Simulated Emotions and Mood as part of Decision Making in a Mobile Agent Society

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Joe Maurice Collenette

January 2020

Abstract

This thesis explores cooperation among self-interested agents that use simulated emotions and mood as part of their decision making. The thesis begins by analysing the psychology literature on emotions and mood, and the models developed. A discussion follows that shows the inherent difficulties in developing these models so that they can be used by computational agents. The first set of experiments shows how a society of emotional agents are affected by the introduction of mobility, in terms of cooperation in the Prisoner's Dilemma and the effects on which emotional characteristics are successful when compared to previous implementations.

The thesis also introduces a new computational model of mood which is demonstrated in two separate implementations. Both implementations are tested in the Prisoner's Dilemma and show improvements in cooperation when compared to the equivalent implementation that does not use the model of mood, where the agents implementing these models are placed in mobile environments. The experiments augment the previous findings that mobility and the environment structure has an an effect agents.

The next set of experiments in the thesis highlighted that mobility reduces the level of cooperation and increases convergence times towards cooperation for the agents. The more open the arena, the more impact the effect of mobility has on the agents. The average scores were noted to only be impacted when the payoff matrix supports this effect.

An Evolutionarily Stable Strategy analysis was conducted on the emotional and moody agents that have been implemented. Emotional agents can be considered stable when allowed time to interact with their neighbours. Moody agents were shown not to be stable. The stability of the emotional agents came at the cost of cooperation, as the ability of moody agents to create cooperation creates the risk that cooperation will not be reciprocal.

In conclusion, the thesis has showed that self-interested emotional and moody agents are able to be both competitive and cooperative in environments where mobility exists.

Acknowledgements

While my name maybe on the front of this thesis, the work produced could not have been achieved without many people's help and support. There are many people I wish to thank, and apologies for any people that I have missed.

Firstly I would like to thank the people who have enabled me to be able to complete my Ph.D. The States of Guernsey education department for contributing significantly to the cost of the degree. My grandfather and my parents for both monetary assistance and support over the years.

There are too many friends in the department over the years to individually thank everyone. Our discussions have been invaluable, including our lunchtime discussion rabbit holes that we often got into. These have often helped keep my morale up during the many tough aspects of completing my Ph.D.

A notable thank you must be given to Hayley, who has managed to put up with me having spent plenty of late nights working, and coping with many changed plans and holidays caused by impromptu work commitments.

Kerry and Scott have been invaluable throughout my Ph.D. by pulling me out of the computer science bubble for brief moments. The coffees, lunches, and beers we have shared have been indispensable. I wish you both the best for your Ph.D.'s

I would also like to thank Dr. Daan Bloembergen who has provided support and guidance in my Ph.D. especially through the formative years when I was getting to grips with research. I have thoroughly enjoyed working with you throughout the Ph.D.

I wish to also thank my second supervisor Professor Karl Tuyls whose input and guidance has been invaluable in shaping this research and my development as a researcher. I have learned a huge amount, and enjoyed our time working together.

A final thank you to my supervisor Professor Katie Atkinson. The number of things to thank you for is extremely long. Your unwavering belief has encouraged me throughout the Ph.D. Your invaluable support and guidance has let me not only develop as a researcher, but occasionally even allowed me to enjoy my Ph.D.

Contents

List of Figures

List of Tables

Notations

The following notations and abbreviations are used throughout this thesis.

RQ Research Question

SRQ Sub-research Question

AI Artificial Intelligence

SARSA State Action Reward State Action

OCC Ortony, Clore, and Collins

PANAS Positive and Negative Affects Schedule

PAD Pleasure-displeasure, Arousal-non-arousal, and Dominance-submissiveness

OCEAN Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism

PD Prisoner's Dilemma

SH Stag Hunt

TFT Tit-For-Tat

WSLS Win-Stay Lose-Shift

ALLCOOP Always cooperate strategy

ALLDEFECT Always defect strategy

GTFT Generous Tit-For-Tat

R Payoff for an agent in mutual cooperation in a social dilemma

P Payoff for an agent in mutual defection in a social dilemma

T Payoff for an agent that takes advantage of the opponent in a social dilemma

S Payoff for an agent in mutual cooperation in a social dilemma

$Q(s,a)$ Q Value for a given state s and action a

s A state

a An action

St The set of all states

A The set of all actions

Iv Invasion strategy

t Time

r Reward

α Learning Rate

γ Discount Factor

$F(s,a,s')$ Function that returns a modified reward for the given state, action, and next state

$V(Sy,Sy')$ Function that returns the expected payoff strategy Sy receives when playing strategy Sy'

Rp Repeating set of Prisoner's Dilemma Outcomes

Cm Mutual cooperation

Dm Mutual Defection

I_i Initial action of agent i

Γ_i^j Mutual action between agents i and j

Ac_i Action of agent i

C Cooperation action

D Defection action

An_i Anger threshold of agent i

G_i Gratitude threshold of agent i

Ag The set of all agents

m_i^t Mood value of agent i at time t

$\eta_{i,j}$ Number of interactions between agent i and j

$E_{i,j}^t$ Action agent i takes against agent j at time t based on the current simulated emotional values

Ξ Function that returns the mood value as a real between 0 and 1, where 0 is a mood value of 100.

$\Omega_{i,j}^t$ Homo Egualis equation

μ_i^t Average payoff of agent i at time t

F_i^t Opponent of agent i at time t

Ψ Function that returns the average payoff for a given percentage of previous outcomes

$Mem_i^a(n)$ Function that returns the payoff agent i received when playing action a at interaction number n

β Function that returns the number of values in Mem that should be uses for Ψ

ϵ A small value

MA Value of Mood Affect

v Value given to MA when set to variable

MV The set of possible mood levels

o Oracle strategy

e Emotional agent strategy

k Moody agent strategy

n Number of rounds

Chapter 1

Introduction

The ability to cooperate with others has allowed humans to flourish as a society. This ability to cooperate derives from an evolved moral system, which requires the need for emotional responses [38]. An individual's choice of deciding whether to cooperate is influenced by the person's emotional state [4, 8, 11]. The decision that is made through an individual process may differ, even when the decision is presented in two identical settings. One of the reasons why the decision may differ is a difference in how the person felt about the counter-part they are choosing to cooperate with or not [52]. Mood is often inter-twined with emotions, with mood influencing emotions and emotions influencing mood [122]. Emotions and mood in the general case are considered a fundamental part of the decision making process in humans.

For example, people have been shown to give away more money than the theoretical minimum payoff in a well known economic experiment (the ultimatum game), which therefore leaves them worse off [55]. However the amount that is given away varies depending on their social environment [88], showing how different moral systems, which are impacted by emotions and mood, can change people's desire to cooperate.

Being able to translate what is considered to be a fundamental part of a person's decision-making into computational systems so that computational agents are better able to cooperate with each other is a topic of relevance for the AI and multi-agent systems' community. The literature has explored the implementation of such agents and how they have succeeded in creating and sustaining cooperation [69, 3]. However we need to be better at translating results that have come from specific implementations into more general results, so that the results become more relevant to the wider community.

Exploring cooperation and decision making processes is an open topic in multi-agent systems. Multi-agent systems have a need to enable cooperation, as multiple agents will often be working to achieve one or many tasks; there has been significant experimental work on specific simulated emotions in multi agent systems [72, 62, 82]. In contrast, simulated mood has often taken a back seat. Simulating mood in conjunction with simulated emotions in the context of multi-agent systems enables a deeper simulation of the emotional process that affects human decision making.

For multi-agent systems to be implemented in real-world scenarios, how the agents should be implemented in the task environment needs to be identified and the effects shown. Humans growing up in different environments will lead to the development of different moral systems, which in turns affects individuals' decision making and their emotional state in different ways [38, 55]. There has not been significant work on how different environments and their structures can affect decision making in multi-agent systems. This thesis therefore aims to yield insights into how the environment affects decision making, with a particular focus on agents that use simulated emotions and mood.

1.1 Research Question

When considering societies of agents that have the ability to move around physically, this thesis is interested in whether agents which implement simulated emotions and/or mood as part of their decision making process can be effective in cooperation. A further consideration is whether the agents cooperate even when individual agents are selfish in their own motivation, i.e. competing to maximise their own payoff. These questions lead to the main research question (**RQ**) of this thesis:

> *Can agents with simulated emotions and simulated mood be cooperative and competitive in a self-interested, mobile society?*

The research question is broad and in order for this thesis to answer the question it is further decomposed into three sub research questions. The three sub research questions are:

SRQ1 *How can we computationally model decision making using emotions and mood?*

SRQ2 *How can we capture cooperation and competitiveness in emotional and moody agent societies?*

SRQ3 *How does mobility affect agents in their interactions within a society?*

To answer the sub research questions, a number of experiments are constructed to test the agents in a number of different environments. The agents themselves use simulated emotions and simulated mood. The simulated emotions are based on previous implementations to allow for effective comparisons between mobile and static agents. The simulated mood includes a computational model of mood which is grounded in psychology, to effectively justify the choices for the implementation of mood. This leads to the second sub research question which requires the agents to interact in a social dilemma, notably the Prisoner's Dilemma, and also requires study of the theoretical concept of evolutionarily stable strategies. To fully answer the final sub research question, the number of strategies that the agents use in the mobile arenas needs to be increased. The experiments to address this are designed around an arena structure, to fully reveal what effects the structures have in the general case.

The sets of experiments conducted to investigate the sub research questions combine together to answer the main research question. Collective consideration of what the experiments show about mobility, simulated emotions and simulated mood, and the evolutionarily stable strategy analysis provides the results needed to answer the main research question. The provided theoretical analysis of the evolutionarily stable strategy shows that simulated emotional agents can be competitive in the Prisoner's Dilemma. The Prisoner's Dilemma captures cooperation among self-interested agents. This is supplemented with a number of experiments in a variety of environment structures, including with and without mobility. Agents using simulated emotions and mood as part of their decision making process are shown to allow cooperation in each of the environments studied, although there are differences caused by the environmental structure. The results show that self-interested emotional and moody agents can be both cooperative and competitive in mobile agent society, answering the overall research question **RQ**.

1.2 Overview

Emotions and mood are a fundamental part of the decision making process in humans. For advances in AI and automated decision making to continue, understanding and applying emotional simulations into both decision making and learning are investigated in order to make effective decisions. In the state-of-the-art there remain difficulties in implementing effective automated decision making, to enable working together, ethical decision making, and fast effective learning. When people are faced with making these kinds of decisions they will rely on emotions and mood as part of an effective decision making process. Simulating parts of emotions and mood can help to tackle these challenges in AI.

Agents are now being placed among the general public and in the real world [1, 18, 77]. Therefore there is a need to understand the differences that occur when moving from the theoretical understanding and basic network implementations of agents to simulated physical environments implementations. When agents are placed in the simulated physical environments the differences between the lab and the real world are often analysed on an ad-hoc basis, with little understanding of the fundamental differences that occur and what they mean for the agent.

The aim of thesis is therefore to advance and justify implementations of simulated emotion and mood into a computational decision making process. There is no universally accepted computational implementation of either emotions or mood, as discussed in Chapter 3. There is a need to analyse both computational implementations of emotions and mood and the psychological understanding of what emotions and mood are. Chapter 2 goes into detail highlighting how emotions and mood manifest themselves and how they affect decision making, in both positive and negative ways. Additionally the analysis highlights how there is a lack of mutual consensus on the definition of mood; in this thesis mood is interpreted as being synonymous with affect. The thesis explores how self-interested agents which use simulated emotions and mood are able to cooperate, and to study this the experimental work focuses on the Prisoner's Dilemma scenario which allows exploration of cooperation among agents in a society.

When considering simulated emotional and moody agents in a multiagent setting, there has been a strong tendency to simulate the agents in a networked setting [3, 72, 126, 82]. This leaves a gap in the literature of how these agents will react when placed in environments that allow the agents to move freely. Chapters 4 and 5 provide a number of experiments that show that mobility affects cooperation and which characteristics are successful for agents that use simulated emotions and mood. Chapter 6 aims to further expand on what mobility and environmental effects have on agents in general, by providing a number of different environments and a variety of different agent strategies.

The literature on simulated emotions and mood has a strong focus on experimental studies [1, 18, 48, 103, 87]. Focusing on purely experimental studies can lead to divergences between people who design the systems and people who focus more on the theoretical aspects of the problem. Chapter 7 aims to bridge this gap by providing an analysis of whether the chosen implementations of simulated emotions and mood are evolutionarily stable strategies. When all the chapters are combined, the answer to the main research question is given; to summarise, simulated emotional and moody agents can be competitive and cooperative in mobile societies.

1.3 Contribution of Knowledge

The aim of the thesis is to show whether self-interested agents that use simulated emotions and mood are able to be both cooperative and competitive when placed in societies that have the ability to move. The thesis will therefore contribute to the understanding of how to simulate emotions and mood in agent systems. The contribution includes an analysis of cooperation and average payoffs of emotional and moody agents when playing social dilemmas. In addition the agents will be tested in arenas that allow free movement of the agents, contributing to the understanding of agents in mobile societies. The contribution of the mobile arenas can be split into how mobility affects the agents and how the structure of the arena affects the agents, along with comparisons to the state of the art knowledge of these effects in networked interactions.

There are a number of different questions that need to be answered and as such the study will be split into distinct chapters which answer specific aspects of the research questions. Chapter 4 focuses on agents that use simulated emotions in mobile arenas. Chapter 5 focuses on simulated mood, with both definitions and psychological ground for the computational model of mood that will be used. Chapter 6 introduces networked interactions and focuses on the differences between networked and arena interactions and the structures of networks and arenas, for all agents. Finally Chapter 7 places emotional and moody agents contextually within the wider literature of self-interested agents, by providing an analysis of whether these particular implementations of emotional and moody agents are evolutionarily stable strategies when playing the Prisoner's Dilemma.

The thesis makes a contribution to the state-of-the-art in agent systems by:

- Experimentally showing that the addition of mobility in a society has an effect on both emotional and moody agents and agents in general.
- Highlighting through experimentation the effect that environmental structure has on agent interactions
- Demonstrating theoretically that emotional agents that use a practical implementation

of simulated emotions come to a mutual action in the Prisoner's Dilemma.

