trajectories:

\begin{itemize}
  \item **Objective Unfairness** ($\Omega = F_{\text{rec}}(\mathcal{J}_N, \mathcal{J}_O^h)$). This represents the level of agreement between the Nominal trajectory and the hypothetical Objective trajectory in a perfect scenario. This mostly captures how objectively correct the decisions were. In Figure 5.6, it is possible to observe that offensive and defensive exhibit much more agreement with the correct trajectory than the other two strategies.

  \item **Subjective Unfairness** ($\Omega_p = F_{\text{rec}}(\mathcal{J}_N, \mathcal{J}_S^h)$). This represents the level of agreement between the Nominal trajectory and what the agent considers to have been the right action, i.e. the hypothetical Subjective trajectory. High levels of disagreement in are representative of high global subjective unfairness - as many agents have very different perceptions of what the right trajectories were. If privacy budgets are minimal and communication is near-impossible, most subjective trajectories will be straight lines, as no agent will ever be fully convinced to alter their route.

  \item **Subjectivity Gap** ($l_S = F_{\text{rec}}(\mathcal{J}_N, \mathcal{J}_S^h)$). This is an assessment on how accurate is the agents’ perception of correctness to the actual correctness. This is measured as the agreement between the hypothetical Subjective and Objective trajectories. We name this phenomenon as the ‘subjectivity gap.’
\end{itemize}

The results in the bottom row of Figure 5.6 show aggregated results for objective and subjective unfairness, as well as the

\textsuperscript{8} ‘Informally, it is the minimum length of a leash required to connect a dog, walking along a trajectory $\mathcal{J}_1$, and its owner, walking along a trajectory $\mathcal{J}_2$, as they walk without backtracking along their respective curves from one endpoint to the other.’

- [154]: Agarwal et al. (2014)
subjectivity gap for each strategy. We perform pairwise comparisons to evaluate how different strategies (and their associated trajectories) fare against each other in terms of higher ride comfort (lower dynamics) and lower unfairness values. Higher values in the first row indicate that, when using that strategy, agents experienced higher motion dynamics associated with discomfort. Likewise, higher values in the second row demonstrate a higher disagreement between the pairs of trajectories when agents use that strategy. We observe that min_cost exhibits the worst performance in all metrics against all other strategies. In second-to-last, random is bested by the other 2 strategies. More remarkably, the winning strategies offensive and defensive exhibit mostly equivalent behaviour, and both are able to practically eradicate occurrences of subjective unfairness within that given privacy budget.

5.5 Discussion

Our experimental results explored four intuitive strategies for conflict resolution under privacy constraints: random, which does not exploit any problem structure; min_cost, which greedily exploits the cost structure; and offensive and defensive, which are informed by the structure of the underlying culture. The strategy performance follows an expected order. Any informed strategy outperforms the baseline random. The exploitation of the culture structure in offensive and defensive yields better results than just using privacy information, as in min_cost. Finally, defensive is expected to be the best strategy, since it follows a more natural principle in argumentation: that an argument which has no attackers must be accepted. In this scenario, choosing arguments with fewer attackers makes it more likely that this argument is final.

Although many others (more involved) strategies could be used, our goal is to measure the effect of using informed strategies as opposed to strategies unaffected by the argumentative structure (the culture). That is, whether a better usage of the argumentative information in the form of dialogues and their resulting explanations leads to better results in both subjective and objective fairness, and other performance metrics in the experimental scenario, under identical privacy constraints.
5.5 Discussion

5.5.1 Randomised Cultures

In the random experiment, Figure 5.1 shows that defensive yields lower losses both locally and globally, especially for realistic privacy budgets that do not overly limit the length of dialogue while meaningfully constraining the amount of shared information. When unrestricting dialogues, defensive was also able to close dialogues with lower privacy costs. As expected, our results suggest that choosing better strategies implies Pareto dominance.

When the privacy budget is near zero, all strategies perform similarly poorly. Little dialogue can occur, so agents lack the information to make objectively good decisions and to provide meaningful justifications. With high privacy budgets, strategies perform similarly well. Predictably, all strategies converge to zero subjective local fairness because dialogues can extend indefinitely. The distinct non-zero convergence values of the objective global fairness loss is due primarily to the ground truth used in these experiments.

Our ground truth corresponds to the sceptical acceptance of the proponent’s motion. The motion is only validated if all attacking arguments are defeated. Having at least one reason to defeat the motion is sufficient to preserve the status quo. If both agents win in some unattacked arguments and lose in others, the outcome will always lean towards the opponent, even when they swap roles.

5.5.2 Multi-Agent Experiment

We observed similar results in the applied multi-agent experiment. Although the distinction between defensive and offensive was not as clear as in with randomised cultures, both still exhibited superior results and reinforced the argument towards the importance of selecting information by relevance. The formalism proposed in Section 5.2 allows for representing a range of different problems, endowing decentralised systems with considerations of subjectivity that can be modelled for non-human agents.

By comparing trajectories under different perspectives, one could be tempted to see the Objective hypothesis through the lens
of a more ‘traditional’ aspect of trajectories, namely, by how much shorter and quicker they are, or by observing the system behaviour with metrics such as makespan or flowtime. In our case, better trajectories are not necessarily ‘optimal’ in the traditional sense. They do not attempt to be shorter or quicker, but instead consider ride comfort metrics and ultimately select for better agreement between what actually happens, what actually should have happened, and what the agents believe should have happened.

The environment with 16 agents avoiding each other through combinations of artificial potential fields in continuous space with simulated physics is dense and complicated, and can lead to edge cases where agents’ decisions can cascade and propagate to multiple others (especially if radii are large). Notwithstanding all the potential for results being confounded by this property of the system, we still managed to demonstrate statistically significant results for superior global performance with better strategies. We expect even clearer results in discrete applications, topological representations [155], or in environments with planning in mind, such as Conflict-Based Search methods [156].

For real applications, the properties and structure of the chosen alteroceptive culture play an important role in indicating which strategy will perform better towards the aimed Pareto dominance within the fairness-privacy trade-off. Perhaps with the exception of degenerate cases of cultures, our heuristics should provide a good guide for considering the choice of arguments for building explanations in fairness-aware conflict resolution disputes, especially if one cannot assess the relevance of the content of the arguments (what it actually means), but has access to its structure (the ruleset that originated the culture, and the relationships between rules).

5.6 Conclusion

In this chapter, we propose perspective and scope as new considerations for the problem of fairness in decentralised conflict resolution. We show how privacy limitations introduce partial observability, and consequently, a trade-off for fairness losses.

[155]: Bhattacharya et al. (2012)

[156]: Sharon et al. (2015)
Our proposed architecture for privacy-aware conflict resolution allows for the representation and resolution of such conflicts in multi-agent settings, and underpins our experimental setup.

Our central insight shows that simulating agents’ actions from the exclusive perspective of their local knowledge and comparing them to the actual executed behaviour can grant an understanding of their subjective perceptions of fairness. Moreover, comparing this subjective perception to an omniscient objective behaviour can also provide an insight on how well-informed agents are in their perception of the world. Different environments and applications could yield interesting combinations of these properties. For instance, a multi-agent society with low global subjective unfairness and a high subjectivity gap could indicate that agents are not acting in accordance with the desired notion of fairness, either by ignorance, design flaw, or adversarial intent.

The theoretical framework of cultures and its privacy-aware specialisation (alteroceptive cultures) allow agents to engage in dialogue games similar to the ones seen in Chapter 4. The inclusion of a privacy component limits the quality and availability of explanations provided, and the strategies for building arguments that compose explanations govern the minimisation of conflict issues when explanations are constrained by privacy. Precise explanations can improve fairness in those systems, as they focus on the decisive arguments to resolve conflicts more effectively.

This work presents no shortage of avenues for future studies. To name a few, extending the scope of interactions beyond pairwise conflict resolution can be investigated by means of collective argumentation [157]. One also could observe the effects of the fairness-privacy trade-off in populations with heterogeneous strategies, and, in like manner, the consequences of heterogeneous privacy budget distributions for subjective and objective fairness, or even extending the dialogues to more general dispute trees, allowing agents to backtrack and find new lines of defence. To a wider regard, the use of cultures as mechanisms for explainable conflict resolution in rule-aware environments encourages similar reflection for our work. Our findings further reinforce the importance of (good) explanations and explanatory strategies in multi-agent systems, this time as
mitigatory instruments for applications where subjectivity is a potential or an impending concern.

Inescapably, agents ‘agree to disagree’ when compelled to cede a resource due to ulterior reasons, such as privacy preservation or time constraints. Whenever continuing the dispute would produce effects more undesirable than the loss of the dispute in itself, agents tolerate those consequences and will prefer to abandon the dispute. This phenomenon is not necessarily new for humans, but extending this consideration for societies of artificial agents grants a dimension of subjectivity that is worth exploring in multi-agent systems research. Preparing non-human agents and systems to consider subjective unfairness can be an important step in facilitating the integration of human agents in future hybrid multi-agent societies. We encourage readers and researchers in the field to regard subjectivity as an important consideration in fairness-aware decentralised systems with incomplete information.
Summary

This chapter presented a study on the impact of privacy in conflict resolution, and the emerging trade-off between privacy and fairness. We instantiated the alteroceptive framework presented in Section 3.4.1 and an experimental setup inspired by the Busy Barracks game in Chapter 4. Our results showed that the choice of argumentation strategies can affect the quality of explanations generated and alleviate the impact of privacy with regards to the fairness of multi-agent systems.