- Defining a general computational model of mood, which is grounded in psychology.
- Demonstrating the model of mood through multiple implementations.
- Evaluating the implementation of emotional and moody agents to show whether they are considered evolutionarily stable strategies.

1.4 Thesis Structure

The thesis is organised into a number of chapters, with a description of each chapter provided below, beginning with a literature review in Chapters 2 and 3. The interdisciplinary nature of the work requires an extensive literature review over several chapters. The original work is provided in Chapters 4, 5, 6, and 7 which uses the literature review chapters as the underlying knowledge of the work. The thesis then concludes in Chapter 8. The next section will give full details of the venues this work has already been published in.

Chapter 2 provides a literature review of the psychology literature to understand emotion and mood. This is in terms of how emotions and mood influence decision making in people, and how emotions and mood are defined by psychologists. Different models of emotions and mood are examined with consideration of the difficultly adapting them into computational models.

Chapter 3 continues the literature review, with this chapter focusing on how emotions and mood have been modelled in multi-agent systems, and how success of the agents has been measured. The chapter will start with a discussion on social dilemmas as an effective way of evaluating cooperation among agents, along with the background of social dilemmas. The different implementations of simulated emotions and mood in agents is also given. Furthermore, the chapter includes a review of previous work which shows what impacts mobility can have on agents. The chapter ends with a summary and the gaps in the literature that this thesis will address.

Chapter 4 starts with an introduction of the testing environment that is used throughout the thesis. Next is an experiment which aims to shows whether there are any impacts that mobility has on the emotional agents. The next part of the chapter will describe

a second experiment which explores in greater detail what the effects of mobility and environment structure are on the emotional agents' success, where success is measured in both societal cooperation and individual payoffs. The chapter will then conclude by presenting the results of both experiments.
The work in this chapter has been published in [22, 21, 24].

Chapter 5 will introduce the Mood model, firstly in the general case, and then followed by a full implementation including the psychology grounding used to justify the implementation. The experiment which uses the model of mood will explore the differences between the emotional agents and the moody agents. The scenarios include the mood model in self-play and the moody agents interacting against a number of different strategies. The chapter will also include a second implementation of the mood model to show how different interpretations of the generic mood model are possible. The second implementation includes a reinforcement learning process and is tested against a popular reinforcement learning algorithm. A summary of the chapter is then presented.
The work in this chapter has been published in [23, 25, 24].

Chapter 6 goes into greater depth on the effects that mobility and environment structure have on agents in general. The analysis is achieved through an additional experiment with extends the experiments of the previous chapters. The experiment has multiple arenas in which agents are placed, multiple network interactions, and the range of possible opponent strategies is increased. After the analysis, which focuses on the differences that mobility and environmental structure have, the chapter will summarise the effect the environmental structure and mobility have on agent interactions.
The work in this chapter has been published in [26].

Chapter 7 aims to provide an evaluation of both the emotional and moody agents in terms of evolutionary stability. An oracle agent will be introduced, which provides an effective opponent against the emotional agent. There will be a number of proofs which will show how emotional agents can be considered an evolutionarily stable strategy, with further proofs that moody agents are not an evolutionarily stable strategy. There will also be a discussion of the implications that the results yield.
This work has been published in [28, 27].

Chapter 8 summarises the thesis. The summary is presented in terms of how each chapter has contributed to answering both the sub research questions and the main research question that has been posed. A discussion is also given on the limitations of the experiments conducted and further questions that have arisen from the work. Finally a conclusion closes the thesis.

1.5 List of Publications

The following papers have been published that contain the research covered by this thesis:

- Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. Mobility effects on the evolution of co-operation in emotional robotic agents. In *Proceedings of ALA'16* (2016), pp. 114–121 [22] Introduces mobility to a simulated emotional agent and shows that different emotional characteristics become successful when mobility is introduced.
- Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. The effect of mobility and emotion on interactions in multi-agent systems. In *Proceedings of STAIRS'16* (2016), D. Pearce and H. S. Pinto, Eds., IOS Press, pp. 39–50. [21] Extends the previous work [22] above and additionally shows that the implementation of the simulated emotional agents converges to a mutual action.
- Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. Modelling mood in co-operative Emotional agents. In *Proceedings of Distributed Autonomous Robotic Systems* (2016), pp. 573–586. [23] Introduces the model of mood with the psychological grounding. An experiment shows that cooperation levels improve when compared to the equivalent emotional agent.
- Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. Environmental effects on simulated Emotional and Moody agents. *The Knowledge Engineering Review 32, e19* (2017), 1–24. [24] Analyses the simulated emotional and moody agents in a wide range of mobile environments, and includes a description of convergence to mutual actions for the simulated emotional agents.
- Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. Mood modelling within reinforcement learning. In *Proceedings of ECAL'17* (2017), MIT Press, pp. 106–113. [25] Implements the mood model within a reinforcement learning

context and shows improvement in cooperation when compared to the popular reinforcement learning algorithm SARSA in both the Prisoner's Dilemma and the Stag Hunt scenarios.

- Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. On the role of mobility and interaction topologies in social dilemmas. In *Proceedings of ALIFE'18* (2018), pp. 477–484. [26] Shows how mobility and different environment structures affect agents in general, over multiple arenas and social dilemmas. Concludes that mobility reduces the level of cooperation and unstructured environments, such as the empty arena, support defection, when compared to structured arenas such as the regular arena.

- Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. Stability of human-inspired agent societies. In *Proceedings of AAMAS'19* (2019), pp. 1889-1891. [28] Describes how the simulated emotional agents can be considered an evolutionarily stable strategy, while simulated moody agents cannot be considered an evolutionarily stable strategy.

- Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. Stability of cooperation in societies of Emotional and moody agents. In *Proceedings of ALIFE'19* (2019), pp. 467–474 [27] Sets out the formal proofs for the discussion of evolutionary stability in [28].

Chapter 2

Defining Emotions and Mood

The aim of this chapter is to survey the psychology literature detailing how emotions and mood are a fundamental part of intelligent decision making in humans, and also to motivate the modelling of these kinds of processes in computation systems for the benefit of the wider community of AI researchers. The chapter will also argue that a notion of fairness affects emotions and mood, and furthermore affects a person's decision making. The main aim of this chapter is to provide an extensive literature review of the current psychology literature regarding how emotions and mood manifest themselves in humans, and what measurable effects emotions and mood, have in terms of decision making. The chapter will begin by providing, in Section 2.1, an introduction to emotions and the current literature on their effects in decision making. Following on from this be an in-depth look at the two main models of emotions in psychology, the first being the Ortony, Clore, and Collins model of emotions, described in Section 2.1.1. Section 2.1.2 will describe the second model of emotions, the Circumplex model of Affect. A comparison of the two models is given.

Following on from this will be a description of mood and the current psychology literature detailing the effects of mood on decision making, in Section 2.2. The next section will show how different mood levels have different affects on decision making and the human psyche. Section 2.2 will, in addition, describe the notion of fairness and equitable outcomes and the impact these have on emotions and mood, and therefore decision making. Lastly a conclusion is given which summarises the argument that emotions, mood, and fairness are part of a human's decision making process and that they are interlinked.

2.1 Defining Emotion

Emotions in humans have been, and continue to be, extensively studied in the field of psychology. While people all have instinctive intuition of the nature of being angry, sad, happy, etc., the ability to define specifically the nature of the emotions and the effects they have on ourselves and on people's actions has proved to be extremely difficult. This section starts with looking at what effects emotions have on decision making and how these can be positive traits.

Keltner and Gross [63] provides the view that emotions are functional, that is, they are adaptations to serve a goal. Keltner and Gross go on to define emotions as short-term, episodic, and biological patterns of perception that are responses to environmental challenges and opportunities. The authors provide a brief history of previous views of what use emotions are to humans. While some believe that emotions have no function and are only a hindrance [63, 33, 54, 66], another view was that the functions of emotion are no longer appropriate to modern society [63, 47]. The most prevalent view among psychologists is that the functions of emotions remain important now since emotions are adaptions to the environment [63].

Given that emotions are important, this leads to the question of what specific effects emotions have on humans, and how emotions are triggered. Emotions occur due to changes in the environment, which helps provide rewards/punishments for the actions taken within the environment to achieve a goal [99]. Rolls [99] provides further insights into emotions and the brain, for both specific parts of the brain and the nature of emotions; this thesis will focus on the latter. The goal orientated nature of emotions allows flexibility in changing behaviours to improve outcomes and behaviours in the environment. The emotions then trigger changes covering changes in decision-making and changes in perception of the environment.

Levenson [67] provides an overview for categorising emotional functions and providing examples of the effects, and why the effects of these emotions are important and beneficial to humans. These functional emotions can be split into two different types, intrapersonal and interpersonal. The intrapersonal emotional functions include effects such as coordination of response systems, which affects facial expressions, voice, and endocrine (internal hormonal distribution such as adrenaline) responses. In terms of decision making, an important interpersonal effect that was shown was that emotions short circuit cognitive processing and that emotion works in conjunction with deliberation [67]. The short-circuiting tends to happen in situations where well-being and integrity are under threat and an action is needed immediately. For example, the commonly known *"fight or flight"* response emotions will short-circuit the more rational thought process as time taken for deliberation will negatively affect the person. When considering the second category, interpersonal functions, this can interlink with intrapersonal effects such as facial expressions allowing communication of emotions across different people [67]. By recognising emotions across individuals, the interpersonal emotional response is able to place things such as objects, people, ideas in context, which in turn affects how people learn.

The account discussed above shows emotions have positive effects in humans. Schwarz [108] gives us a deeper overview of how emotions in general affect decision making and cognition. The author looks at how emotion affects decision making at different points in time, including when the decision was made and after the action has been taken, including how emotions are reflected in memory of previous choices and how anticipation of emotion is a part of decision making. For example, if you are presented with a choice of two different coffees when feeling negative, you choose brand A. When it comes to making the choice a second time you are reminded of the negative emotions you felt when choose brand A, and so are more likely to choose the other brand [108].

Emotions affect the decision making process while the decision is being made. When an individual is in a positive mood, then the decision they are deliberating upon will be viewed more positively [108], if they were in a negative mood, the decision will be viewed negatively. This effect has far reaching consequences; to take for example, punitive governmental social policies, these are more preferred by angry people, while more reparative governmental social policies are preferred by happier people [108]. Once the decision has been made and the action taken, there is an emotional response, which can include regret, relief, etc. Therefore emotions are bi-directional, meaning that decisions and their actions affect the emotional makeup of a person, and in turn the emotional makeup of a person will affect their decision making process. In addition to the positive and negative associations of emotional decision making, the action taken will also consider how to avoid negative emotions such as disappointment and try to enhance positive emotions [108]. Memory of what emotions were felt when the action was taken previously will affect what action will be chosen by the current decision making process, however this only takes into account the strongest emotion at that time, and the emotion that resulted.

The literature shows that emotions are important to humans and that emotions affect the decision making process. The next part of this section will explore in more depth the effects emotion has on learning and memory. Through study on the human amygdala, which are two almond shaped parts of the brain in the temporal lobe, and are associated with emotions, Phelps explores the link between emotion, memory, and learning [90]. Phelps shows that emotional learning, where an entity becomes associated with an emotion, exists and that direct experience is not required for the emotion to become associated. For example, fear could become associated with dogs, if that person was bitten by a dog, but in addition the person could show the same fear response if they are told about a dog that bites [90]. The author shows that the ability to remember an event is affected by emotions, with more intense emotions allowing the memory to become more vivid and easier to recall. The vividness of the emotion increases along with the confidence that the memory is correct, even when the memory is incorrect. The first part of a person processing the world around them is what to pay attention to and how the world is perceived. Emotions affect the perception of the world by processing negative parts of the environment so that fearful or threatening stimulus gets processed first [90]. Processing negative emotions first is reflected in the processing human faces, with negative facial expressions being processed first. From the literature, emotions have been shown to be goal-orientated and are affected by changes in environment that reinforce behaviours of actions. The effects of emotion are both physical and psychological, such as the racing of the heart in a fearful scenario and the increase in cooperation when grateful to another [31]. Furthermore, emotions have effects on decision making, learning, and memory, making emotions a system which influences itself.

Now that the nature of emotions has been expressed, the following sections will look at two models of emotions, the Ortony, Clore, and Collins (OCC) model of emotions [89], and the Circumplex model of affect [91]. The purpose of looking at these models is to see how emotions can be represented, and the specifics of which particular emotion should affect a particular action, along with how individual emotions should be affected by the environment. Given that the aim of this thesis is to analyse simulating emotions in an agent system, there will be a very brief mention of the relative difficultly in representing these models in computational systems.