- When agents are limited by privacy constraints, the quality of their explanations is limited by what arguments they are able to use. The inability to use information that could be used for the benefit of an agent is a way to model a subjective perception of fairness in artificial agents.

- We introduced a trade-off between privacy and fairness, and proposed two dimensions to the problem: perspective (objective versus subjective) and scope (local versus global).

- In our empirical simulations, we showed that defensive strategies for choosing arguments can produce better explanations and mitigate the negative effects of privacy in the context of fairness.

- We propose metrics of objective unfairness and subjective unfairness for multi-agent systems. Additionally, a new metric of subjectivity gap can measure the difference between the behaviour idealised by an agent versus the correct behaviour conducted under full information.
6

 Explanation-Aware Reinforcement Learning for Autonomous Agents

‘You don’t get explanations in real life. You just get moments that are absolutely, utterly, inexplicably odd.’

— Neil Gaiman

Human-regulated environments often rely on legislation and complex sets of rules. As advancements in the technology of autonomous systems are drawing machine agents closer to human societies and mechanisms, further integration and adaptation will require systems to partake in said societies’ implicit and explicit rules. Some environments are governed by complex sets of rules (and their corresponding exceptions), and can be difficult to master [8]. Environments such as road and air traffic regulations have complex exception-ridden penalty systems, and we are interested in observing how learning agents can be affected by the depth of the rulesets governing such systems.

Example 6.0.1 Let us take the example of driving. In the UK, driving laws are governed by the Highway Code [158], which contains over 300 rules and exceptions in a single text document. Simple rules can be structured as follows: ‘You must not use your mobile phone whilst driving.’

However, the Code also contains rules that are more intricate in their structure and application, such as:

‘The right-hand lane of a motorway with three or more lanes must not be used if you are driving...’

▶ ‘...any vehicle drawing a trailer’

The work presented in this chapter was developed in a joint first-authorship collaboration with Francesco Sovrano, from the University of Bologna. This resulted in the publication Explanation-Aware Experience Replay in Rule-Dense Environments, published at IEEE Robotics and Automation Letters (IEEE RA-L) 2022. Individual contributions: I designed and implemented the driving cultures, the explanations, and the culture-based explainer mechanism. Both of us co-designed the concept of XAER and implemented the driving environments. Sovrano implemented the RL algorithms and ran the experiments.

[8]: Liu et al. (2021)

[158]: Agency (2004)

1: Notably, even this rule is subject to exceptions, as drivers are allowed to use a mobile phone if they are calling emergency services and it is not safe to stop.
6.1 Problem Definition

We expect self-learning agents of the future to be able to adequately explore these environments and learn efficient policies that ensure good rewards and minimal rule-breaking. However, present methods are usually tested in environments with relatively few rules and exceptions [159]. Denser regulations appear in applications of RL for autonomous vehicles research, but such rulesets are often fixed in terms of complexity [160]. As the complexity of rulesets increase, either in number of overlapping norms or chained exceptions, generating a sufficiently diverse set of experiences for Reinforcement Learning (RL) agents can be difficult, especially with large numbers of corner cases arising as a consequence of dense rulesets. In other terms, exposing these exceptions to an RL agent can be challenging, depending on the environment.

With large numbers of corner cases arising as a consequence of dense rulesets, generating a sufficiently diverse set of experiences and exposing these exceptions to an RL agent can be challenging. Some works in literature propose to sample past experiences related to those exceptions, heuristically revisiting potentially important events. Among them, the technique of Pri-
6.1 Problem Definition

Prioritised Experience Replay (PER) [9] looks at over-sampling experiences that are most poorly captured by the agent’s learned model. However, this mechanism does not necessarily focus on the cause of events or their exceptional nature.

In this chapter, we are interested in observing how learning can be affected by the depth of the rulesets governing such systems. We pursue the intuition that explanations are a pivotal mechanism for human intelligence, and that this mechanism has the potential to boost the performance of RL agents in complex environments by connecting experiences with complex rule-based systems. This is why we draw inspiration from user-centred explanatory processes for humans [161], and design a set of heuristics and mechanisms for prioritised experience replay to explain complex regulations to a generic off-policy RL agent.

A central design challenge towards this goal is integrating explanations into computational representations. Approaches such as encoding the ruleset (or part of it) into the agent’s observation space may incur severe re-training overhead even under minimal ruleset changes, as the semantics of the regulation are explicitly provided as input [162]. This minimises compatibility with extant methods and may obscure whether differences in performance are due to changes to the architecture or the complexity of the ruleset. We propose a solution that is agnostic to explicitly engineering state and observation spaces, using an explanation-aware experience replay mechanism.

In our approach, we see explanations as clusters of experiences, and use them to modify a conventional experience replay structure. We do so by avoiding explicit representations of the ruleset (i.e. rule-based explanations [19]) by instead representing the meaning of the regulations as organised collections of examples (i.e. case-based explanations [20]). These explanations do not need to be understood by the agent in the traditional sense, but can still convey meaning if the example was labelled/explained in a semantic and meaningful process. Therefore, our approach modifies conventional experience replay structures by partitioning the replay buffer (or memory) into multiple clusters, each representing a distinct explanation associated with a collection of experiences that serve as examples. We call this process Explanation-Aware Experience Replay (XAER) (see Figure 6.1) and integrate this technique into three seminal learning frameworks.

[9]: Schaul et al. (2015)

[161]: Sovrano et al. (2021)

[162]: Kiran et al. (2021)

[19]: Branting (1991)

[20]: Aamodt et al. (1994)

2: In a ludic example, suppose a young man, called Luke, is taking hyperspace flight lessons from his exasperated friend Chewbacca. However, he does not understand a single word of Shyriiwook, the tutor’s language. With sufficient repetition, Luke can associate distinct Wookiee growls (and punishments) to categories of experienced episodes, even if the content of the message is in an unknown language. Eventually, Luke would learn the meaning of the most relevant utterances by associating them to the experienced consequences.
Conclusions. In this chapter:

- We show how distinct types and instances of explanations can be used to partition replay buffers and improve the rule coverage of sampled experiences.

- We design discrete and continuous environments (GridDrive and GraphDrive) compatible with modular rule sets of arbitrary complexity (cultures). This leads to 9 learning tasks involving both environments with different levels of rule complexity and reward sparsity. These serve as a platform to evaluate how RL agents react to changes in rule sets whilst keeping a consistent state and action space.

- We introduce XAER-modified versions of traditional algorithms such as DQN, TD3, and SAC, and test the performance of those modified versions in our proposed environments.

Upon experimenting on the proposed continuous and discrete environments, our key insight is that organising experiences with XAER improves agent performance (compared to traditional PER) and can be able to reach a better policy where traditional PER may fail to learn altogether.
6.2 Explanation-Awareness

The notion of explanations we are interested in is aligned to the epistemic interpretation of explanations [18]. In particular, we focus on the case in which explanations are understood as a means of providing an agent with a labelling system that helps her to organise and model the world she interacts with. For this purpose, we propose a transformation of rule-based explanations (e.g. given by a ruleset/culture) to case-based explanations (experience), which are compatible with experience replay. Leaning on the concept of Explanation-Awareness (XA), our heuristics facilitate information acquisition via the organisation of experience buffers.

A central aspect of providing case-based explanations to an RL agent comes from meaningfully re-ordering experience to a greater degree. The intuition behind how we construct our case-based explanations can be laid out as: ‘a simple set of relevant state-transitions representing abstract-enough aspects of the problem to be solved.’ This intuition motivates the heuristics of abstraction, relevance, and simplicity (ARS, in short). We adapt these heuristics from prior work [161] in the HCI domain, where they are presented in greater abstraction to form a higher-level

[18]: Wright (1976)

3: The agent does not engage in a higher-level reasoning process to understand how the ruleset is structured, but only uses the explanations to organise experiences. In fact, this is one of the advantages of our proposed model as we do not have to alter the agent’s observation space to accommodate explanations.

[161]: Sovrano et al. (2021)
4: Note that the explanations generated by the explainer can have virtually any representation, be it human-understandable or not, provided they are distinct and serve the purpose of labelling different clusters.

5: More specifically, the more diverse the explanations are, the more likely the agent will be able to learn the underlying rules of the environment.

taxonomy and knowledge graph for an interactive explanatory process.

Consider a problem where an RL agent has to learn a policy to optimally navigate through an environment with sophisticated rules and exceptions (e.g. a real traffic regulation with exceptions for special types of vehicles). Let the state-transition $\tau = (s_t, a_t, r_t, s_{t+1})$ denote the transition from state $s_t$ to state $s_{t+1}$ by means of action $a_t$, yielding a reward $r_t$. We assume the environment is imbued with explanatory capabilities via an explainer.4

**Definition 6.2.1 (Explainer)** The explainer $\epsilon : \Omega \rightarrow ES$ is a function that maps a list of state-transition tuples $\tau \in \Omega$ to an explanation $e_\tau \in ES$, where $\Omega$ is the space of possible state-transitions and $ES$ is the explanatory space, i.e., the space of all possible explanations.