2.1.1 Ortony, Clore, and Collins Model of Emotions

The OCC model of emotions [89] is a hierarchical model, as shown in Figure 2.1, with twenty-two emotions defined as a valenced reaction to some specific entity. This entity can

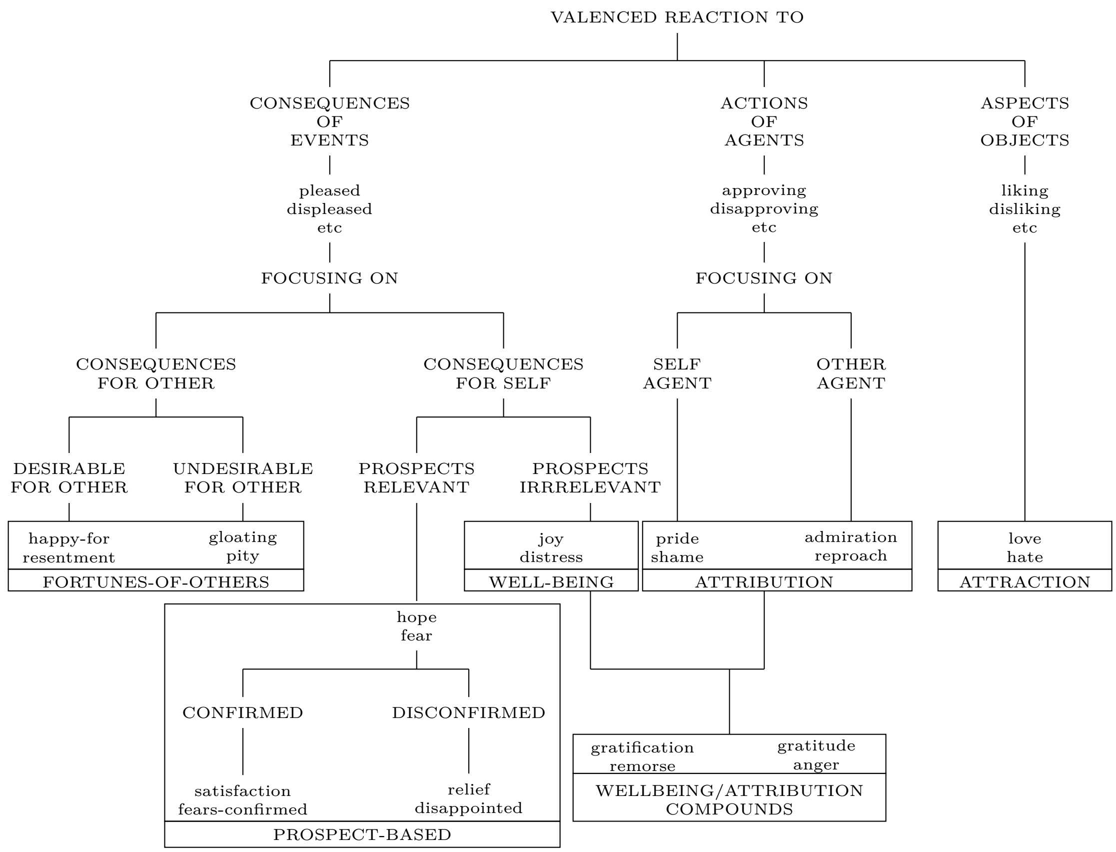


Figure 2.1: Hierarchical representation of the Ortony, Clore, and Collins model of emotions [89].

be defined as an event, or an agent/person, or an object. There is flexibility in how a real world object can be perceived, changing the category of that entity. To give an example of how an object can be perceived as different entities, consider how a car can be perceived as either an object, or an agent. The car will be represented as an object when a person is evaluating what kind of car to buy when looking around a car showroom. An example of the car being treated as if it were an agent rather than an object; You are running late for work, you jump in your car and it is failing to start. After multiple attempts of trying to start the car, it finally bursts into life. The person starting the car has an outpouring of gratitude towards the car, as if it were an agent.

Using Figure 2.1 helps explain the validity of the example. Starting from the top of the figure, the car starting produces a valenced reaction. The car is an agent, and this is an action that is being approved of, and the car is an "other agent". The reaction doesn't stop at the attribution section as the example concerns the well-being of the car, eliciting the gratitude emotion.

The valenced reaction becomes an emotion when the action causes the *potential* value of the reaction to reach the emotional threshold, which will then change a person's behaviour. The *intensity* of the emotion is in relation to how far past the threshold the *potential* has gone. The OCC model describes how the potential of each emotion is evaluated. As an example, the Joy emotion requires that the event is desirable for the person or agent. Equation 2.1 describes how Joy is elicited, by the authors [89].

$$
\begin{aligned}
&IF(desire(person, event, t) > 0)\\
&\{\\
&\quad joyPotential(person, event, t) =\\
&\qquad f_j(desire(person, event, t) + globalIntensity(person, event, t))\\
&\}
\end{aligned}
\tag{2.1}
$$

The Joy potential for a given person/agent during an event at time t is the result of the Joy function f_j which takes the desirability the person finds the event at t which is then increased by how intense the person finds the event at t. The resulting action that would be taken would depend on which threshold the Joy potential has reached and how far past the threshold the potential has gone; this would give the intensity.

The authors of the OCC model state that the reason that emotions are elicited is to help achieve a person's goals. Using the example of the car starting, the gratitude is elicited because there is a current goal of moving elsewhere using the car. The OCC model categorises goals into three different categories, which are:

Active-Pursuit Goals that a person or agent will actively put effort into achieving. Examples include: "get to point B", "writing a book", etc.

Interest Goals that a person or agent has an interest in achieving as they will benefit from the goal being achieved. Examples include: "sports team wins", "keeping fit", "winning the lottery", etc.

Replenishment Goals of a person or agent that become more heavily prioritised as the resource for the goal becomes increasingly depleted. Examples include: "staying hydrated", "keeping phone charged", etc.

In addition to emotions affecting the elicitation of goals, the model also includes standards and attitudes. These are included to allow the necessary differences in people's values, likes and dislikes, and opinions. Standards represent a person's values. Attitudes represent people's opinions and individual tastes.

The goals that the OCC model creates can be viewed through the lens of multi-agent systems, with goals that need to be worked towards, and resources that need to be maintained. In addition, the flexibility of representing the values of the agent, and the thresholds at which the emotions become elicited allow an agent designer to create a computational model. The computational model can be modelled by the designer to produce output in a deterministic way, allowing predictability in action selection. Additionally being able to justify the choices made to the computational model using the OCC model, gives the designer sufficient psychological grounding to call aspects of the model simulated emotions.

2.1.2 Circumplex Model of Affect

Russell provides a different model of emotions, termed the Circumplex model of affect [101]. The model uses a dimensional theory of emotions, that an emotion is a linear product of valence and arousal. Valence is the dimension that represents how pleasant the emotion is to the person, and arousal is the dimension that describes how active the emotion is. The result of these two dimensions is then interpreted as an emotion. For example, joy is the

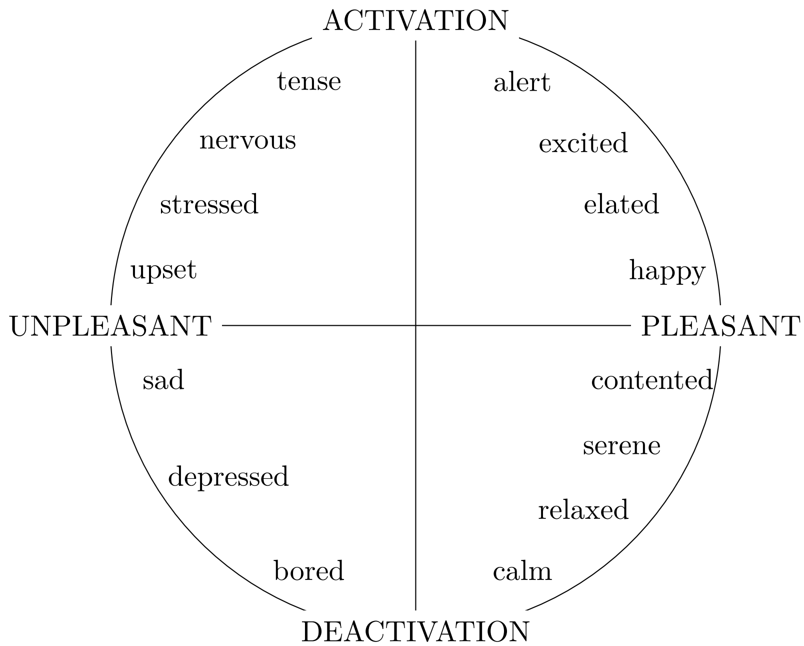


Figure 2.2: Circumplex model of affect [91].

emotion that has a strong positive valence and has a moderate activation of arousal. A representation of the Circumplex model is shown in Figure 2.2.

The model was achieved by taking twenty-eight different emotional words and creating an experiment where people place where they thought each emotional word should be placed [91]. The authors noted that the edges of each boundary were fuzzy as there is an overlap between different words but there is a consistent circular ordering.

Posner et al. analysed this model when theorising how to move from the model of basic emotions to the Circumplex model of affect, and the implications this would have in psychology [91]. The authors also point to the variability of cognitive schemas that people have and how this affects the model, with emotions overlapping in the model due to differences in valence that people naturally have, and natural changes over time. Examples of valence differences are given in Figure 2.3.

When agent designers are looking at how to use the Circumplex model of affect in a computational model, there are a number of issues that make the model harder to implement in comparison to the OCC model. The model does not include how individual emotions and goals are integrated together and the effects on the outside world, which makes validating

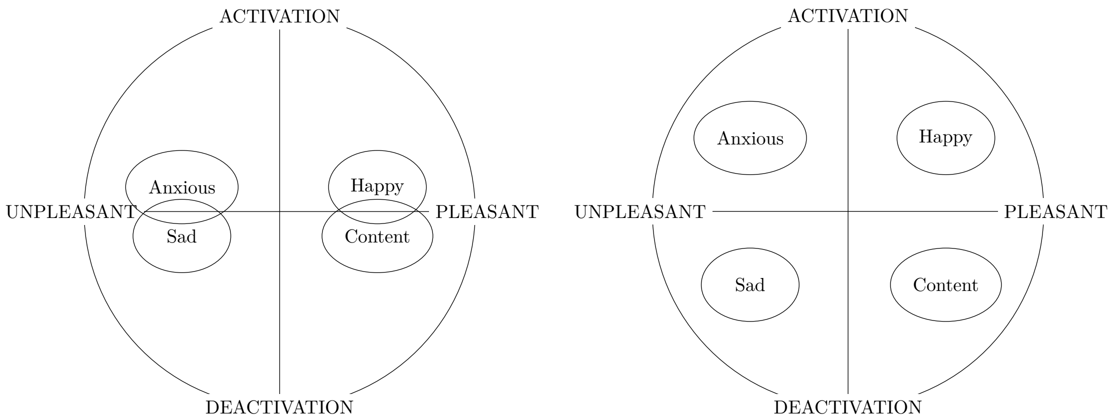


Figure 2.3: Valence difference example in the Circumplex model of affect [91].

a particular implementation extremely difficult and would necessitate other psychological research. The possible overlapping emotions as well as the fuzziness of these emotional definitions would provide further complications in implementation since determining which emotion should be elicited becomes non-deterministic.

2.2 Mood

The chapter up to this point has reviewed emotions and how they affect multiple parts of the human psyche. Furthermore, the chapter has focused on two different psychological emotional models and given some initial insight into how they can be implemented in computational models along with any difficulties in doing so. This section argues that in addition to emotions, mood is a necessary concept to be examined as a person's mood interlinks with emotions and affects their decision-making in a similar way by affecting learning, memory, and judgement.

When showing that mood affects people's decision making as well as other parts of the human psyche, the following sections will first show how the literature defines mood and how this differs from emotion which the psychology literature has taken into consideration. Studying the psychology literature of mood, in order to define what exactly mood is, introduces a larger issue. Psychologists do not have a fixed definition of mood, where the definition of mood is dependant on a particular author's interpretation. Watson and Tellegen [124] highlight this difficulty and offer their particular interpretation of mood, as positive and negative affect.

Firstly looking at Russell's work on the dimensions of affect [100], where people rated emotional words in terms of pleasantness, arousal, and dominance, the results show that there are at least two dimensions within these emotions. These were the *pleasantness* and *arousal* dimensions. The author also suggests that there may be more than two dimensions, and that these further dimensions are descriptions of how future events are judged and who controls the event.

Diener et al. [34] provides further detail, showing that the positive and negative affect are inversely correlated, and that the arousal[1] dimension explains how positive and negative affect can appear to be independent of each other. The authors make the reference to mood in their experimental details, describing how it has positive or negative affect over a significant period of time, in their particular case, a full day [34].

While mood has not been mentioned in terms that are intuitive, psychologists often lack a consensus between intuitive names that are easily recognizable and more domain specific terms. Zevon and Tellegen [127] help bridge this gap through their studies, by defining mood as positive or negative affect which in addition support multiple dimensions as a structure. They also provide further evidence of mood affecting aspects of personality such as anxiety and depression. The study gave credence to the intuitive understanding of mood as positive and negative affect, that it is bipolar and anchored at both poles [127]. They effectively describe mood as having three aspects: positive, negative, and neutral, with neutral being that I am not in a good mood, but that does not necessarily mean I am in a bad mood [127].

Furthermore Watson's book *Mood and Temperament* [122] provides the most explicit case of mood being described as positive and negative affect. This book also shows how mood and emotions are interlinked, and that they affect a large range of aspects in the human psyche. In addition to Watson's book, Hepburn and Eysenck [56] describe how the two dimensional mood description as positive and negative affect is an effective description of mood. Their experiments provide empirical evidence of this, as the description was able to predict that positive affect relates to extroversion and negative affect predicts neuroticism. Hepburn and Eysenck also showed how someone's personality affects the nature of their mood, with stable introverts having less mood variability, and neurotic extroverts showing the largest mood variability.

[1] Arousal is referred to as *intensity* in [34]

Now that the gap between the intuitive understanding of mood and the domain specific terms used in the psychology literature has been bridged for this thesis, the chapter can now analyse how positive and negative affect impact upon multiple parts of the human psyche, in a similar manner to the section on emotions.

The previous section on emotions reviewed the literature that shows that emotions affect memory and learning. Bower provides evidence [16] that mood in addition to emotion affects people's memory. People remember events more reliably when their current mood is similar to the mood the people were in at the time the event happened. Additionally people attach more importance and therefore remember events better, when the event matches their current mood. More evidence of this effect is provided by Mayer and Hanson [79], who in addition showed that when a person's mood changes this then in turn affects their judgement of the event.

There is further support for mood affecting memory and learning by Rinck et al. [98] who showed that mood-congruent learning occurs when the emotionally-toned learned words were strong, whereas mood-incongruent learning occurred when the events were only slightly emotionally-toned. The results show that emotions and mood both have an effect on learning. Rusting [102] expands on this by showing that a person's personality affects their mood, which then in turn affects their emotional processing of the event. Further highlighting how mood, emotion, learning, and memory are interlinked. Baxter et al. [9] show that memory is a fundamental part of learning, which interacts with cognition. Eldar and Niv [37], have also shown the effects that mood has on people's perception of rewards and the effects that it has on the brain, and that these effects were amplified when the person was in an unstable mood.

Now that this section has reviewed the studies that showed that mood affects memory and learning, it is worth noting that Baumeister et al. [8] shows us that "bad" results, such as negative emotions, negative affect, and negative events have a stronger effect on humans than positive aspects. The negative results are therefore more memorable and have a greater effect on people's psyche.

The general effects that mood has on memory and learning have been described, and it has been shown that mood can be made of positive, negative, and neutral moods. Aspinwall [4] provides an extensive review of the effects of mood, and the differences between positive and negative moods. In the general sense, positive moods allow people to be more creative, open minded, and willing to cooperate with others. While intuitively people think that being in a positive mood is always beneficial, when in a positive mood processing information actually becomes more difficult. The difficulty in processing information in a positive mood only gets worse as the information becomes more negative and not in line with the current mood. The end result is that people in a positive mood take riskier actions [4]. When considering negative mood, the processing of information becomes more considered, especially when that information mimics the mood (both the mood and the information are negative to the person). Aspinwall also gives evidence that the mood serves as an input to the action chosen, which affects the mood, leading to a circular dependency for mood [4].

The different approaches that mood has on decision making and information processing was further shown by Hertel et al. [57], with decisions being made more quickly when people were in a positive mood, highlighting the use of heuristic processing. Consequently when in a negative mood, decisions took longer, highlighting a more systematic approach. In the context of social dilemmas, where working together for the greater good, rather than personal gain is the aim[2], the results [57] showed that mood does influence decision making, but positive mood does not necessarily mean a higher level of cooperation.

Lount [74] gives further evidence of the occurrence of heuristic decision making when in a positive mood, through experiments in trust. While intuition may say that being in a positive mood would make you more trustful of everybody, the experiments showed that positive mood increased trust only for those that are part of the “ingroup” (Social group that an individual identifies with.). Trust for people that were considered to be part of the “outgroup” (Social group that an individual does not identify with.) was lower, specially when they exhibited untrustworthiness cues [74]. For people in a neutral mood, they take a more considered approach by considering what actions the group previously took, rather than just considering the out-group to be distrustful [74].

Expanding on social dilemmas and how these are used to measure mood and its affects, Haley and Strickland [53] performed an experiment where depressed women interacted in a social dilemma to see what effects extremely low moods had on the outcomes. The people were shown to be more self-critical of their performance in a subsequence task. The people in the experiment were described as being more aggressive with their action choices in the social dilemma. In this particular instance the social dilemma was the Prisoner’s Dilemma and the aggressive action was defection. Defection is one of the two action choices, it can be considered aggressive as it aims to take advantage of the opponent. Defection can also be considered the more rational choice as defection is the best choice of action if both players

[2]Further discussion on social dilemmas is given in Chapter 3.

are playing to their full ability, also known as a Nash Equilibrium[3].

Further evidence of the increased critical evaluation and feelings of betrayal when depressed in the Prisoner's Dilemma is given by Gradin et al. [52]. The people playing the game in these experiments also showed decreased satisfaction with their payoffs in the current game, as well as increased feelings of guilt when they did take advantage of their opponent [52, 10]. The experiment also showed that between depressed people and the control group, their reaction time did not differ significantly [52].

There has been a mention of depression, which is characterised as an extremely low mood, and it was shown that these extremely low moods reflect the descriptions of the low moods, but taken to the extreme. Malhi et al. [76] gives an extensive review of the different neurocognitive models that underpin mood disorders, as well as giving intuitive explanations of mood disorders. Firstly, depression and secondly, mania which is characterised as an extremely high/positive mood [76], and finally bipolar disorder which is when someone's mood moves between positive and negative moods quickly [76].

For mania it is not immediately intuitive what the negatives of a constantly very positive mood would be; Leahy [65] gives insight into these negatives. As stated before, mania is a very high/positive mood, therefore the people with mania will rely more heavily on heuristic decision making. In turn this affects their ability to effectively assess the risks of their actions, leading to behaviours such as excessive impulsiveness, and aggression [65]. One of the difficulties in treating mania lies in the fact that the people experiencing mania may enjoy the heightened mood levels [65], and as people try to keep themselves in a good mood, mania may not be recognised as a problem by them.

Now that each different level of mood has been reviewed, the following sections will now look at how mood levels change. Mood levels will change slowly, as when mood levels change quickly this is defined as mood instability [19]. Broome et al. [19] looks at how mood instability is relatively common and how the precise definition is difficult and lacks consensus, with the two defining questions being:

1. Do you have a lot of sudden mood changes?

2. Have you suffered this symptom over the last several years?

These questions show that while the exact definition is difficult to define, Broome et al. [19], Bilderbeck et al. [11], and Bonsall et al. [15] agree that having the mood levels change

[3]As noted previously, an in-depth look at social dilemmas, such as the Prisoner's Dilemma is given in Chapter 3

quickly is a negative trait for people. Bilderbeck et al. show that emotional processing biases in patients with bipolar disorder are affected by mood instability [11]. Bonsall et al.'s experiments [15] show how both mood levels that are unstable over large periods of time, as well as large episodic jumps, are negative as they produce high rates of suicide. From the literature we can infer that a healthy mood over time changes slowly, as Bipolar disorder is mood changing quickly over time and is considered an unhealthy trait [11, 15, 19].

To summarise, the section on mood has reviewed the literature that shows that mood is characterised as positive and negative, is interlinked with emotions but is independent [20]. Mood does not have a particular focus as it affects emotional processing, which affects everything including such things as food choices [50]. Personality affects a persons possible range of mood, and that mood changes slowly over time, with fast changes being extremely negative for a person [11, 19, 15].

In addition this section has reviewed the literature that shows that a person's personality affects their mood, and their decision making. A part of someone's personality that will have an effect on the mood is how people are not perfectly self-interested agents as people want equitable outcomes [45]. However when a person is achieving a good outcome in a society, their willingness to give up some of their reward to reduce inequity is lower than if they were doing poorly in society. There are differences between individuals relating to their backgrounds that have an effect on how much they value efficiency over equity, or equity over efficiency, along with the type of game chosen [44]. To prevent cooperation and trust breaking down between individuals, due to a minority of selfish players, people will punish selfish players, even if the punisher is not affected by the actions of the selfish players. This is termed *strong reciprocity* [43]. To summarise, a notion of fairness and equity is an intrinsic part of a person's personality. An individual's interpretation of fairness and equity will affect how they determine whether an outcome is positive or negative, which will in turn affect their mood and decision making.

The previous sections have looked at how mood is defined and the effects that mood has on decision making, in terms of both positive and negative moods. The following paragraphs will now look at two models which psychologists have used to measure mood. These are Watson et al.'s Positive and Negative Affects Schedule (PANAS) [123], and Mehrabian's Pleasure-Displeasure, Arousal-Non-arousal, and Dominance-Submissiveness (PAD) Temperament Model [81].

The PANAS scale [123] allows scientists to measure someone's mood, through rating the applicability of 20 words which are plotted on a 5 point scale. The 20 words are given

Table 2.1: PANAS positive and negative words.

Positive affect words	**Negative affect words**
Attentive	Hostile
Active	Irritable
Alert	Ashamed
Excited	Guilty
Enthusiastic	Distressed
Determined	Upset
Inspired	Scared
Proud	Afraid
Interested	Jittery
Strong	Nervous

in Table 2.1, along whether they are considered positive or negative. Once all the words are rated by an individual, you get a positive affect rating and a separate negative affect rating which ranges as a value between 10 and 50. The PANAS scale is effective over different timescales depending on the measurement that needs to be taken. This allows accurate comparisons of mood levels over hours or days, with the PANAS scale also being effective when measured over weeks and months.

While the PANAS scale is an effective tool to use when analysing the mood of people over a given time frame, the ability to use the PANAS scale as an effective part of a computational decision making process is not trivial. The scale does not show how different levels of moods will affect the decision making process, nor how actions or the environment will affect how the different words are rated. When considering how to implement this, the implementation details are completely open to almost any interpretation, making the justification for any particular implementation rely heavily on other psychology literature.

The PAD Temperament Model [80] is a 3 dimensional model which includes **P**leasure, **A**rousal, and **D**ominance. The PA part of the PAD model mirrors that of the Circumplex model of affect. The extra dimension of dominance measures whether an emotion is dominant, for example Anger is an emotion which is dominant while fear is the same emotion (Low pleasure, High arousal) but submissive in nature. Mehrabian [81] also found that the PAD model allowed effective analysis of personality traits, such as emotional empathy. Additionally multiple temperaments were defined and how this interacted with the OCEAN[4] (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) personality model [35].

When looking at the PAD Temperament model in terms of simulation models, there are strong similarities to the Circumplex model of affect. The addition of a third dimension does not change the difficulty in integrating the model into a computational model, as the same issues arise as from the Circumplex model of affect.

To conclude, this section has shown that mood can be characterised as positive and negative. Mood influences emotions and decision making, and in turn personality, emotions, and the environment affects mood. In addition it was shown that positive moods allow decision making to be more heuristic, while negative moods make people more considered in making their decision. Mood affects a persons learning behaviours as people are more readily able to recall events that match that person's current mood, as well as events that matched the mood people were in. Mood does not change quickly, and mood instability negatively affects people's lives. As personality affects mood, a notion of fairness will in turn affect their mood, by colouring the perception of an outcome.

2.3 Summary

The aim of this chapter was to show that both emotions and mood are fundamental aspects that people use as part of their decision-making process. This chapter has reviewed the literature that describes emotions as short-term feelings towards a specific entity. When considering how to model emotions, two different models of emotions were reviewed, the OCC model and the Circumplex model of affect. These models take different approaches to modelling emotions, with the OCC model taking a more goal-orientated view and considers what outward effects individual emotions have on actions, while the Circumplex model of affect takes a more inward looking approach with emotions being a descriptor of a two dimensional model of valence and affect.

In looking at what effects mood has on the human-psyche, there is a notable difficulty in defining mood, with psychologists often using domain specific knowledge that may conflict with a broader intuitive understanding of mood. Through looking at the literature on mood, this thesis will define mood as being synonymous with affect. While there is no explicit model of mood, the chapter reviewed two models which incorporate mood, the

[4] Also known as the big 5 personality traits.

PANAS scale and PAD Temperament Model. The PANAS scale is useful for measuring what a person's current mood is but does not provide the internal structure which allows us to predict what the mood will be after a given input. The PAD Temperament model is effective in describing how personality traits can be attributed to different temperaments.

This chapter has highlighted that while there are emotional models that can be implemented in computational models, they often do not include mood. There is also a need for a mood model that takes into account the outward effects of mood, to allow a computational model of mood that can capture the essence of mood in a psychologically-grounded way.

Chapter 3

Computational Accounts of Emotions

The prior literature will now be placed in context of computational models of emotion and mood. The aim of this chapter is to provide a full overview of the models of emotional and mood computation that exist in the literature. The study of the literature establishes the necessary background to provide a starting point to answering the main research question set out in Chapter 1. In addition this chapter will also provide the background to two social dilemmas and how effects external to the agent can affect that agent's decision making. The chapter will begin with an extensive overview of computational models of emotion and mood in Sections 3.4 and 3.5 respectively. I consider emotional/moody computation models in this thesis to be an adaptation of a psychology model that is represented as a computable function. Following in Section 3.1, will be an introduction to social dilemmas, namely the Prisoner's Dilemma and the Stag Hunt, and how they are used in order to analyse cooperation through human-inspired models. Section 3.6 will then provide the background to what kind of effects the environment can have on decision-making agents. Finally Section 3.7 will provide the conclusion of the chapter and explain how Chapters 2 and 3 lead into the experimental work.

3.1 Social Dilemmas

Now that we have shown that emotions and mood have and continue to be modelled in a computational setting, this section will focus on how these models can be useful to computer scientists and the problems they can assist in solving. Looking at the computational models and Chapter 2, emotions and mood have often been studied under the context of social dilemmas. Social dilemmas are situations with two or more people/agents, where there is a temptation to be selfish, however if everyone is selfish then the whole group of people/agents loses out. This section will be looking at social dilemmas from an agent perspective, other perspectives have be used such as using argumentation frameworks [10]. The agent perspective has been chosen as this thesis will be analysing mobility and environmental structure, which the agent perspective naturally aligns itself with.

3.1.1 Prisoner's Dilemma

One of the most common social dilemmas is the Prisoner's Dilemma, popularised through the influential Axelrod's tournament [5], where two players pick between cooperating with the other player, or trying to take advantage. The Prisoner's Dilemma gets its namesake from the scenario it presents. Two people are taken into custody and interrogated separately, the officers only have enough evidence to charge both of them with a minor crime. They are both offered a deal whereby if they testify against the other then they are able to go free, while the other gets charged with a major crime. If they both testify they both go to prison for a major crime. If only one testifies then the testifier goes free while the other goes to prison for an even longer time than if they both testified. If neither testify then they both go to prison for the minor crime.

In an agent setting the action choices are described as cooperation (Do not testify) and defection (Testify). Both agents choose an action simultaneously, independent of one another, and with no prior communications. If both agents choose to cooperate this is the best payoff for both agents. If one player chooses to defect against the other then they receive the best individual payoff, and the cooperator receives the worst individual payoff. If both agents choose to defect they both receive the worst payoff.

The payoff matrix for the Prisoner's Dilemma is shown in Table 3.1, example values are also provided. The payoffs in a Prisoner's Dilemma game must have the following restrictions hold [97] to be a valid game:

- (T)emptation $>$ (R)eward $>$ (P)unishment $>$ (S)ucker
- $R > (S + T)/2$

Axelrod's Tournament [5] is the seminal work for analysing the iterated Prisoner's

Table 3.1: Payoff matrix of the Prisoner's Dilemma, R, P, T, S represent the payoffs which must be of the form $T > R > P > S$ and $R > (S + T)/2$. Example payoff values are given.