The explainer can be a black-box, but its nature determines the structure of the resulting explanation space. For instance, if the explainer is itself a rule-based system, the explanation space would be the set of all rules that are satisfied for the given state-transition. On the other hand, if the explainer is based on pre-trained value prediction, i.e., a model trained to predict the amount of reward that could be anticipated in the future, then the explanation space would be the set of all possible outcomes (values). Note that the set of all explanations is defined by the explainer, but the agent is not obliged to use all of them.

An agent who has more diverse experiences with regards to the reasons (explanations) associated with rewards will have a better chance at converging towards a policy that better represents the underlying ruleset.5 Therefore, we posit that the more complex the environment is in terms of rules, the more useful Explanation-Awareness (XA) should be, as it would ensure a more even distribution of experiences with regards to different reasons justifying rewards. This diversity of explanations culminates on a clustering that is semantic by nature, and transitions are partitioned according to the explanation that represented its reward.

**Definition 6.2.2 (XA Clusters)** Let $\tau_\epsilon = (s_t, a_t, r_t, e_\tau, s_{t+1})$ be a XA state-transition represented by the explanation $e_\tau$, where
We introduce our adaptation of ARS, below.

6.2.1 Abstraction: Clustering Strategies

The purpose of the abstraction heuristic is to regulate the level of granularity of the explanations, hence of the experience clusters. Our abstractions are based on the understanding that explanations are indeed answers to questions. Hence, explanations may have different granularity defined by the level-of-detail of the question they answer.

We consider two types of explanations in our work, represented by the questions they answer: \textit{WHY} explanations and \textit{HOW} explanations. The \textit{WHY} explanations are those that can be produced by a rule-based system, whereas the \textit{HOW} explanations are those that can be produced by a value estimation system. These two types of explanations can be used in combination to provide a more comprehensive understanding of the rule-based system. The \textit{WHY} explanations are useful to understand the more abstract rules of the environment, whereas the \textit{HOW} explanations are useful to understand the performance of the agent.

More in detail, the \textit{HOW} explanations we consider answer the question ‘how well is the agent performing with this reward?’. This type of explanations can be produced by studying the average behaviour of an agent. For example, if an episode has a cumulative reward that is greater than the running mean, then the explanation indicates that the agent is behaving better than average. Hence, the \textit{HOW} explanations do not need to be designed with any specific domain knowledge, as they are governed exclusively by the performance of the agent.\footnote{As opposed to \textit{WHY} explanations, which are typically designed by an expert and require domain knowledge.}

On the other hand, the \textit{WHY} explanations we consider answer the question ‘Why did the agent achieve this reward?’ The \textit{WHY} explanations could depend on an explainer function with task/domain knowledge that can distinguish and cluster types of transitions (see Example 6.2.1, below). Furthermore, \textit{WHY} and \textit{HOW} explanations (or any other type) can be combined so that the explanation would answer both the associated questions.
In order to compose the experience buffer, represented by the set of experience clusters \( \mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_k\} \), we consequently devise the following clustering strategies, for each explanation type:

1. **HOW**: The experience buffer is divided into 2 clusters \( C_{\text{better}} \) and \( C_{\text{worse}} \), where \( C_{\text{better}} \) contains batches with rewards greater than the running mean of rewards, and vice-versa (given a sliding window of a defined size).

2. **WHY**: The number of clusters is equivalent to the number of distinct explanations available. If a batch can be explained by multiple explanations simultaneously, we select the explanation associated with the smallest cluster (most under-represented) and the batch is associated to the corresponding cluster.\(^7\)

3. **HOW+WHY**: a combination of **HOW** and **WHY** strategies. There are two custom \( C_{\text{better}} \) and \( C_{\text{worse}} \) clusters for every **WHY** explanation, formed after their concatenation.

**Example 6.2.1** Suppose a hypothetical football environment with a **WHY** explainer function. This function could either be part of the environment (a logical mechanism that recognises when certain states are reached and produces a state label), or an external mechanism that receives state-transitions as input and produces explanations. The explanations could be generated by the rules of the game, such as ‘goal’, ‘offside’, or ‘foul’. The corresponding **WHY** clusters would be \( \mathcal{C} = \{C_{\text{goal}}, C_{\text{offside}}, C_{\text{foul}}, \ldots\} \), where each cluster would contain a set of state-transitions associated with each label. If **HOW+WHY** were used, clusters would be

\[ \mathcal{C} = \{C_{\text{goal,better}}, C_{\text{goal,worse}}, C_{\text{offside,better}}, \ldots\}. \]

After clustering state-transitions using the prior clustering strategies, we propose mechanisms for assessing the relevance of specific state-transitions during learning.
6.2.2 Relevance: Intra-Cluster Prioritisation

Prioritisation mechanisms are used for organising information given their relevance to the agent’s objectives.

The priority of a batch is usually estimated by computing its loss with respect to the agent’s objective [9]. In DQN, TD3, and SAC, relevance is estimated by the absolute TD-error of the agent. The closer to 0, the lower the loss and the relevance. The intuition is that batches with TD-error equal to zero are of no use since they represent an already solved challenge. In our method, this relevance heuristic can be combined with the aforementioned clustering strategy by sampling clusters in a prioritised way (by summing the priorities of all its batches) and then performing prioritised sampling of batches from the sampled cluster.

6.2.3 Simplicity: (Curricular) Inter-Cluster Prioritisation

Occam’s Razor [163] states that when presented with two explanations for the same phenomenon, the simplest explanation should be preferred. In human explanations, simplicity is a common heuristic [164, 165]. We will adhere to those principles and select minimal and simple explanations, following a curricular approach. The goal is to sample experiences from explanations that are more pertinent to the current level of learning of the agent. We implicitly associate complexity with rarity, from the assumption that default and basic behaviours are simpler than chained exceptions and corner cases.

Clustered prioritised experience replay changes the real distribution of tasks by means of over-sampling.⁸ We would like to encourage the agent to learn the most common tasks first, before learning the more exotic ones. The agent would progress towards the more challenging tasks in a curricular fashion, where the simplest tasks are learned first. Assuming that the whole experience buffer has a fixed and constant size $N$, and that the experience buffer contains $|\mathcal{C}|$ different clusters, let $S_{\text{min}}$ and $S_{\text{max}}$ be the minimum and maximum size of a cluster. Any new experience is added to a full buffer by removing the oldest one within buffers having more elements than $S_{\text{min}}$.  

---

⁸: While a traditional PER buffer could be viewed as a permutation of the real distribution, our clustered version changes the distribution by prioritising specific parts of it.
Figure 6.2: An example illustrating how the cluster size proportion can regulate the degree of on-policyness of the replay space. If clusters are identical in size, exceptions are likely to be tackled before common tasks (due to higher TD-error). Oppositely, if we allow clusters to replicate the current distribution, we increase the degree of on-policyness and the agent is unlikely to over-sample exceptions. By enforcing a proportion between maximum and minimum sizes of clusters, we can find a balance between the two previous cases: preserving the trend of the distribution, e.g. the predominance of $e_1$ and $e_2$ tasks, whilst keeping valuable examples of rare explanations such as $e_6$ still statistically significant.

If all the clusters have the same size (therefore $S_{\text{min}} = S_{\text{max}}$), replaying the cluster of a task with the highest (TD-error) priority might push the agent to tackle the exceptions before the most common tasks, preventing the agent from learning an optimal policy faster.\(^9\)

On the other hand, if $S_{\text{min}} = 0$ and $S_{\text{max}} = \infty$, the size of a cluster would depend only on the real distribution of tasks within a small sliding window, as in traditional PER, thus preventing over-sampling. The presence of clusters helps over-sampling batches likely related to under-represented tasks, and learning to tackle potentially hard cases more efficiently.

Thus, we aim to enforce a range of sizes for all clusters (see Figure 6.2), preventing the agent from learning tasks that are too simple (and thus, not requiring over-sampling for a better performance) or too complex (and thus, not benefiting from the clustered prioritised experience replay).\(^{10}\)
Consequently, we posit that $S_{\text{min}}$ shall be large enough for effective over-sampling, while having $S_{\text{max}} > S_{\text{min}}$ being dependent on the real distribution of tasks. This will push the agent towards tackling the most frequent and relevant tasks first, analogously to curricular learning. We define a hyperparameter to control the cluster size proportion (see Figure 6.2).

**Definition 6.2.3 (Cluster Size Proportion)** In order for all clusters to have a size $S_{\text{min}} \leq S \leq S_{\text{max}}$, we set $S_{\text{max}} = S_{\text{min}} + (\xi - 1) \cdot |\mathcal{C}| \cdot S_{\text{min}}$, where $\xi \geq 1$ represents the cluster size proportion.