$Cooperate, Cooperate$	$Defect, Defect$	$Defect, Cooperate$	$Cooperate, Defect$
R, R	P, P	T, S	S, T
$3, 3$	$1, 1$	$5, 0$	$0, 5$

Dilemma. People were able to submit strategies to the tournament where they would all compete against each, and they were allowed to remember their past encounters. The most successful strategy was the Tit-For-Tat strategy, which initially cooperates and copies the opponents' previous move [5]. Axelrod gives a number of conditions that need to be met into order for a strategy to be successful, which are [5]:

Nice The strategy should be optimistic by being initially cooperative. The strategy should not be the first to defect.

Retaliating The strategy should not be overly optimistic. If an opponent defects, defect back at the opponent.

Forgiving If the opponent starts cooperating after defection, the strategy should start cooperating with the opponent.

Non-envious The strategy should not aim to score more than the opponent.

Win-Stay Lose-Shift[1] [86] is another Prisoner's Dilemma strategy developed after Axelrod's Tournament. The strategy changes the current action only if the outcome of the previous action was the worst-case scenario for that action. Win-Stay Lose-Shift has been shown to improve on Tit-For-Tat, and Generous-Tit-For-Tat[2] in terms of sustained cooperation as it can move back to cooperation from a mutual defection [86]. Hilbe et al. [58] provides an extensive consideration of all strategies with a memory of one previous action such as Tit-For-Tat and Win-Stay Lose-Shift. The authors classified the strategies into 3 different categories and found the payoffs were relative to their grouping. The differences in payoff between the groups were unaffected by the opponent strategy [58]. The categories and their relationship to the payoff they receive are:

[1]Win-Stay Lose-Shift is also referred to as Pavlov
[2]GTFT requires two defects before changing to defection

Table 3.2: Payoff matrix of the Stag Hunt, R, P, T, S represent the payoffs which must be of the form $R > T \geq P > S$. Example payoff values are given on the next row.

$Cooperate, Cooperate$	$Defect, Defect$	$Defect, Cooperate$	$Cooperate, Defect$
R, R	P, P	T, S	S, T
$3, 3$	$1, 1$	$2, 0$	$0, 2$

Partner Strategies are strategies which are initially cooperative, but will switch to defection. The payoff for the society of agents will be maximised; this is the social optimum. The society will not receive this social optimum if there are opponents that prefer an unfair strategy. Both strategies will lower their average payoff as they switch to mutual defection.

Competivitive Strategies are strategies that are initially defecting.

Zero-Determinant Strategies are given a set of probabilities of taking an action based on the previous outcome, however they need to know the opponent strategy to come up with effective probabilities.

The authors of [58] concluded that competitive strategies never obtain less than the opponent, while a player using the zero-determinant strategies forces a linear relation between its own payoff and the opponent's. The authors also showed that these properties held even if the opponent was not using a memory-one strategy.

As documented in the previous sections, there have been a number of different strategies that can achieve good outcomes for an agent in the iterated Prisoner's Dilemma. Another focus in creating effective strategies is to create an adaptable strategy by optimising your actions against specific opponent strategies, with the literature having a heavy focus on using machine learning as a strategy to adapt to the opponent.

Another example of a social dilemma is the Stag Hunt [110], where a group of hunters have tracked a stag and if they all work together then they successfully hunt the stag and eat well. If not everyone hunts the stag then they go hungry, however a hunter can guarantee that they themselves will not go hungry if they hunt rabbits rather than the stag. Table 3.2 has an example payoff matrix, where hunting the stag is cooperation and defection is hunting the rabbit. This differs from the Prisoner's Dilemma by having the temptation payoff for an individual be lower than the mutual cooperation payoff for an individual, i.e. $R > T$. This creates two Nash equilibria [84] for the Stag Hunt of mutual cooperation and mutual defection. The literature on the Stag Hunt is not as extensive, however the Stag Hunt is often combined with the Prisoner's Dilemma as a means of comparison [29, 116, 112].

3.2 Reinforcement Learning

Reinforcement learning [115] prescribes how an agent can learn to optimise her behaviour by repeated trial-and-error interaction with the environment. At each time step, the agent takes an action based on the current state of the environment, and observes its effect in terms of a reward signal and resulting state change. Behaviours that yield high rewards will be reinforced, whereas behaviours that cause low rewards or penalties will be reduced. The goal of the learning agent is to maximise her expected future rewards.

One of the starting points in using Reinforcement Learning in a social dilemma situation was to see if different types of Reinforcement Learning are able to predict what strategies were used by the agents [40]. Erev and Roth [40] showed that even very basic learning models were able to be an effective predictor of future moves, in a variety of games and strategies.

One of the most well-known reinforcement learning algorithms is SARSA. SARSA learns state-action values, $Q(s_t, a_t)$, which represent the expected sum of (discounted) future rewards after taking action a in state s at time step t. Definition 1 gives the Q update function for SARSA given the immediate reward r_{t+1} and the expected future rewards, which are estimated recursively by the value of the next state-action pair $Q(s_{t+1}, a_{t+1})$ discounted by γ.

Definition 1 (SARSA [115].)**.** Let St be the set of states with $s \in St$. Let A be the set of actions with $a \in A$. Let t be the time, r represent the reward, α the learning step size and γ the discount factor of future rewards. Then, SARSA updates $Q(s_t, a_t)$ using the following equation:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] \tag{3.1}$$

When applying SARSA to a social dilemma setting, commonly the payoff is given as a reward r and cooperation or defection is given as action a. The variations are commonly in how the state is represented, often the state will include the agent and some of their previous actions. For example, Vassiliades et al. [118] represented the state as a combination of agent and previous action.

Given SARSA's success, modifications have been made to the learning to increase the speed of learning. Reward shaping supplies an additional reward to the state update step that would normally be received for a particular action [85]. For the SARSA algorithm the update rule will be altered to the following.

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + F(s_t, a_t, s_{t+1}) + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] \qquad (3.2)$$

where F returns a separate reward based on the state action pair and the next state. The reward is defined by the designer of the system. This technique can allow reinforcement learning to scale up to more complex problems [32, 92].

With agents that use reinforcement learning as their strategy, Crandell and Goodrich [29] note that many of these strategies are myopic as they are reactive. For a strategy to be considered effective it needs to meet two properties [29]:

Security The strategy never produces an expected payoff below the minimax value of the game.

Mutual Cooperation The agents should work together for the benefit of the society of agents, sacrificing their own individual payoff if necessary.

M_Qubed [29] was an adaption of Q-Learning [121] that showed the first property, and demonstrated the second property when all the agents used this strategy. In essence M_Qubed behaves predictably when the game is not competitive and unpredictably when the game is competitive. However as the strategy aims for cooperation, it is easy to take advantage of it, before it can adapt to the selfish agents.

The next problem Masuda and Ohtsuki [78] aimed to solve was to see if an agent could learn the different strategies that are used in a society. Once the agent has learned the strategies, the strategy should take advantage of naïve opponents and protect its own payoff. The authors achieved this strategy through temporal difference learning [78], however while the strategy was successful, the time needed to adapt was large.

By introducing multiple different strategies, there are also multiple agents. This introduces the ability to learn a strategy through multiple agents, as multiple agents can observe the actions and payoffs of a variety of different opponents. Vassiliades et al.

[118] applied multiple different implementations of learning algorithms, which have been categorised between agents that implement spiking [75] and those that don't. One aspect that was mentioned briefly by Vassiliades et al. [118] was how different strategies performed with different valid payoff matrices, in this particular instance having negative values for the punishment (P) payoff and the sucker payoff (S) was beneficial to learning.

3.3 Evolutionary Stable Strategies

There are many different strategies that utilise not only a single agent's perspective but many perspectives, requiring the ability to compare a large range of different agents. An effective method is to determine whether a given strategy can be considered evolutionarily stable.

A strategy can be described as evolutionarily stable when given that the majority of the agents in a society are using a particular strategy, that strategy cannot be invaded by any other invasion strategy that is initially rare [111].

Definition 2 (Evolutionarily Stable Strategies [111])**.** Let $V(Sy, Sy')$ be the expected payoff strategy Sy receives when playing against strategy Sy'. A strategy M is considered evolutionarily stable where M is the majority and the following holds for all invasion strategies Iv where $M \neq Iv$.

$$V(M, M) > V(Iv, M) \text{ OR } (V(M, M) = V(Iv, M) \text{ AND } V(M, Iv) > V(Iv, Iv)) \quad (3.3)$$

The stability argument has been used in biology much in the same way as game theory concepts have been applied to biological phenomenon [117, 30]. Evolutionary stability in the Prisoner's Dilemma has been extensively analysed from an agent perceptive [14], with no pure strategy being evolutionarily stable in the iterated version of the Prisoner's Dilemma [17]. Furthermore no Tit-For-n-Tats is an evolutionarily stable strategy[42], and neither are any reactive strategies [73]. The predictability of these kinds of strategies allows invasion strategies to be successful. This shows that evolutionary stability is an extremely demanding criterion to place on a strategy in the iterated Prisoner's Dilemma.

Summarising this review of social dilemmas, the Prisoner's Dilemma is a useful game theoretical approach to analysing cooperation, especially in self-interested agents. There are many different approaches to developing successful strategies, where success is measured as creating and sustaining cooperation over time. Success is also measured in terms of payoff for a given strategy. These measures of success are useful to provide context and examples of how the strategy behaves. Analysis of evolutionarily stable strategies provides a useful measure in addition by allowing a broader view of how the strategy will react and thrive.

3.4 Models of Emotional Computation

There are a number of different implementations that use different models of emotions as part of a decision making process. This section will describe a number of different models that have been implemented in a computational model. To begin, one of the uses for simulated emotions is the ability to adapt in human-agent interactions, either by simulating the emotions for a particular agent or by analysing the current emotions of the human. André et al. [1] present a number of experiments which implemented simulated emotions inspired by the OCC model in addition to simulated affect and personalities. The simulated emotions, affect, and personalities are used to constrain the number of options that are available to choose. Their first experiment *Puppet* shows how different emotional states can change a character's behaviour, and how actions can affect the emotional state of the characters [1]. The aim of creating these agents with emotions and personality was to show children that a persons' action can affect how they feel. The outcome of their experiments concludes that hand crafted responses that are stereotypical for the emotions allow a complex and consistent model of an agents' affect state over vastly different interactions and application domains.

The next piece of work uses emotions in human-agent interactions to help an agent learn; Broekens and Haazebroek [18] implement smiling as a positive reward and frowning as a negative reward which they incorporate into a reinforcement learning task. The agents' task is a reasonably simple task of learning a path towards a goal, in this case *food*. The authors compared two different reinforcement agents, one with the emotion interpretation and one without. They noted that when the human emotion is given, the learning of the agent increases, however once the stimuli is removed the speed of learning reduces to the same level as the agent without the emotional learning [18]. The authors then allowed the agent to learn what emotional stimuli would be given for a state. This resulted in the agent learning quicker as it did in the previous experiments, but now it no longer slows down once the emotional stimuli is removed [18].

Gadanho and Hallam [48] provided a different implementation of emotions in a reinforcement learning agent. The robot in this case had a number of sensations such as *hunger* which is the energy deficit and *pain* which is high when the robot is colliding. These sensations, in addition to a number of others, informed its emotional state, which is given as 4 different possibilities; happy (everything is ok), sad (energy is low), fearful (bumping into obstacles), angry (not moving) [48]. Due to the agent being applied to a real robot rather than a simulation, the reinforcement behaviour occurs when the dominant emotion changes. The experiment compared the agent against four other kinds of robots which did not use the emotional controllers. They found that the emotional agent had significantly less collisions and was significantly happier for longer than the other agents [48]. These experiments show us that simulated emotions can be incorporated into decision making models and can improve the results when compared to singularly using the model.

Jacobs et al. [62] performed a similar experiment to Gadanho, however instead of using the 4 simulated emotions, Jacobs et al.'s experiment aimed to simulate a subset of the OCC model. The subset of simulated emotions included Joy, Distress, Hope, and Fear and these were incorporated into a reinforcement learning framework, finding that the agent was able to learn an effective policy and that the rules of the OCC model were respected [62].

Using other psychological frameworks to simulate emotions is also possible. Marsella and Gratch [77] provide a means of using appraisal theory to help detect emotions in humans and simulate emotions in their agents [77]. One of their key insights was to show how psychological theories of emotions serve as a specification for an agent's design in order to simulate emotions and affect. The authors also delve into the difficulty of evaluating these simulated models as the intended outcomes can differ drastically, and there is a need to tailor specific scenarios that allow empirical evaluation of the models [77].

Agent interpretation can be without a strong psychological grounding, as the number of possible implementations for a interpretation is vast. Steunebrink et al. [114] tackle this issue by providing a logic which encompasses the OCC model. This logic allows objects, actions, and events to be formalised in the context of the OCC model of emotions, leading to a change in an agents' behaviour [114]. A key issue with providing a logic is that there is still a further step in order to provide an implementation of this model, leaving some interpretation of how a given action fits a particular simulated emotion.

3.4.1 Simulated Emotions in Social Dilemmas

We limit the range of previous work by now only looking at agent interactions, without any human intervention, in social dilemma situations. When looking at social dilemmas such as the Prisoner's Dilemma, Ashlock and Rogers [3] looked at whether an extremely simplistic model of emotions would have an effect on cooperation. The experiment included simulations with and without noise in the action selection. Noise was given as a percentage change of the chosen action changing to another action. The model of emotions that the authors used in [3] is simulated as a bit, where 1 is the opponent has cooperated more than defected, otherwise it is set to 0. In their iterated simulation of agents using lookup tables and artificial neural nets, where they are evolved with and without the emotion bit, they found that most of the results were not surprising. The level of cooperation drops when noise is added, and neural nets were not as cooperative as lookup tables. The addition of the emotion bit changed the evolved strategies, with some improvements and some negative results [3]. Improvements included an increase in cooperation in the society, the negatives included an increase in defection when the noise increased. Even though there was a lack of definition in how the emotion bit affected the evolved strategies, they managed to show that even extremely simple emotions can have large effects in the iterated Prisoner's Dilemma.