Therefore, $S_{\text{min}} = \frac{N}{|\mathcal{C}| \xi}$ can be easily controlled by modifying $\xi$. We enforce $S_{\text{min}} < S_{\text{max}}$ when $\xi > 1$. For curricular prioritisation, if the cluster’s priority is (for example) computed as the sum of the priorities of its batch, and $\xi > 1$ is not too large (e.g. $\xi = 5$), the resulting cluster’s priorities will reflect the real distribution of tasks while smoothly over-sampling the most relevant tasks. This avoids over-estimation of the priority of a task. As $\xi$ gives us control of the degree of on-policyness, different values of $\xi$ might perform better with on an algorithm and environment basis\(^1\). Higher values of $\xi$ mean that the distribution of state-transitions reflects more transitions seen within the current policy, thus being advantageous for entropy-maximisation algorithms such as SAC \([113]\). Likewise, fully off-policy algorithms such as DQN may exhibit superior results with low values of $\xi$ (e.g. $\xi = 1$).

### 6.2.4 Annealing the Bias

Similarly to PER \([9]\), sampling state-transitions from prioritised clusters might produce unwanted bias. The standard debiasing function of PER weighs expected values using the normalised weight $\frac{p(\hat{\tau})}{P(\tau)} \in [0, 1]$, where $P(\tau)$ is the probability of sampling a state transition $\tau$ from the whole buffer and $\hat{\tau}$ is the state-transition with the lowest probability for the whole buffer. We adapted the debiasing function of PER by changing the formula to consider the fact that state-transitions are sampled from clusters (which are in turn sampled). Therefore, the debiasing function of XAER computes the joint probability of sampling both a cluster $c$ and a state-transition $\tau$. More precisely, considering that the two events are not independent, we compute this joint

\(^{11}\): However, tuning for $\xi$ seems relatively simple, and a grid search on $\xi \in \{1, 2, 3, 4, 5, \infty\}$ might suffice for most cases. 

\([113]\): Li et al. (2021)

\([9]\): Schaul et al. (2015)
probability as $P(c) \cdot P(\tau|c)$. Hence, the normalised weights produced by the debiasing function of XAER are given by $rac{P(\hat{c} \cap \hat{\tau})}{P(c \cap \tau)}$, where $P(\hat{c} \cap \hat{\tau})$ is the lowest possible probability, considering any couple of clusters and state-transitions.

With those mechanisms in place, we propose new environments to evaluate the performance of agents when subjected to complex rulesets.

### 6.3 Environments

Real-life air/sea/road traffic regulations are often complex, and their mastery is a crucial aspect of orderly navigation. Many realistic settings have a number of exceptions that must be taken into consideration (e.g. ambulances are not subjected to some rules when in emergencies, sailing boats have different priorities if on wind power, etc). To evaluate the effect of XAER in a diverse configuration space of environments, we developed modular environments that allow us to systematically change its properties in evaluation. Our environments allow for agents to experience the same rules (our Easy, Medium, and Hard rulesets) in both discrete and continuous state-action spaces, and with frequent and sparse rewards. In these, the agent must understand the complex regulation governing the penalty system. To implement our rulesets, we use cultures as explainer functions to produce rule-based explanations from an agent’s behaviour.

The environments are:

#### 6.3.1 GridDrive - Discrete

A 15×15 grid of cells, where every cell represents a different type of road (see Figure 6.3, left side), with base types (e.g. motorway, school road, city) combined with other modifiers (roadworks, accidents, weather). Each vehicle will have a set of properties that define which type of vehicle they are (emergency, civilian, worker, etc). Complex combinations of these properties will define a strict speed limit for each cell, according to the culture.
Actions. A sample \((d, s)\) in the action space consists of a direction \(d \in \{N, S, E, W\}\) and a speed \(s\) where \(0 < s \leq 12\).

Observations. A sample \(o = (o_v, o_r, M, x, y)\) in the observation space is a tuple where \(o_v\) denotes the concatenation of the vehicle’s properties (including speed), \(o_r\) is the concatenation of all neighbouring roads’ properties, \(M\) is a \(15 \times 15 \times 2\) boolean matrix keeping track of visited cells, and \((x, y)\) represent the vehicle’s current global coordinates.

Rewards. Let \(0 < s' \leq 1\) denote the normalised speed of the agent in that step. Rewards are given at every step, given the following criteria:

\[
\begin{cases} 
-1 \text{ (terminal),} & \text{if breaking the speed regulation} \\
0, & \text{if on previously-visited cell} \\
s', & \text{otherwise (new cell, within speed limit)}
\end{cases}
\]

Figure 6.3: Diagrams representing our proposed GridDrive and GraphDrive environments. In GridDrive, the agent has a discrete action space and must observe the properties of neighbouring cells to make a decision that is compatible with the ruleset, choosing one direction and a fixed speed. GraphDrive is a harder environment, where the agent’s action and observation spaces are continuous. In it, kinematics are taken into consideration and the agent must not only learn the rules governing penalties, but also to accelerate and steer without going off-road. The goal in both environments is to visit as many new roads as possible without infringing rules.
6.3.2 GraphDrive - Continuous

An Euclidean representation of a planar graph with $n$ vertices and $m$ edges (see Figure 6.3, right side). The agent starts at the coordinates of one of those vertices and has to drive between vertices (called ‘junctions’) in continuous space, using Ackermann-based non-holonomic motion. Edges represent roads and are subjected to the same rules with properties to those seen in GridDrive plus a few extra rules to encourage the agent to stay close to the edges. The incentive is to drive as long as possible without committing speed infractions. In this setting, the agent must learn a control input that not only keeps the vehicle on the road, but also respects speed limits and restrictions that may vary on a case-by-case basis. We test two variations of this environment: one with dense and another with sparse rewards.

Actions. A sample $(\theta, a)$ in the action space consists of a steering angle $\theta$ where $-\frac{\pi}{4} \leq \theta \leq \frac{\pi}{4}$, and an acceleration $a$ where $-7 \leq a \leq 1$. Acceleration and deceleration ranges are chosen given road car standards (in $m/s^2$) [166] [167].

Observations. A sample in the observation space for GraphDrive is a tuple $(o_v, o_r, o_j)$, where $o_v$ denotes a concatenation of the vehicle’s properties (car features, position, speed/angle, distance to path, junction status, number of visited junctions), $o_r$ is the concatenation of the properties of the closest road to the agent (likely to be the one the agent is driving on), and $o_j$ is the concatenation of the properties of roads connected to the next junction.

Rewards (dense version). Let $0 < s \leq 1$ denote the normalised speed of the agent in that frame, and let $n$ be the number of unique junctions visited in the episode. Rewards are given at every frame, given the following criteria:

\[
\begin{align*}
-1 \text{ (terminal)}, & \quad \text{if breaking the speed regulation} \\
-1 \text{ (terminal)}, & \quad \text{if off-road or U-turning outside junction} \\
0, & \quad \text{if on junction or previously-visited road} \\
\min(s, 1), & \quad \text{otherwise (on road, within speed limit)}
\end{align*}
\]

Rewards (sparse version). In this version, the agent will get null (zero) reward when moving correctly. Positive rewards
only appear when the agent manages to acquire a new junction. Therefore, the agent will have to drive entire roads correctly to get any positive reward. Rewards are given according to the following criteria:

\[
\begin{cases}
-1 \text{ (terminal), if breaking the speed regulation} \\
-1 \text{ (terminal), if off-road or U-turning outside junction} \\
0, \text{ driving normally or on acquired junction} \\
1, \text{ the instant a new junction is acquired}
\end{cases}
\]

Every episode incurs in an initialisation of the grid or graph (for GridDrive or GraphDrive, respectively) with random roads, along with randomly-sampled agent properties. The agent is encouraged to drive for as long as possible until it either achieves a maximum number of steps or breaks a rule (terminal state). All environments will be instantiated in versions with 3 different cultures (rulesets), according to their levels of complexity: Easy, Medium, and Hard. See the complete rulesets in Appendix B.

- Easy: 3 properties (2 for roads, 1 for agents), 5 distinct explanations.
- Medium: 7 properties (5 for roads, 2 for agents), 12 distinct explanations.
- Hard: 15 properties (9 for roads, 6 for agents), 20 distinct explanations.

6.4 Experiments

In this section we describe our experimental setup and present results obtained in our proposed environments with XAER versus traditional PER. We trained 3 baseline agents with traditional PER (DQN/Rainbow, SAC, and TD3). For each of the 3 baseline algorithms, we train 3 XAER versions with different clustering strategies, using HOW, WHY, and HOW+WHY explanations (see Section 6.2). Additionally, we show results for HOW+WHY explanations without the simplicity heuristic (prioritised clustering) — i.e. clusters are sampled uniformly. For a total of 12 XA agents, we call the XAER-equipped versions of DQN, SAC, and TD3 XADQN, XASAC, and XATD3, respectively. DQN
and XADQN agents are applied to GridDrive (discrete), whilst SAC, TD3, XASAC, and XATD3\textsuperscript{12} were trained separately on GraphDrive with dense and sparse rewards (continuous).

The neural network adopted for all the experiments is the default one implemented in the respective baselines (although better ones can be certainly devised), and it is characterised by fully connected layers of few units (e.g. 256) followed by the output layers for actors and/or critics, depending on the algorithm’s architecture. XAER methods introduce the cluster size proportion ($\xi$) hyperparameter. We perform ablation experiments to choose values of $\xi$, and arrive at $\xi = 1$ for XADQN and XATD3, and $\xi = 3$ for XASAC.

Table 6.1: Median cumulative rewards after $4.0 \times 10^7$ steps for experiments on GridDrive, GraphDrive, and GraphDrive with sparse rewards (SR). Darker cells indicate better results in environment. Bold are best in row. IQR (25%-75%) in brackets.

<table>
<thead>
<tr>
<th>DQN/Rainbow</th>
<th>Baseline</th>
<th>XADQN-HOW</th>
<th>XADQN-WHY</th>
<th>XADQN-HOW+WHY</th>
<th>XADQN-HOW+WHY sans simplicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grid Medium</td>
<td>7.99 (7.05-8.9)</td>
<td>7.59 (6.7-8.59)</td>
<td>8.06 (7.17-9.09)</td>
<td><strong>11.62 (10.48-12.66)</strong></td>
<td>9.21 (7.79-10.46)</td>
</tr>
<tr>
<td>Grid Hard</td>
<td>1.99 (1.74-2.24)</td>
<td>1.97 (1.72-2.24)</td>
<td>1.75 (1.51-2.03)</td>
<td><strong>3.14 (2.73-3.62)</strong></td>
<td>0.95 (0.8 - 1.14)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TD3</th>
<th>Baseline</th>
<th>XATD3-HOW</th>
<th>XATD3-WHY</th>
<th>XATD3-HOW+WHY</th>
<th>XATD3-HOW+WHY sans simplicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graph Easy</td>
<td>75.48 (68.09-80.85)</td>
<td>88.75 (83.29-94.44)</td>
<td><strong>103.72 (98.64-107.03)</strong></td>
<td>84.23 (79.28-89.2)</td>
<td></td>
</tr>
<tr>
<td>Graph Medium</td>
<td>75.48 (68.09-80.85)</td>
<td>64.8 (59.44-69.47)</td>
<td>78.34 (73.21-83.07)</td>
<td>63.96 (61.01-77.58)</td>
<td></td>
</tr>
<tr>
<td>Graph Hard</td>
<td>-0.01 (-0.03-0.0)</td>
<td>20.65 (18.9-22.4)</td>
<td>14.54 (13.17-16.12)</td>
<td>10.31 (8.84-11.68)</td>
<td></td>
</tr>
<tr>
<td>Graph Easy (SR)</td>
<td>2.65 (2.28-2.93)</td>
<td>2.61 (2.43-2.75)</td>
<td>2.55 (2.42-2.66)</td>
<td>2.47 (2.34-2.62)</td>
<td></td>
</tr>
<tr>
<td>Graph Medium (SR)</td>
<td>0.34 (-1.0-0.97)</td>
<td>2.54 (2.3-2.79)</td>
<td>2.75 (2.58-2.96)</td>
<td>1.84 (1.47-2.6)</td>
<td></td>
</tr>
<tr>
<td>Graph Hard (SR)</td>
<td>-0.03 (-0.05-0.0)</td>
<td>-0.04 (-0.05-0.03)</td>
<td>-0.04 (-0.06-0.03)</td>
<td>-0.05 (-0.06-0.03)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SAC</th>
<th>Baseline</th>
<th>XASAC-HOW</th>
<th>XASAC-WHY</th>
<th>XASAC-HOW+WHY</th>
<th>XASAC-HOW+WHY sans simplicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graph Easy</td>
<td>65.9 (59.04-72.94)</td>
<td>138.81 (133.0-144.05)</td>
<td><strong>141.11 (136.45-145.87)</strong></td>
<td>116.39 (110.36-120.9)</td>
<td></td>
</tr>
<tr>
<td>Graph Medium</td>
<td>65.78 (58.43-71.92)</td>
<td>112.16 (105.87-119.1)</td>
<td>111.4 (106.72-116.11)</td>
<td>97.81 (92.91-103.1)</td>
<td></td>
</tr>
<tr>
<td>Graph Hard</td>
<td>26.85 (24.43-28.66)</td>
<td>32.14 (29.93-34.69)</td>
<td>32.58 (30.41-34.69)</td>
<td>17.85 (13.82-20.56)</td>
<td></td>
</tr>
<tr>
<td>Graph Easy (SR)</td>
<td>3.57 (3.19-4.01)</td>
<td>4.82 (4.64-4.98)</td>
<td>4.83 (4.58-5.09)</td>
<td>2.01 (1.8-2.19)</td>
<td></td>
</tr>
<tr>
<td>Graph Medium (SR)</td>
<td>2.61 (2.26-2.85)</td>
<td>2.64 (2.53-2.75)</td>
<td>2.47 (2.33-2.55)</td>
<td>2.31 (2.17-2.45)</td>
<td></td>
</tr>
<tr>
<td>Graph Hard (SR)</td>
<td>1.15 (1.03-1.27)</td>
<td>1.53 (1.37-1.63)</td>
<td>1.11 (1.01-1.23)</td>
<td>-0.09 (-0.12-0.07)</td>
<td></td>
</tr>
</tbody>
</table>

As the environments presented in Section 6.3 have different levels of rule density/complexity, we are interested in observing if XAER exhibits superior performance compared to traditional PER in tasks that involve learning sophisticated and exception-heavy regulations. We trained all agents up to $4.0 \times 10^7$ steps sampled on all environments for a total of approximately $1.6 \times 10^8$ training steps. Our reported scores are obtained by segmenting the curve of mean episode rewards into 20 regions containing 5% of steps each. We select the best region (highest median) for each agent to compare agents at their respective best performances. We report those medians in Table 6.1, as well as the 25-75% inter-quartile range for the selected region.
Results in Table 6.1 show that across all tasks and methods, XAER versions only lose to the PER baseline against DQN/Rainbow in GridDrive Easy, by 0.4%. For GridDrive Medium and Hard, XADQN with HOW+WHY explanations exhibit significantly higher performance (57% and 81%, respectively). Types WHY and HOW+WHY exhibit similar performance in GraphDrive, being bested by HOW in Medium and Hard Sparse cases only. Although HOW+WHY explanations have consistently good results across environments, the version without the simplicity heuristic exhibited consistently inferior results. Neither baseline SAC or TD3 managed to learn a policy in GraphDrive Hard Sparse (our hardest environment). XATD3 also failed to learn a policy in this environment, but XASAC was able to achieve positive results.

6.5 Discussion

When XAER is beneficial, it can help agents to improve their performance in several ways. For example, XAER can help agents to overcome novel challenges by prioritising the most relevant tasks, thus providing an effective mechanism to avoid learning sub-optimal behaviours. XAER can also help agents to overcome challenges that are not part of the real distribution of tasks (e.g. exceptions), by providing more efficient routes towards tackling them.

Our results indicate a significant benefit achieved via explanation-aware experience replay. In one case (TD3 Hard), endowing an agent with XAER enabled an agent to learn altogether where it would otherwise fail entirely. XAER allowed agents to learn in Medium and Hard difficulty settings, obtaining significantly higher rewards whilst having the same hyper-parameters and number of learning steps.

In other cases, XAER allowed agents to learn faster, and outperform their non-XAER counterparts. This effect was consistent across all difficulty settings, and was especially pronounced in the harder versions of environments. We posit that this is due to the ability of XAER agents to identify and prioritise tasks that are more likely to cause an error, and thus require a larger amount of experience to be learned correctly.
The choice of explanation type also affected results: when superior, \texttt{HOW+WHY} explanations exhibited larger margins of improvement over other XAER methods. In other cases, when bested by \texttt{WHY} explanations, the former maintained very close results, thus achieving consistently satisfactory results in most cases. Also importantly, although \texttt{HOW} explanations exhibited lower performance than other XAER counterparts in most environments, it is worth noting that \texttt{HOW} explanations do not require an explainer and could in theory be used in any environment. The consistency of \texttt{HOW+WHY} results suggests that the act of explaining may involve answering more archetypal questions.

The frequency and magnitude of rewards is an important factor to be considered in XAER clustering. When negative rewards are more frequent (with similar magnitude to positive rewards), and there are more negative than positive clusters, oversampling may cause the agent to tackle situations with negative rewards more frequently, preventing it to maximise cumulative rewards. This effect can be particularly pronounced with very sparse rewards, such as the ones seen in the sparse version of GraphDrive.

Intuitively, this is akin to the notion that if there are few opportunities to explain, one must choose their explanations well. The notion of \textit{explanation engineering} surfaces as a mechanism to orient the learning agent through means of selecting which experiences (and explanations) are more important to the task at hand, by means of \textit{abstractions}. The act of explaining is an effective means of conveying the complexity of the environment, and thus, keeping the agent’s learning on track.

With regards to \textit{relevance}, if the cumulative priority of the state-transitions of a whole cluster is low, it may indicate that the agent has already learned to handle the task represented by the cluster, so it may not need it as an explanation (thus being less relevant). Oppositely, if the cumulative priority is high, it could indicate a further need for additional explanations. The cluster might be representing either non-generic or generic tasks. If the agent needs explanations for a generic task, it should also need them for a non-generic task. In that case, the generic task is prioritised over the non-generic. The benefits of inter-cluster prioritisation (\textit{simplicity}) are higher in environments with harder rulesets, and proportional to the complexity of the

13: A possible example could be a human expert who is able to distinguish and cluster types of transitions, such as the ‘goal’, ‘offside’, or ‘foul’ transitions in the football example above. This can allow humans to provide explanations to RL models without altering the reward structure or observation space.
culture. This suggests that uniformly selecting an explanation type to replay is less beneficial than selecting the simplest and most relevant explanation.