Continuing with the iterated Prisoner's Dilemma, this section will now consider work on implementing a formal model of emotions from psychology into a framework which can be computed, where a number of emotions from the OCC model of emotions are implemented into the decision making process of the agent. Lloyd-Kelly et al. [71] initially modelled two emotions as a subset of the OCC, which were Gratitude, and Anger. The simulated emotions were implemented to respond to the opponents' actions in the Prisoner's Dilemma. Gratitude is increased by a cooperating opponent, and the agent will cooperate when the gratitude threshold is reached, while anger is increased when the opponent defects, again causing the agent to defect when the anger threshold is reached. They were able to replicate the popular Tit-For-Tat strategy [5] using these emotions, and some strategies were effective at adapting to the variety of opposing strategies.

Lloyd-Kelly et al. continued by adding a number of other simulated emotions to the subset; admiration [70], and hope [72]. An agent's admiration increases when another agent has a greater score than itself which occurs after five interactions and when the admiration threshold is reached the agent will use the opponent's strategy. Hope increases when the agent and opponent are both cooperating, however the hope value will reset when the outcome is not mutual cooperation. When the hope threshold is reached the agent will defect against the opponent, with the aim of achieving the highest payoff for itself, as the agent is self-interested. These additional emotions help the authors show how simulated emotions are able to aid a decision-making process which can avoid some of the pitfalls of social dilemmas [69], such as protection against greedy opponents.

3.4.2 Simulated Emotions within Reinforcement Learning

Section 3.2 described reinforcement learning and two of the major implementations. There has been some work on how to incorporate simulated emotions into a reinforcement learning framework. Yu et al. [126] uses simulated emotions in addition to a number of different reinforcement learning algorithms. The emotions the authors have chosen to model are social fairness and individual fairness, both given as a number between -1 and 1 [126]. Additionally the authors tested these agents in a number of different social dilemmas, and a number of different network configurations including: Small-world, Scale-free, and Grid based [126]. The outcome of these experiments showed that this interpretation of simulated emotions assisted in focusing the reinforcement learning task to improve cooperation when compared other reinforcement learning algorithms, such as Q-Learning [126].

Moerland et al. [82] provide a full review of emotions used in reinforcement learning agents in a number of different scenarios. The review looks at how emotion can benefit three groups of communities involved in reinforcement learning, the machine learning community, the human-robot interaction community, and finally the affective modelling community. Focussing on the sections with purely agent-agent interactions, the review shows how emotions can be a benefit to the agents [82]. The benefits improve learning efficiency by using simulated emotions to provide inspiration for motivation, exploration, and parameter tuning [82]. The emotions can affect differing parts of the reinforcement learning process such as:

Reward Emotions affect the reward received, has similarities to reward shaping [85].

State Emotions are part of the learning state.

Meta Emotions are an inherent part of the model, with parameters of the model representing emotions.

Action Emotions affect how the action is selected.

Simulated emotions in agents have a multitude of different implementations ranging from simple deterministic actions, to implementations in a larger non-deterministic model [82]. The ability to call the values used in these implementation a reasonable representation of a simulated emotion also differs, with some that seem to be simulated emotions in name only, to implementations that stick as close as possible to the psychological models.

3.5 Models of Moody Computation

Similarly to the simulated emotional model implementations, this section will now look at some of the implementations of moody models and how they have been defined. Santos et al. [106] simulate mood, in addition to emotions and personality using the PAD[3], OCC[4], and OCEAN [35] models respectively. The implementation integrated each of these models together as part of a decision making process within an argumentation protocol with the aim being to help facilitate agreement between agents [106]. There are a few issues with this account, namely that the agents need to know the other agents' personality and the lack of experimentation with these agents.

A similar implementation which uses mood, emotion, and personality was conducted by Sakellariou et al. [103]. In addition they included emotional contagion, which is how the emotions of a particular agent can influence another agent's emotions. The authors chose to use a different implementation of emotions and moods compared to the Santos implementation. Emotions and mood were both implemented using the Circumplex model of affect, showing that these models have an effect on a multi-agent El-Farol problem. The El-Farol problem is when an agent has choice of going out to the El-Farol bar or staying at home. If more than 60% of agents go the bar then the bar is too busy and they would be better off staying at home. If less than 60% go then they have a great time at the bar. All agents must choose at the same time. Specifically looking at the mood implementation, the value assigned to mood was static over the experiments and influenced the emotion value by dragging this value slowly over towards the mood value.

Continuing with implementing mood using the PAD model, Arellano et al. [2] tackle a different problem. Instead of implementing emotions as part of agents' decision-making processes, they experimented with taking simulated facial expression and mapping them to points on the PAD model. The authors set out to show that each mood type can be mapped to an expression, and found that Pleasure and Arousal are relatively easy to calculate from an expression but Dominance proves to be more challenging as the expressions were static

[3] "Pleasure-Displeasure, Arousal-Non-arousal, and Dominance-Submissiveness Temperament Model" described in Section 2.2

[4] "Ortony, Clore, and Collins model of emotions" described in Section 2.1.1

images [2]. They also showed that each mood type has an associated expression. The authors used the same set of mood types and description of mood as Santos et al. [106].

In terms of mood and emotion in human-robot interaction, Ojha and Williams [87] have designed an Ethical Emotion Generation System for use in a robot or agent when interacting with a human. The outcome of the work shows how the emotions of the robot are different depending on the age of the human that is interacting with the robot [87]. For example when the user kicks the robot, if the user is an adult the robot becomes angry, however if the user is a child then the robot would instead become sad. The authors use a different model of emotions and mood based on appraisal theory [107]. The authors modelled mood using the same model as the emotions. The difference between the mood and emotions is that the mood value changes slower than the emotions [87], the simulated mood and emotions both influence each other's values.

There are a range of different mood model implementations, which are often implemented in conjunction with emotional models and occasionally personality models. Although there is a wide range of mood models, there is a general lack of psychological grounding. Some models make use of the PAD model, but often lack the justification for how the model changes the agents' action.

3.6 Environmental Effects on Decision Making

The aim of this section is to consider the nature of the environments the strategies operate within and the effects that different environments can have on a given strategy's success. The majority of the prior literature already surveyed mainly dealt with the theoretical outcomes of the strategy. In order to effectively analyse strategies in social dilemmas, there is a need to look at how the strategy reacts in both networked simulations and mobile simulations. Agents in a network are nodes that interact with their immediate neighbours. Agents in a mobile simulation are free to roam a simulated environment. This next section will now review the literature on what the thesis will term *environmental effects* on interactions.

Firstly this section will start by looking at networked interactions and how they can affect agents in social dilemmas. Ranjbar et al. [95] provides an introductory insight into these effects. The authors introduced a Prisoner's Dilemma game, where the payoff[5] is

[5] Ranjbar et al. use the term Fitness rather than Payoff [95]

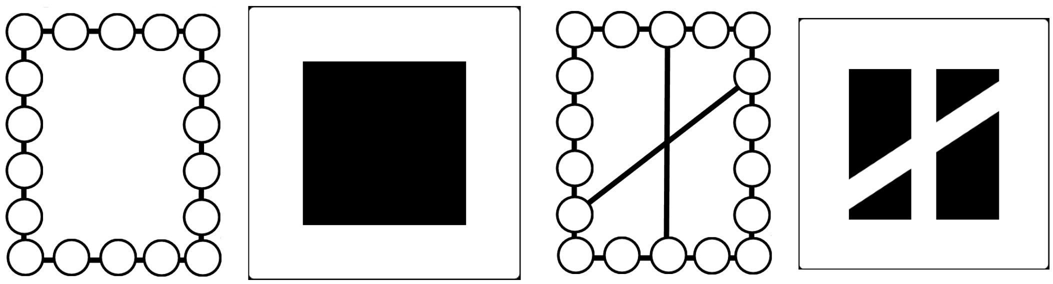


Figure 3.1: Example graphs with equivalent arenas, for the regular and small world environments respectively.

measured by how far a robot has moved. Cooperation was implemented as the agent moving out of the way of the opponent, defection was the agent continuing moving straight on. The designs of the arenas for the robots and the networks is an important aspect to consider. In Ranjbar et al.'s work [95] they have used mobile arenas which are designed so that they physically represent a network structure with the connections providing the navigable areas. The network structures they looked at were the small-world and regular structures. Figure 3.1 demonstrates an example of a small world network and a possible equivalent in a mobile arena, along with an example of regular networks and arenas.

Next is the definition of regular, scale-free, small-world, and random network structures in terms of how they are constructed.

Regular All vertices have the same degree. That is, each vertex has the same number of edges. For random regular networks, the connections to specific vertices may differ.

Scale Free For any given node the number of edges that extend from it have a power law distribution [6]. Intuitively each node in the network is well connected but there are a number of "hub" nodes which have a very high number of connections.

Small World There is a high clustering of vertices and the path length between any two nodes is the logarithm of the number of vertices [125]. Intuitively this is a network where the nodes have very few neighbours but the distance between any two given nodes is also small.

Random A network where an edge between any two vertices in the network has some given probability of existing.

Ranjbar et al. [95] used 20 robots, and simulated them walking in the mobile environment. They found that there is a difference between the regular and small-world arenas, which mimics the same differences found in the equivalent networked interactions [125, 7]. One of the unexpected behaviours that affected the interactions was how robots had difficulty finding opponents when there was a high density of robots, which caused them to act more like defectors as they crashed into each other.

For the experimental test-bed, which will be introduced in Chapter 4, there is a need to know what effects the different network structures have. This knowledge will allow the analysis of the results to attribute the observed effects to the correct origin, such as the introduction of mobility or the structure of the arena. The work uses a range of different experiments and social dilemmas but the chapter will focus on what effects can be attributed to the differences in network structure, which may only be in passing, and the differences between the experiment and previous work. All of the following work uses networked agents which interact with their immediate neighbours.

Bloembergen et al. [13, 12] studied learning agents in small-world and scale-free networks, finding that in their experiments with the Stag Hunt that cooperation spreads best in the networks with high connectivity [12]. The authors also found that an agent's influence spreads best when a node has high connectivity and there are relativity few of these types of nodes in the Prisoner's Dilemma [13].

Ranjbar et al. [93, 94, 96] provides further evidence of high connectivity affecting cooperation when comparing scale-free and small-world networks [93], and also shows that the same effects are present in regular networks [94]. The authors also showed that when a network is constructed over time, and the agents choose where the next node goes, the end network has scale-free properties [94]. Starnini et al. [112] provides a useful addition by changing the game to the Stag Hunt, the authors showed that there are some underlying effects the structure of the network has on the game that should be explored. In addition the authors also showed that the network that is created when the underlying game has changed is still a scale-free network [112].

Moving onto random graphs, there is a notable difference between graphs which have high connectivity and graphs with low connectivity. Durán and Mulet [36] found that for graphs that have a low connectivity, the number of cooperators that exist in the Prisoner's Dilemma is dependant on the initial conditions. When the connectivity is increased, the number of cooperators becomes independent of what the initial condition of the network is [36].

Hofmann et al. [60] brings a lot of this together, testing grid, random, scale-free, and small-world networks with agents playing the Prisoner's Dilemma. The main insight from Hofmann et al.'s work [60] is that general statements on cooperation in networked interactions cannot be made and that cooperation depends on what kind of network is used, along with state update rules and what fraction of the society are initially cooperative.

Interactions between agents have assumed that the actions chosen were what the agent intended; "noise" is a percentage chance that the action unintentionally changes. Vukov et al. [120] has showed that noise has an effect on what can be expected from structural effects of regular networks, with noise decreasing the number of cooperators. Taking a different interaction model where agents can create a random edge with another player and edges can be deleted, Szolnoki and Perc [116] found that by allowing time between the deletions and the creation of new edges, cooperation between the agents can be supported.

Having a diverse set of public good games that the agents conduct in the same network, allowed Santos et al. [105] to show that regular graphs distributed payoffs evenly, while payoffs in scale-free networks follow a power law. The scale-free resulted in fewer poor agents and more rich agents for a fixed cost to each individual agent. When implementing the Prisoner's Dilemma Santos et al. [104] showed that cooperators in a scale-free network were better supported and could flourish when compared to a regular network where the cooperators died off.

When using a simulation where the agents are able to move in a network, Ichinose et al. [61] introduces nodes with no agent associated with them, and agents can now change positions to an empty node. The movement supported cooperation in regular graphs and scale-free networks [61]. The authors suggest that cooperation can supported when a network allows movement [61].

There are a number of effects that a network can have on a society of agents, with the structure of the network proving to be significant. Studying these effects proves to be difficult, given the wide range of possible social dilemmas, strategies, and networks. In order to study agents that have mobility, these effects need to be isolated in order to appropriately attribute each noted effect to its correct cause.

3.7 Summary

In conclusion, this chapter has reviewed four separate bodies of work, namely simulated emotions in a computational model, simulated mood as a computational model, social dilemmas in general, and how environmental structure has an effect on social dilemmas. For simulated emotional models, there is a wide body of work which has shown that what can be termed an "emotion" in a simulated system varies widely. The simulated mood showed again that what can be considered "mood" in a computational setting varies significantly across the literature, with few examples of agents that use mood exclusively as a decision maker. Social dilemmas, such as the Prisoner's Dilemma, provide an effective means of analysing cooperation among agents. Evolutionary stability analysis allows an effective comparison of human-inspired agents with strategies that do not use simulated emotions or mood. Environmental structure proves to be an important factor when analysing agents in different network types, and the environment is an important consideration for agents that have full mobility.

The chapter has identified the following gaps in the literature which the research questions of this thesis will be able to address:

- Does mobility affect decision making in emotional and moody agents? How do mobile environments affect cooperation?
- How to design computational model of mood grounded in psychology? Is there an implementation of the mood model which enables the evolution of cooperation?
- Are simulated emotional and moody agents evolutionarily stable? Which attributes of these agents break or maintain stability?

The knowledge from Chapters 2 and 3 provides the background information that is needed to construct effective experiments that allow this thesis to answer the main research question **RQ**.