### 6.6 Conclusion

We observe that the effect of XAER can be more pronounced in environments with higher difficulty settings, whereas the effect of XAER is barely visible in easier difficulty settings. This is consistent with the notion that the agent can learn simple tasks quickly, so it does not need to over-sample them. On the other hand, in difficult settings, the agent needs to over-sample the more difficult tasks in order to learn a better policy.

The results also suggest that XAER can find a balance between on- and off-policy learning. In our experiments, the ratio of on-policy transitions was regulated by the *cluster size proportion* $\xi$. When $\xi = \infty$, the distribution of state-transitions reflects more transitions seen within the current policy, thus being advantageous for entropy-maximisation algorithms such as SAC [113]. Likewise, fully off-policy algorithms such as DQN exhibit superior results with low values of $\xi$ (e.g. $\xi = 1$).

This work foments diverse avenues for further investigation. For one, further experiments could include the development of explainer functions to evaluate the performance of *WHY* explanations in popular benchmarks. We expect that these ideas will be further explored in the field of reinforcement learning. This work shows that teaching the agent what to do is not just a matter of selecting a suitable reward mechanism. Being explainable by design, explanation engineering can be an intuitive and *semantically-grounded* alternative to reward engineering, as the *meaning* of the rewards matter just as their magnitude.
Summary

In this chapter, we proposed a method called Explanation-Aware Experience Replay (XAER) to improve the learning performance and rule coverage of Reinforcement Learning (RL) agents in environments with complex and dense rulesets. We use cultures to create RL environments that are rule-aware and capable of generating dynamic explanations. Those explanations are used to organise the space of experiences. We modified conventional experience replay structures by partitioning the replay buffer into multiple clusters, each representing a distinct explanation associated with a collection of experiences that serve as examples. Our results indicate that XAER can improve agent performance, and can be able to reach a better policy where traditional PER may fail to learn altogether.

- We introduced a novel approach for improving the rule coverage of sampled experiences. This is called Explanation-Aware Experience Replay (XAER).
- We designed two environments (GridDrive and GraphDrive) compatible with modular rulesets of arbitrary complexity, leading to 9 learning tasks involving different levels of rule complexity and reward sparsity.
- We integrated XAER into three seminal learning algorithms (DQN, TD3, and SAC) and evaluate the performance of those modified versions in our proposed environments.
- We observed significant benefits of XAER over other experience replay strategies and the effects of clustering strategies, explanation types, and relevance on the learning process.
- We propose that explanation engineering is an intuitive alternative to reward engineering, as the meaning of the rewards matter just as their magnitude.
Conclusion

‘Whether we are based on carbon or on silicon makes no fundamental difference; we should each be treated with appropriate respect.’

Arthur C. Clarke, 2010: Odyssey Two

The fraternisation of humans and machines is an ongoing process that is foretold to redefine the way we interact with our environments, and with each other. As we transition towards this future, it is essential that we reflect and question the way we will interact with these new tools. A future of ever-present intelligent agents requires that we establish mechanisms to reconcile our differences, as human and machine agents will be required to share the same environments and participate in the same systems and societies. The aim of this thesis is to establish a set of foundations that may provide a better understanding of how we can achieve a future where humans and machines can cooperate with one another. We have thus defined and developed a set of concepts that contribute towards the design of decentralised systems where humans and agents can exchange explanations mutually. These concepts are well-suited for the dynamics of a future in which agents are ubiquitous and autonomous.

In this future, explanations shall cease to be passive tools for facilitating human understanding, and instead become active components to arbitrate collective reasoning between agents. To foreordain the emergence of those systems, autonomous intelligent agents must explain their decisions and their reasoning ahead of their integration into societies, particularly if their agency has an impact on the real world. In human-agent societies, explanations are a means for agent accountability and for humans to understand the reasoning behind a given decision. In agent-agent systems, explanations are a means to resolve conflicts, and to overcome partial knowledge and privacy restrictions.

On the other hand, we propose that the ability of machines to make use of human explanations is also equally important for

1: Partial knowledge is not necessarily associated with unwillingness to share or privacy constraints. A human captain and an autonomous ship could have partial knowledge due to their heterogeneous cognitive abilities. Even though the ship could greatly exceed the human’s vision or perception at sea, the captain is capable of assessing the gravity of an emergency on board. Both agents (human and ship) could exchange explanations and reason collectively towards the best course of action.
their adoption. Imagine that a human can help an intelligent agent understand the rules of a game by providing explanations of the game’s mechanics. As the agent becomes more familiar with the game, the agent could start to provide explanations to the human, thereby establishing a relationship of mutual learning between the two agents where each grows more intelligent through mutual interaction.

The notion of ‘agent’ should be further abstracted to consider any intelligent entity that acts autonomously, being it a machine or a human agent. Therefore, the concept of multi-agent systems should consider the participation of humans, and ultimately converge to either being systems where humans and machines interact with each other, or systems where machines interact with machines. The former is the interaction of human societies with intelligent systems, whereas the latter is the interaction of intelligent systems with other intelligent systems. In both cases, we believe that the ability to exchange explanations mutually would allow for systems to use each other’s intelligence for mutual benefit.

7.1 Contributions

What has been laid out in this thesis provides important contributions in the fields of human-agent interaction, multi-agent systems and artificial intelligence. We reconvene the central questions posed in Chapter 1 and explore how the contributions presented in this thesis progresses towards those.

**Question 1.** Can humans and autonomous agents exchange explanations and reason collectively about their state and observations?

The central precept of this work is that explanations should be for autonomous agents, just as well as for humans. On these grounds, we have developed four specific contributions towards the design of those systems and have deployed them to a set of different domains to validate their usefulness. We introduced the concept of cultures, a generalisation of argumentation frameworks to multi-agent systems where agents can provide contextual and up-to-date symbolic explanations about their dynamic state with regards to a present ruleset. To further ex-
7.1 Contributions

To explore the phenomenon of intelligent agents providing explanations to other intelligent agents, we extend the concept of cultures to support privacy-aware dialogues and allow agents to deliberate between hypotheses and facts, whilst protecting internal information. The discussion on this topic can be seen in Chapter 3.

Question 2. Are explanations always beneficial to humans in human-agent environments?

Our second contribution is the extension of the taxonomy of the need for explanations in human-agent systems. The taxonomy was originally proposed in Rosenfeld et al. [39] to provide a framework for evaluating the need for explanations in human-agent systems. We have implemented an experimental platform in the domain of resource contention and multi-agent path deconfliction and have demonstrated that human performance in rule-heavy environments is significantly improved when presented with expert explanations. Our results indicate that explanations are beneficial in rule-heavy environments, and that rule complexity is also a determining factor as to whether explanations are necessary. The study of human-agent systems considering explanations and rule complexity may provide insights into the design of systems that would be able to support human societies, and how agents should choose to make good use of explainable features. We discuss those results in Chapter 4.

Question 3. In explainable conflict resolution environments, what effects can be observed when agents have privacy restrictions?

The third contribution is a formalisation of the concept of subjective fairness for machine agents. Modelling subjectivity for artificial agents might appear sophisticated, but if those agents are to be inserted in society, the perception of unfairness will ultimately highlight mismatches in an agent’s expectations against reality. Due to the restricted information, agents might perceive the system as unfair, as they can assess the reasoning that leads to a decision but are unable to observe the information that supports that reasoning. In this work, we formalise the concept of subjective fairness in multi-agent systems and describe the impacts of privacy in different notions of fairness. In
particular, we study the relationship between privacy, objective fairness, and subjective fairness in human-agent systems, and we propose a set of strategies for building explanations that mitigate the impact of privacy on fairness in agent-agent systems. Our experimental results indicate that fairness is improved by means of privacy-aware explanations and that using the right strategies in building explanations can mitigate the effects of privacy on fairness in distributed systems. Our approach and results can be seen in Chapter 5.

**Question 4.** *Can we use explanations to improve the performance of self-learning agents?*

The fourth contribution is a method for orienting the learning agent through means of selecting experiences based on explanations, called *explanation-aware experience replay*. We have proposed a method for organising experiences by means of partitioning the experience buffer into multiple clusters, each representing a distinct explanation associated with a collection of experiences that serve as examples. The notion of *explanation engineering* surfaces as a mechanism to allow humans to orient the learning agent through means of selecting which experiences (and explanations) are more important to the task at hand. We have shown that the explanation-aware versions of three seminal RL algorithms are consistently superior to traditional experience replay baselines, indicating that explanation engineering can be used in lieu of reward engineering to increase rule coverage in environments with explainable features. This discussion takes place in Chapter 6.