Chapter 4

Emotions and Mobility

In this chapter the generic outline of the experiments and agent setups that will be used throughout this chapter and thesis are given. Each choice involving the experimental setup will be justified, to enable the experiments to fully contribute to the research question set out in Chapter 1. The aim of this chapter to construct a number of experiments that provide results that provide answers for the sub research questions **SRQ1**, **SRQ2**, and **SRQ3**. Starting with Section 4.1, this section gives the outline of the generic experiment and the design of the simulated agents, and the implementation of simulated emotions, along with justification for each part. The first experiment validates the implementation of the simulated emotional agents and is described in Section 4.2 and 4.3. Once the agents have been validated by the first experiment, a second experiment will be described and conducted to provide results for **SRQ1**, **SRQ2**, and **SRQ3**. The outcome of the second experiment is detailed in Section 4.5, which also gives the specific implementation. In addition to the two experiments, a proof that any two emotional agents will converge to a repeating outcome is given in Section 4.4.3. The chapter will then conclude in Section 4.6 which will summarise the results and show how these experiments have provided a step towards answering the research question. The conclusion will summarise the chapter and also motivate the later chapters. The work in this chapter has been published in [22, 21, 24].

4.1 Experimental Setup

This section will set out the design of the experiments. The start of the section will define the design of the arenas and how they have been constructed, and how the arenas are

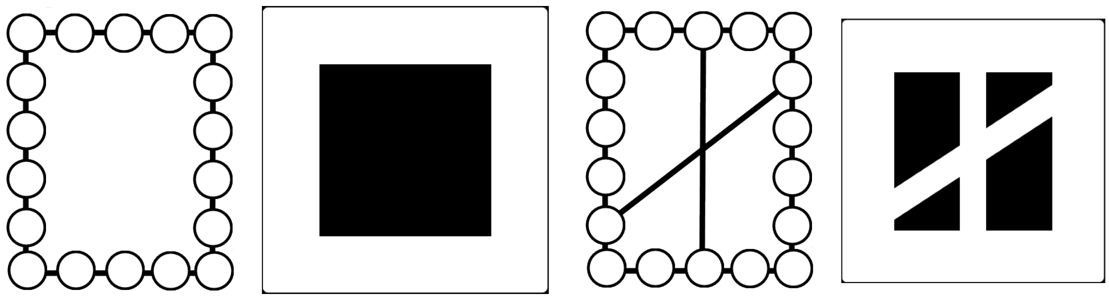


Figure 4.1: Graph followed by arena for the regular and small world arenas respectively, where both arenas are $5m^2$

based on their networked equivalent. The next part will move onto the design of how the agents will move and how the agents interact in the arenas, before describing how the emotional model will be implemented into the agents. There are a number of factors that will be specific to each experiment, namely the number of agents and the number of times the experiment will be repeated. The agents themselves are simulated in the Stage [49] simulator. The simulations throughout this thesis were programmed in C++, and conducted on a Late 2013 27-inch 3.4 GHz iMac with 32GB of RAM. A single simulation of 10 minutes typically took 5 seconds to complete when no visual output was requested.

4.1.1 Arena Design

The purpose of the arena design is to enable investigation into how the environment structure affects the interactions between agents. The two structures that have been selected for the experiments are the regular and small world structures. These structures have been selected to enable comparisons with the results of previous work that used similar environments [95]. Given a network, the edges will be converted to traversable areas within an arena. Figure 4.1 shows the regular and small-world networks and the arenas they have been converted into. Both arenas are five meters square, to allow agents space to freely move even when the area becomes crowded.

Each agent that is placed in the arena will be given a randomised starting location, such that they do not overlap each other or are located within any walls. The agents in the experiments need to have a number of different attributes distributed among them evenly, such as initial action and admiration threshold. Algorithm 1 gives the pseudocode of how the attributes are distributed.

Algorithm 1: Algorithm to distribute attributes evenly among agents

```
Input: List of Agents, in a random order
Input: List of Attributes, with each having a list of the tuples ⟨Setting, Set, Ratio⟩
Result: Agents with Attributes Evenly Distributed
while There is an Attribute to Distribute do
    i ← 0;
    foreach tuple in Attribute do
        numToSet ← i + (Ratio * |Agents|);
        for k ← i to numToSet do
            Agents[k].Setting ← Set;
        end
        i ← numToSet;
    end
    Shuffle Agents;
end
```

There will now be a demonstration of the algorithm using two examples of attributes that will be distributed. The first one will set the initial action, which is a basic yes or no type of attribute. The second example will set the attributes for the admiration threshold, this threshold can be set to one of three levels; high, medium, or low. This second example shows how attributes with multiple possible settings get set.

Example 1 - Initial Action The actions need to be distributed so that 60% of the agents initially choose to defect and 40% choose to initially cooperate. The Settings list will have two tuples; ⟨Action, Defect, 0.6⟩ and ⟨Action, Coop, 0.4⟩. The first 60% of agents in the list get their initial action set to defection, and the last 40% of agents get their initial action set to cooperation.

Example 2 - Admiration The distribution of the admiration thresholds must be that 50% of agents are set to a high threshold ⟨Admiration, High, 0.5⟩, 25% to a medium threshold ⟨Admiration, Medium, 0.25⟩, and 25% to a low threshold ⟨Admiration, Low, 0.25⟩. The first 50% of agents in the list are set to a high admiration threshold, much like the first example. The next 25% are set to a medium threshold. At this point 75% of agents have been assigned their admiration threshold. Finally the last 25% of agents in the list are assigned the low admiration threshold.

Combining the two examples, after the initial action has been distributed, the list of agents will be randomised before the admiration is distributed. Randomising the order of the list of agents between setting each attribute prevents combinations of attributes being over represented. In this case the list was not randomised then all the high admiration agents would have their initial action set to defection, which may affect the results of any experiments.

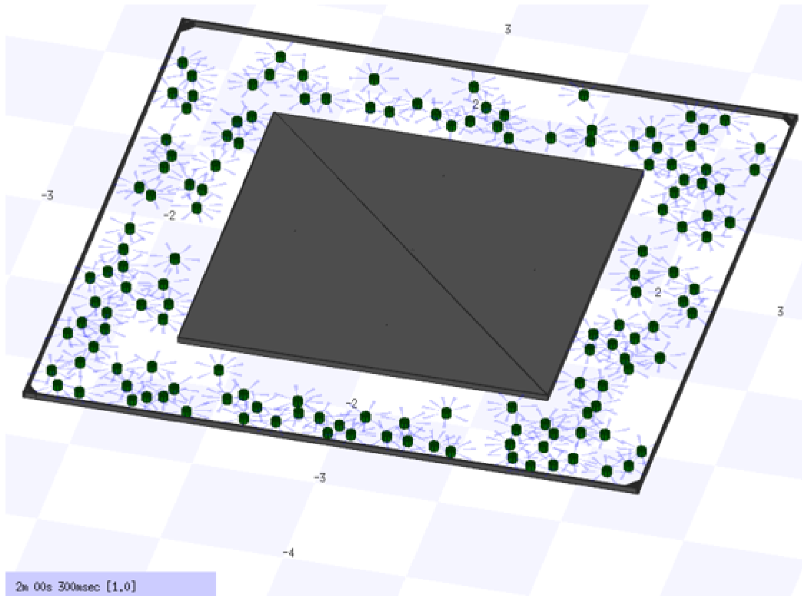
4.1.2 Physical Agent Design

The agents are simulated as e-puck [83] robots where each agent has proximity sensors to detect walls and obstacles, located at {-90, -45, -15, 15, 45, 90} degrees with respect to the robot's heading. E-puck robots were chosen for their small size, quick movement, simplicity in simulation, and possibility of moving to a real world implementation.

The agents are given a random walk behaviour with some basic obstacle avoidance procedures. If the sensors on the left detect anything, the agent will stop and then turn to the right, and the reverse for the right sensors. The agents' speed is set to 10 centimetres per second and they can turn at speeds of up to 45 degrees per second. When the agent detects no obstacles, a turn speed between -45 and 45 degrees per second is selected while continuing to move forward. A new heading is generated each time the robot receives sensor data, which in the simulation is every millisecond. When the turn speed is selected, this affects the power that goes to one of the wheels. For example if the turn speed selected is 20, then the power to the right wheel is reduced, allowing the robot to arc towards the right. Given how often the turn speed is changed, the robot will move erratically around the arena. This also causes the robot to always move forward, allowing for the exploration of the full environment. Using the random walk also allows the exploration of the effects of the arena construction without being concerned whether a particular effect was due to a movement selection.

One round of the chosen social dilemma is initiated between two agents whenever these two agents are within 20cm of each other, and both have line of sight of their opponent. Each round is performed instantaneously, with no effects on the physical movement of the agents. This is to ensure that the processing time does not affect the results. To prevent the agents playing multiple games as they come into contact, there is a one second delay before the agent can play against the same opponent again. The delay allows the agents to have an interaction and move to another space effectively. A screenshot of a number of agents being simulated is provided in Figure 4.2.

Figure 4.2: Image of e-pucks being simulated in the Stage [49] simulator.



After the round is completed the agents will then continue their random walk behaviour. The agents have no knowledge of the payoffs or the number of games to be played, and will purely use the strategy that the agent has been assigned. The agents have no knowledge of what strategies are being used by their opponents, but can differentiate between the agents. The strategy that has been assigned to an agent will be applied to each opponent rather than all opponents, unless noted otherwise. Finally all the agents have no knowledge of the environment; they will only use the random walk behaviour driven by their sensor inputs.

In some situations the agents might be required to obtain some information from the opponent, notably the average payoff for the admiration threshold for the emotional strategy. In this case the agent can freely request the information from the opponent. The opponent will accurately give any information requested by an opponent. While fully truthful agents maybe an unrealistic proposition, exploring how lying affects the agents hinders the ability to answer the main research question.

4.1.3 Emotional Agent Implementation

Lloyd-Kelly's [70, 71, 69, 72] implementation of emotional agents has been chosen as the basis for the implementation in this work. This implementation has been chosen due to:

Table 4.1: Emotional agent thresholds, including descriptive names

Anger Threshold	*Gratitude* Threshold	Character	In Lloyd-Kelly et al. [72]
1	1	Responsive	E1
1	2	Active	E2
1	3	Distrustful	E3
2	1	Accepting	E4
2	2	Impartial	E5
2	3	Non-Accepting	E6
3	1	Trustful	E7
3	2	Passive	E8
3	3	Stubborn	E9

the simplicity in simulation, the body of work that is available so that the results can be compared, and that the simulated emotions are grounded in psychology, based on a subset of the OCC model of emotions [89]. There are four emotions that have been simulated by Lloyd-Kelly [72]; *Anger*, *Gratitude*, *Admiration* and *Hope*. In this work the subset of the OCC model that will be simulated are the *Anger*, *Gratitude* and *Admiration* emotions. *Hope* will not be simulated as this reduces the body of work that is available for comparison, in the experiments that will be conducted. Additionally *Hope* introduces a randomness element to the decision making which will add noise to the data.

Each emotional agent can have one of many emotional "characteristics", each characteristic has different thresholds for both *Anger* and *Gratitude*. The full list of the characteristics and their thresholds is given in Table 4.1. *Admiration* thresholds are rated as high (three), medium (two) or low (one). The *Admiration* thresholds are distributed among the characteristics randomly, as described in Section 4.1.2.

The number of emotions simulated from the OCC model is intentionally kept low, so as to minimise the difficulty in analysing effects from when mobility has been added. The specific emotions and their thresholds, along with their effects on an agent's actions are repeated from Lloyd-Kelly's previous work [71, 70]. The example will be given in the context of the Prisoner's Dilemma. After an action, a simulated emotional value may increase. If this value is equal to or greater than the corresponding simulated emotional threshold, then the value gets reset and the agent's behaviour changes due to the threshold being reached.

As an example, take the Prisoner's Dilemma game, where if an opponent defects the *Anger* value increases and if the opponent chooses to cooperate the *Gratitude* value will increase. Consider for example the Accepting characteristic, which is initially cooperative, and an opponent that defects against the Accepting agent. Accepting's *Anger* value will increase by one. In the next round Accepting will continue to cooperate with the opponent. Assume the opponent defects again: Accepting's *Anger* value is now two and has reached the *Anger* threshold. Accepting's *Anger* value will reset back to zero, and when the Accepting agent plays another round with that particular opponent, the Accepting agent will choose to defect.

Admiration works in a similar manner; when the *Admiration* threshold has been triggered the agent will take on the characteristics of the opponent which triggered the *Admiration* threshold. In essence the Admiration agent becomes the opponent agent. In Lloyd-Kelly et al. [70], the *Admiration* threshold increases when an agent compares its total payoff against each of its neighbours after every five games. For the agents in the mobile arenas, the neighbours are not well defined because the agents will be moving constantly, which changes who they are near to at a particular time. In order to account for the differences between the mobile and static arenas there will be a modified version of the trigger for *Admiration*, which is described below.

After an agent has five interactions with any combination of opponents, the agent will then request the current average payoff of the next opponent. The agent will compare the average payoff of the opponent to its own average payoff. The *Admiration* threshold will increase towards whichever agent has the higher average payoff, if they are equal the threshold will not increase. The average payoff is used rather than total payoff, which was used by Lloyd-Kelly et al. [70], because there is no guarantee that each mobile agent has engaged in the same number of games as its opponent.