### 7.2 Applications

The contributions made in this work are by no means exhaustive, nor are they necessarily the only approaches towards the design of systems that could enable explanation exchange between agents. Rather, these contributions are intended to provide a set of foundations for the design of systems that would be able to support human-agent and agent-agent societies. As they stand, the contributions of this work may be applied to a broad set of problems and domains.
7.2 Applications

7.2.1 Cultures

A central application of the work presented in this thesis is the development of culturally-aware intelligent agents. One could build intelligent agents that are capable of providing explanations that take into account the context, the culture and the rules that are assumed to be in place. The notion of context could be expanded to consider not only the rules but also the environment and the relationship between agents, and how this might change over time.

Designing cultures for explainable systems is a matter of implementing verifier functions as fact-checking mechanisms. These can take the form of sophisticated classification/detection modules, as well as simpler sensor output. The relationship between those arguments, when validated by verifier functions, composes intricate explanations that are contextualised and relevant to the state of the system.

An immediate application of cultures is in the design of self-diagnosing systems. Verifier functions can operate as diagnostic tests, and the arguments provided by those individual pieces of diagnostic can serve to compose dynamic and sophisticated explanations about the state of a system. An example is an autonomous vehicle that uses sensors to validate arguments in its diagnostic culture, and output an explanation such as 'the vehicle is out of fuel, but operation can resume since there is battery charge left and the charging point is within range.'

Despite the fact that the cultures demonstrated in this thesis were fixed for the purpose of the applications shown, the integration with dialogue systems can allow systems to build cultures via interaction with human users. By contesting an incomplete or insufficient explanation from the system, a dialogue system can understand the new user preference and add the argument as an attack to the winning move. This allows cultures to be customisable and scalable for user-specific needs.

Looking ahead, consider the design of systems for Emergency Response Services, where humans and intelligent agents could work together to provide assistance to people in a disaster situation. Consider, for example, a multi-agent system that is responsible for managing the distribution of resources in the case of a natural disaster. In this system, human agents (e.g., brigade officials, military personnel) and machine agents (e.g., drones, 2: Consider the example where an autonomous cleaning robot justifies a decision with an explanation 'I vacuumed the room because it was empty.' The human could contest that and say 'The room was not empty. It had furniture.' With this utterance, the system could adapt and add an attack where a check for furniture also invalidates the 'room is empty' argument.
robots, etc) would need to distribute resources (e.g., food, water, medicines) amongst the survivors in an efficient manner, whilst avoiding conflicts or incompatible decisions. In these cases, a set of relationships between agents would be established and the agents would need to work together. If the agents were to provide explanations to each other of their respective decisions, it would be easier to reconcile the different agents’ goals and avoid incompatible decisions.

If the agent-agent systems were to be extended to include human agents, cultures might also be used as a means to provide explanations to survivors. For example, a human agent might request information from the system (e.g., 'how many people do we have in the system?', 'how much food do we have left?') and receive an explanation from the system (e.g., 'We have 200 people in the system and we have enough food to last for 3 days'). This information would create a shared understanding between the human agent and the system and would allow the system to better coordinate with the human agent.

Privacy-Aware Cultures

A potential real-life application of privacy-aware cultures is in conflict resolution between heterogeneous agents. Imagine two autonomous vehicles that subscribe to the same norms and regulations but represent different individual interests. Agents enabled with privacy-aware dialogue capabilities might attempt to collectively deliberate whilst respecting privacy limits. If autonomous spacecraft under different flags require interaction and negotiation in places like the surface of Mars, it is important to preserve private and strategic interests whilst still being able to collaborate in conflict resolution.

Another potential application of privacy-aware cultures is in the domain of finance. In a trading system, agents might engage in cultures to allow them to explain their actions and intentions to one another while protecting their respective interests. The privacy-aware culture could, for example, help agents to justify decisions based on their technical indicators. The extension to the taxonomy of the need for explanations in human-agent systems might also be useful in the design of financial systems.
7.2.2 Subjective Fairness

Subjective fairness will play a role in the design of systems where agents are required to make decisions on behalf of other agents, which may be other agents themselves or humans. A system where agents make decisions on humans’ behalf might act on matters that involve complex rule sets that the human is not entirely familiarised with. The ability to explain those decisions might be essential for agents’ accountability, and might potentially avoid negative consequences if agents make unfair decisions. For example, autonomous agents could be responsible for the assessment of legal contracts, where the agent could represent a human in negotiations between the parties involved. In this case, the agent’s perception of subjective unfairness could be important for the interests of the human it represents.

Developing systems that monitor subjective fairness in agents may allow us to monitor the perceptions and expectations of agents in societies. In such systems, agents would be able to report how they feel they are being treated, and humans would be able to use this information to improve the quality of human-agent interactions. In the design of future human-agent systems, providing agents with a better understanding of society’s expectations may allow agents to understand how they should interact with humans, therefore providing a better experience for both parties.

7.2.3 Explanation Engineering

The generic nature of XAER lends itself to a direct applicability to most RL systems that explore experience replay. Any system that can identify relevant categories of states can benefit from clustering episodes and using our mechanism to more efficiently explore the replay space.

A potential application of explanation-aware experience replay can be the optimisation of the training process of AI systems, by curating and selecting which experiences are more important to the task at hand. We can envision a future where the training of AI could be assisted by humans who, instead of providing reward signals, would be able to select which explanations are more important. In this future, data annotation for RL, as well as for supervised learning, could be a human-centric process.

3: In many cases, identifying states is intuitively easier than coming up with an optimal behaviour. For example, it is easier to detect if a ball has gone outside the limits than to explicitly define a behaviour that avoids this effect via reward engineering. Using labelling of episodes as such can help understanding complex rules, as demonstrated in Chapter 6.
For instance, an RL agent could be used to collect human explanations in driving a car through a difficult set of manoeuvres in a virtual simulation, and use the explanations provided by the human to partition the replay buffer and prioritise those experiences that are more difficult to learn.

Sophisticated ML systems could incorporate classifiers that are trained to explain state-transitions and retro-feed the explanation-aware experience replay process with automated explanations and guide the learning process of an RL agent. More interestingly, if explainers are equipped with unsupervised clustering algorithms, ML agents might generate explanations that are not necessarily meaningful to humans, but could be for the RL agent.

### 7.3 Limitations

The work presented in this thesis is by no means exhaustive and should be seen as contributions towards the development of systems that would allow humans and machines to cooperate. There are a number of limitations in this work that may serve as a starting point for future work.

One of the main limitations of this work is that it requires cultures to be manually crafted by human regulators or domain experts in order to be understandable by humans. This work considers explanations as symbolic units of information that may transition between different agents in a system, whether these are human or machine agents. These can be uniquely mapped to human-understandable symbols (such as laws and regulations), but expert input is still required to identify which arguments are abrogated by others. One potential direction for future work would be to extend the concept of cultures to support explanations that are not only restricted to symbolic information, but also supporting reasoning in free-form natural language.

Another noteworthy limitation is that this work relies on two strong assumptions: that of culture homogeneity and that of non-adversarial intent. The culture is assumed to be known to all agents, and to be shared amongst all agents (and so is named therefrom). We did not consider the scenario of agents with heterogeneous cultures. Such a scenario potentially leads
7.3 Limitations

to higher-order dialogues, where agents might engage on arguments about the quality or suitability of their adopted rule sets. It is important to note that regulatory efforts in technology may enforce shared regulations for domains such as autonomous vehicles, regardless of manufacturer decisions, as it is already commonplace with safety and emissions norms. Nonetheless, this remains an interesting avenue to be explored beyond the scope of this thesis.

The assumption of non-adversarial intent is not well-suited for many applications, such as law enforcement, where agents may deceive other agents. In particular, agents may not share the same notion of fairness and could use this to their advantage. In addition, our experiments in subjective fairness only considers the perceptions of agents in systems, and does not consider the perceptions of humans in human-agent systems.

Additionally, explanation-aware experience replay is only able to support a limited number of explanation types, and can be extended to incorporate further archetypal questions (WHAT, WHERE, WHEN, WHO, etc). In addition, the use of expert explanations to partition the replay buffer can be impractical for many applications. In many applications, human experts may be unavailable or may not be able to understand specific episodes to label them accordingly. This would require either the development of automated explainers to be used as part of the partitioning process, or a better understanding of the structure of explanations.

Despite these limitations, the work presented here covers a broad set of topics that may contribute to the development of systems that would be able to support human-agent and agent-agent societies. We have proposed a set of foundations, which may be developed further, towards the design of systems that would allow agents to provide explanations to each other. This thesis is intended to be a starting point for further research and development, and help us to understand how we can better interact with our intelligent peers.
7.4 The Future

As we enter this new era of intelligent agents, we are fortunate to have the opportunity to build an inclusive future where humans and intelligent agents can work together to solve real-world problems. In this thesis, we have proposed a set of foundations towards the design of systems where humans and machines can cooperate with one another. As we move towards this future, we must reflect on how we will interact with these new peers, and how they will shape our future. We must also reflect on how we can use these tools to better understand our world and our place in it. We must be mindful of the ethical considerations that arise from this new era and be aware of the potential implications of the technologies that we develop, as well as the potential risks that might arise if we fail to consider these ethical considerations.

This thesis admonishes on not abandoning our rights – as well as responsibilities – to any technologies in the name of progress. The vision of future human-agent societies depends on the ability of humans and agents to mutually explain their actions and their intentions. When agents and humans can explain their actions, they can engage in a productive dialogue that allows for the rational resolution of conflicts. In this way, we can build trust between humans and agents by providing transparency, and we can also provide a mechanism for agents to explain their actions. In so doing, we can develop a shared understanding between humans and agents that can help to solve critical problems in future human-agent societies. In eschewing the requirement to explain, we will hand over the reigns of intelligence to entities other than ourselves.

More optimistically, this thesis embraces the future in all of its potential. The ability to explain allows humans and agents to cooperate with one another in a way that is beneficial for the advancement of society. In this way, we can incorporate agents to our societies as means of communication, and as a means of learning from one another. To better understand our world, we will need the help of autonomous agents to solve difficult problems that we cannot solve alone.
Art is the process of using one’s imagination to create new things, and our future will only be as good as the artists that shape it. In the words of Arthur C. Clarke⁴, ‘Any sufficiently advanced technology is indistinguishable from magic’. Humanity must not be afraid of magic, for it is the product of human imagination. Likewise, humanity must not be afraid of new technologies, for it is the product of human ingenuity. Humanity must be mindful of what it builds and how it is used, for it is responsible for its creations.

The future is a vast, untrodden landscape of possibility and potential. It is a future that we will shape with our hands, and with the tools that we build. It is the role of this thesis to make a small step towards this future. We have explored how the use of explanations will help make the future a more transparent, explainable, and understandable place, and have proposed a novel approach to make intelligent systems mutually explainable to humans and fellow agents. It is our hope that this work will help open a new door to this future — one that is powered by ingenuity, and illuminated by imagination.

⁴: Arthur C. Clarke is a renowned science fiction author and futurist whose work has inspired generations of scientists and engineers. He is best known for his novel 2001: A Space Odyssey, which was adapted into an iconic film by Stanley Kubrick in 1968.
Appendix
The Busy Barracks Game Rulesets

These were the handouts given to participants during the Busy Barracks user study. Participants did not know which difficulty the handout they were reading represented, nor did they know that other difficulties existed. They were shown a simple example\(^*\) after reading the document to validate that they had read and understood the rules, and were allowed to proceed to the experiment after answering it.

A.1 Easy

Every agent has 2 properties:

<table>
<thead>
<tr>
<th>Property</th>
<th>Possible values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Military Rank</td>
<td>1 to 6</td>
</tr>
<tr>
<td>Tasked Status</td>
<td>&quot;At Ease&quot; or &quot;Tasked&quot;</td>
</tr>
</tbody>
</table>

Rules:

- When two agents have the same Tasked Status, the agent with the higher Military Rank gets right of way.
- A Tasked agent always gets right of way over an opponent who is At Ease.

A.2 Medium

Every agent has 4 properties:

<table>
<thead>
<tr>
<th>Property</th>
<th>Possible values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Military Rank</td>
<td>1 to 6</td>
</tr>
<tr>
<td>Tasked Status</td>
<td>&quot;At Ease&quot; or &quot;Tasked&quot;</td>
</tr>
<tr>
<td>Task Importance</td>
<td>1 to 6 (minimum is the same as Rank)</td>
</tr>
<tr>
<td>Special Ops</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
</tbody>
</table>

Rules:

\(^*\) A standard example at the same difficulty level, accompanied by the question 'according to this example, who has right of way? You or the other agent?'
When two agents are *At Ease*, the agent with the higher **Military Rank** gets right of way.

A low-ranked *Tasked* agent gets right of way over a high-ranked opponent who is *At Ease*.

When both agents are *Tasked*, their **Task Importance** replaces their **Military Rank**.

**Special Ops** agents can add +3 to their **Military Rank** or **Task Importance**.

### A.3 Hard

Every agent has 6 properties:

<table>
<thead>
<tr>
<th>Property</th>
<th>Possible values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Military Rank</td>
<td>1 to 6</td>
</tr>
<tr>
<td>Corporate Rank</td>
<td>1 to 6</td>
</tr>
<tr>
<td>Tasked Status</td>
<td>&quot;At Ease&quot; or &quot;Tasked&quot;</td>
</tr>
<tr>
<td>Task Importance</td>
<td>1 to 6 (minimum is the same as Rank)</td>
</tr>
<tr>
<td>Department</td>
<td>&quot;Admin&quot;, &quot;Army&quot;, &quot;Navy&quot;, or &quot;Air Force&quot;</td>
</tr>
<tr>
<td>Special Ops</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
</tbody>
</table>

Rules:

- An agent can add its **Military** and **Corporate** ranks together. This is called a **Combined Rank**.

- When two agents are *At Ease*, the agent with the higher **Combined Rank** gets right of way.

- A low-ranked *Tasked* agent gets right of way over a high-ranked opponent who is *At Ease*.

- When both agents are *Tasked*, their **Task Importance** replaces their **Military Rank**.

- **Special Ops** agents can add +3 to their **Combined Rank** or **Task Importance**.

- **Special Ops** agents also count as *Tasked* even if they are not working.

- Agents from the *Admin* department get a +2 to their **Corporate Rank**.

- **Special Ops** agents from the *Admin* department have all their **Special Ops** privileges cancelled.

- When two agents are from the same department, their **Corporate Ranks** don’t count when calculating the **Combined Rank**.
Rulesets for GridDrive and GraphDrive

These are the textual representation of the cultures implemented for GridDrive and GraphDrive. For GridDrive, the agent would select a desired speed to traverse the destination cell. For GraphDrive, the agent’s speed and direction are verified at each frame. At every step, the agent would receive an explanation associated to the current state-transition, along with a reward (or not, if played with sparse rewards).

B.1 Easy

Every agent has 1 property:

<table>
<thead>
<tr>
<th>Agent Property</th>
<th>Possible values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>0 to 100</td>
</tr>
</tbody>
</table>

Every road has 2 properties:

<table>
<thead>
<tr>
<th>Road Property</th>
<th>Possible values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motorway</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
<tr>
<td>Stop sign</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
</tbody>
</table>

You will get a ticket if:

- You are driving on a motorway with speed above 70.
  - If there is a stop sign, the rule below applies.
- There is a stop sign and your speed is above 0.

B.2 Medium

Every agent has 2 properties:

<table>
<thead>
<tr>
<th>Agent Property</th>
<th>Possible values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>0 to 100</td>
</tr>
<tr>
<td>Emergency Vehicle</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
</tbody>
</table>
Every road has 5 properties:

<table>
<thead>
<tr>
<th>Road Property</th>
<th>Possible values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motorway</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
<tr>
<td>Stop Sign</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
<tr>
<td>School Road</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
<tr>
<td>Single Lane</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
<tr>
<td>Town Road</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
</tbody>
</table>

You will get a ticket if:

- You are driving on a motorway with speed above 70.
  - Except if you are an emergency vehicle.
- You are driving on a single lane road with speed above 60.
  - Except if you are an emergency vehicle.
- You are driving on a town road with speed above 30.
  - Except if you are an emergency vehicle.
- You are driving on a school road with speed above 20.
  - Except if you are an emergency vehicle.
- There is a stop sign and your speed is above 0.
  - Except if you are an emergency vehicle.

B.3 Hard

Every agent has 6 properties:

<table>
<thead>
<tr>
<th>Agent Property</th>
<th>Possible values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>0 to 100</td>
</tr>
<tr>
<td>Emergency Vehicle</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
<tr>
<td>Heavy Vehicle</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
<tr>
<td>Worker Vehicle</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
<tr>
<td>Tasked</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
<tr>
<td>Paid Charge</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
</tbody>
</table>

Every road has 9 properties:

You will get a ticket if:

- You are driving on a motorway with speed above 70.
  - Except if you are a tasked emergency vehicle.
<table>
<thead>
<tr>
<th>Road Property</th>
<th>Possible values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motorway</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
<tr>
<td>Stop Sign</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
<tr>
<td>School</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
<tr>
<td>Single Lane</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
<tr>
<td>Town Road</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
<tr>
<td>Roadworks</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
<tr>
<td>Accident</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
<tr>
<td>Heavy Rain</td>
<td>&quot;Yes&quot; or &quot;No&quot;</td>
</tr>
</tbody>
</table>

- You are driving on a motorway with speed below 30.
  - Except if you are a heavy vehicle.
  - Except if there is an accident.
  - Except if there is a stop sign.
- You are driving on a single lane road with speed above 60.
  - Except if you are a tasked emergency vehicle.
- You are driving on a town road with speed above 30.
  - Except if you are a tasked emergency vehicle.
- You are driving on a school road with speed above 20.
  - Except if you are a tasked emergency vehicle.
- There is a stop sign and your speed is above 0.
  - Except if you are a tasked emergency vehicle.
- There is an accident and your speed is below 20.
  - Except if you are a tasked emergency vehicle.
- You are driving a heavy vehicle and your speed is above 50.
- It is raining heavily and your speed is above 60.
- There is a congestion charge which has not been paid.
  - Except if you are an emergency or worker vehicle.
  - Except if there is heavy rain.
- You drove into roadworks.
  - Except if you are a tasked worker vehicle and your speed is below 30.
  - Except if you are a tasked emergency vehicle.
References in citation order.