In summary for a social dilemma, when triggered, each emotional threshold will change the agent such that:

Anger Starts defecting against the opponent

Gratitude Starts cooperating against the opponent

Admiration Changes the *Anger*, *Gratitude*, and *Admiration* thresholds to match the opponent's

4.2 Validation Experiment

The aim of this first experiment is to show that the implemented simulated mobile agents have the same emotional response and outcomes as the static agents reported by [71]. In this experiment the subset of emotions will be further reduced to only include *Gratitude* and *Anger*, as these were the emotions used in the original experiment [70], and to stop the *Admiration* threshold from altering any of the emotional responses. The emotional agents will play the iterated Prisoner's Dilemma against a fixed-strategy agent that does not use emotions. The emotional agents will be set to cooperate initially. The non-emotional agents have the same knowledge of the world as the emotional agents. They have the same random walk behaviour and the same limited knowledge about their neighbours. The fixed strategies that the emotional agents will be tested against are the traditional ones from Axelrod's tournament [5] and are described in [70] but are reiterated here:

Mendacious Always defects

Veracious Always cooperates

Random Equal chance of defection or cooperation

Tit-For-Tat Initially cooperates, then mimics the opponent's last move

Joss Tit-For-Tat with a 10% chance of defection

Tester Defect on round n, if the opponent defects play Tit-For-Tat until the end of the game, otherwise cooperate until round $n+2$ then repeat the strategy by Defecting at $n+3$ as if $n+3$ is round n

In this experiment there are only two agents in the environment: the emotional agent, and the fixed-strategy agent. For each emotional characteristic shown in Table 4.1, 10 runs will be performed against each fixed strategy in turn. A run consists of 200 rounds of the Prisoner's Dilemma game, and then the mobile simulation will stop. The number of rounds is equal to the number of rounds used in Lloyd-Kelly [71]. Additionally the simulation of the agents will be restricted to the regular arena. The hypothesis is that the average payoffs will be identical to the average payoffs found in Lloyd-Kelly's work [71]. The addition of movement has no direct effect on the decision-making process, so introducing a delay between each interaction of the two agents should make no difference. In addition there is

Table 4.2: Average payoffs for a strategy vs an emotional agent. Results for the static agents are reprinted from [71]

vs	Responsive Static[71]	Responsive Mobile	Trustful Static[71]	Trustful Mobile
Mendacious	204, 199	204, 199	212, 197	212, 197
Veracious	600, 600	600, 600	600, 600	600, 600
Random	451, 449	459.4, 457.4	630.4, 372.4	618.6, 367.4
Tit-For-Tat	600, 600	600, 600	600, 600	600, 600
Tester	533, 533	533, 533	668, 443	668, 443
Joss	233.4, 228.4	256.3, 251.3	523.4, 449.4	531.2, 467.2

the hypothesis that there maybe some small variations in the Random and Joss strategies, as they use randomness to decide their action.

4.2.1 Validation Results

After the simulation the results generated will be compared to those in [71]. Responsive and Trustful are the only results provided as they are the only results presented in Lloyd-Kelly et al. [71].

Table 4.2 shows that the implemented mobile agents have the same emotional reaction and therefore receive the same average payoff. From the table it can be seen that against agents which do not have randomness as part of their decision making, the mobile agents perform identically to their non-moving counterparts. When the emotional agents interact against agents which have a random element in their decision making, it can be seen that the average payoffs between the two types of agent are sufficiently close, and that the same strategy is dominant by having a larger average payoff. With the knowledge that the implementation of the emotional agents is valid, larger amounts of emotional agents can be introduced and the experiments can begin to answer the research question.

4.3 Initial Experiment

The initial experiment sets out to that mobility has an effect on the success of emotional agents, which begins to answer **SRQ3**. The experiment will be conducted in two different arenas which are shown in Figure 4.3. Success of the emotional agents is measured by the

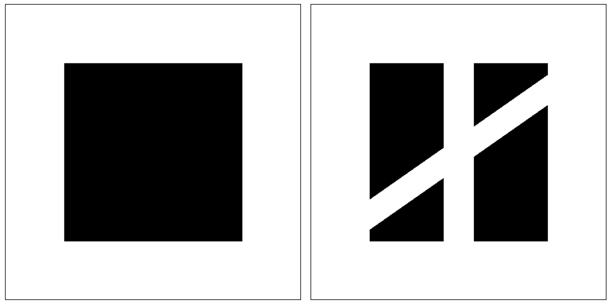


Figure 4.3: The two arenas used for the initial experiment. The regular arena is on the left, with the small world arena on the right

level of cooperation among the agents, as well as which emotional characteristic has become the most represented. The results of this experiment will be compared to Lloyd-Kelly's work [71], which conducted the same experiment only there was no mobility for the agents. The scenarios that will be conducted will vary the ratio of initial defectors, as well as varying the level and ratios of the *Admiration* thresholds. The full list of scenarios and their ratios are shown in Table 4.3.

Each scenario will be conducted against a number of sub-scenarios, which specify the number of agents that will be simulated. The aim of varying the number of agents being simulated is to analyse whether the density of agents has any effect on the experiment. The number of agents will range from 9 to 144, and the emotional characteristics will be equally distributed among the agents. The emotional agent characteristics are shown in Table 4.1. Table 4.4 shows the sub-scenarios, and the number of simulated agents, along with how many of each characteristic will be present at the beginning of the simulation. Keeping the initial distribution of characteristics equal prevents any characteristic becoming the most represented due to being the most represented initially.

Each combination of scenario and sub-scenario will simulated 10 times, with each simulation allowing the agents to move and interact for 10 minutes which allows for sufficient interactions and replication to take place.

The hypothesis for this is experiment is that as the density of agents increases, the range of opponents an agent can be expected to face will reduce. The cause for the hypothesis is that as the number of agents in the arena increases, the agents will be unable to move away due to the other agents blocking the exits. This will lead to the agents facing the same opponents more often. An additional hypothesis is that the most represented agent will be

Table 4.3: Scenarios for the initial experiment, where the ratios of initial actions and Admiration thresholds vary. These scenarios are identical to the ones in Lloyd-Kelly et al. [70]

Scenario	Initially Defects %	Initially Cooperates %	Admiration % High	Admiration % Medium	Admiration % Low
1	90	10	34	34	32
2	70	30	34	34	32
3	50	50	34	34	32
4	30	70	34	34	32
5	10	90	34	34	32
6	50	50	50	25	25
7	50	50	70	15	15
8	50	50	90	5	5
9	50	50	25	50	25
10	50	50	15	70	15
11	50	50	5	90	5
12	50	50	25	25	50
13	50	50	15	15	70
14	50	50	5	5	90

Table 4.4: Sub-scenarios of the scenarios presented in Table 4.3, showing the number of agents along with the total number of each emotional characteristic.

Sub-scenario	Density Description	Number of Agents	Number of each emotional Characteristic
1	Very low	9	1
2	Low	36	4
3	Medium	72	8
4	High	144	16

the Trustful agent, as in Lloyd-Kelly et al.'s work [71]. The least successful characteristics are hypothesised to be in a different order, due to the larger range of opponents that will be faced. If all the given hypotheses are true, then the experiment can conclude that mobility does effect which emotional characteristics are successful but the best performing characteristics are able to overcome the inclusion of mobility.

4.4 Initial Experiment Results and Analysis

This section will report and analyse the results of the experiment described. Firstly this section will consider what the effect was on the average payoff of the emotional agents in general when varying the ratio of initial cooperators and defectors. Then it will be shown how the outcome between two emotional agents becomes fixed, with no variation. The algorithm to determine this will be given, along with an exhaustive explanation on why the algorithm is true. After, the analysis will consider which characteristics become the most represented in the society. The effects density has on the emotional agents will be described next, before finally analysing how the arena structure affects the society of emotional agents. The conclusion of the experiment will be given, along with how the results fit to the wider research question of this thesis.

4.4.1 Effects of initial actions

Table 4.5 shows the average payoff of an emotional agent for a given arena. Looking specifically at the differences in the ratio of initial cooperators and defectors (Scenarios 1-5). Scenario 1 has 90% of the emotional agent initially defect, and Scenario 5 has 90% initially cooperation. Scenarios 2-4, are ratios between these two extremes, the full description is given in Table 4.3. The results show that the average payoff increases as the proportion of initial cooperators increases. The results show the same ordering as in Lloyd-Kelly et al. [70]. The average payoff increasing as the number of initial cooperators increases is an expected result. The number of cooperators increases the chance of a mutual cooperation outcome, while the chance for a mutual defection outcome also decreases. Mutual cooperation has a higher payoff than mutual defection leading to a higher average payoff for the agents when the number of initial cooperators increases. The differences between the Regular arena and the Small World arena will be discussed in detail in Section 4.4.7.

Table 4.5: Average Payoff (Standard Deviation) of an emotional agent in each arena for the initial experiment

Scenario	Regular Arena	Small World Arena
1	1.2269 (0.0407)	1.2162 (0.0454)
2	1.6522 (0.1065)	1.6243 (0.0738)
3	2.0536 (0.1213)	2.1156 (0.1231)
4	2.4590 (0.0791)	2.4020 (0.0936)
5	2.8099 (0.0611)	2.8171 (0.0260)
6	2.0986 (0.1230)	2.0903 (0.0330)
7	2.0397 (0.0951)	2.1252 (0.0544)
8	2.1608 (0.0855)	2.1215 (0.0769)
9	2.0108 (0.1199)	2.1117 (0.1166)
10	2.1027 (0.0890)	2.0289 (0.0955)
11	2.0953 (0.0728)	2.1333 (0.0637)
12	2.0355 (0.0558)	2.0028 (0.0916)
13	1.9732 (0.0643)	2.0215 (0.0455)
14	1.9342 (0.0749)	1.9312 (0.0992)

4.4.2 Mutual Outcomes

Considering the interactions between pairs of agents, there are a number of patterns that emerge. When two emotional agents start with identical initial actions, the result of the game will either be continued mutual cooperation or defection without deviation. The agents will not deviate as the emotional thresholds that would change the action will not increase. When the initial actions are different, the emotional agents will have a number of asymmetrical outcomes which will then turn to mutual defection or cooperation and continue the mutual outcome indefinitely.

There is a third possible outcome for two emotional agents, that they will continue asymmetrical outcomes indefinitely. The indefinite asymmetrical outcome occurs when two conditions hold, which are based on their anger and gratitude thresholds. The first condition is that the emotional agent's anger threshold is equal to the opponent's gratitude threshold. The second condition that must hold is that the emotional agent's gratitude threshold is equal to the opponent's anger threshold. The two emotional agents will continuously swap which agent is the cooperator and which agent is the defector as both the cooperators' anger threshold and the defectors' gratitude threshold are met at the same time. This third outcome will happen irregardless of which agent is the initial cooperator or defector.

Table 4.6: The mutual outcomes that occur between two agents i and j with differing initial actions. I_i is the initial action of agent i, Cm is mutual cooperation, Dm is mutual defection and Rp is an asymmetrical outcome.

Character	Responsive	Active	Distrustful	Accepting	Impartial	Non-Accepting	Trustful	Passive	Stubborn
Responsive	Rp	Dm	Dm	Cm	I_j	I_j	Cm	I_j	I_j
Active	Dm	Dm	Dm	Rp	Dm	Dm	Cm	I_j	I_j
Distrustful	Dm	Dm	Dm	Dm	Dm	Dm	Rp	Dm	Dm
Accepting	Cm	Rp	Dm	Cm	Cm	I_j	Cm	Cm	I_j
Impartial	I_i	Dm	Dm	Cm	Rp	Dm	Cm	Cm	I_j
Non-Accepting	I_i	Dm	Dm	I_i	Dm	Dm	Cm	Rp	Dm
Trustful	Cm	Cm	Rp	Cm	Cm	Cm	Cm	Cm	Cm
Passive	I_i	I_i	Dm	Cm	Cm	Rp	Cm	Cm	Cm
Stubborn	I_i	I_i	Dm	I_i	I_i	Dm	Cm	Cm	Rp

The outcome that the two emotional agents will settle on is dependant on the gratitude and anger thresholds of both agents. Table 4.6 shows what the mutual action will be between two emotional agents that use the emotional characteristics defined in Table 4.1. The algorithm to determine what the mutual outcome will be between two emotional agents is given in Definition 3. To validate Equation 4.1 an analysis of all possible situations is shown.

Definition 3. $\Omega_{i,j}$ returns the mutual action of emotional agents i and j. A_i is the anger threshold of agent i. G_i is the gratitude threshold of agent i. Ac_i returns the current action of agent i. C is cooperation. D is defection.

$$\Omega_i^j = \begin{cases} C, & \text{If } (Ac_i = Ac_j = C) \text{ or } (Ac_i = C \text{ and } G_j < A_i) \\ D, & \text{If } (Ac_i = Ac_j = D) \text{ or } (Ac_i = D \text{ and } A_j < G_i) \\ NotMutual, & \text{If } A_i = G_j \text{ and } G_i = A_j \\ \Omega_j^i & \text{Otherwise} \end{cases} \tag{4.1}$$

4.4.3 Exhaustive Analysis of Mutual Outcome Equation

Assumptions Emotional agents are paired to play iterated PD games, both start with zero anger and gratitude. The assumption of zero anger and gratitude only holds in

Table 4.7: All possible interactions for emotional agents and the matching mutual outcomes

Current Outcome	Condition	Result
(C, C)	Any	Both players' gratitude increases, hence they keep co-operating and remain in (C, C) indefinitely.
(D, D)	Any	Both players' anger increases, they keep defecting and remain in (D, D) indefinitely.
(C, D)	$A_c = G_d$	After $N = A_c = G_d$ rounds of (C, D) both players switch strategy and play (D, C) thereafter.
	$A_c < G_d$	A_c is reached before G_d, hence the cooperator switches to defection after $N = A_c$ rounds and (D, D) is played thereafter.
	$A_c > G_d$	G_d is reached before A_c, hence the defector switches to cooperation after $N = G_d$ rounds and (C, C) is played thereafter.

two-player interactions, for multiple players the admiration emotion starts playing a role as well.

Notation Actions are C (cooperate) and D (defect); the anger threshold is A, subscript denotes a player, e.g. A_c is the anger threshold of the cooperating player; and similarly the gratitude threshold is G. Each time A or G is reached its value is reset to zero.

Enumerating all possible interactions Based on initial actions of both players and conditions on their values of A and G, all possible outcomes can be enumerated. This is shown in Table 4.7

Example Suppose Responsive meets Active. If Responsive plays C and Active plays D, they will switch to (D, D) after 1 round since $A_c = 1 < 2 = G_d$. If Responsive plays D and Active plays C, they will swap strategies after 1 round since $A_c = G_d = 1$ and play (C, D) for 1 round (since now $A_c = 1 < 2 = G_d$, as before), and then switch to (D, D) again.

4.4.4 Most Represented Characteristic

Looking at which characteristic is most represented among the emotional agents, Figure 4.4 shows the percentage of each characteristic where they were the most represented after the

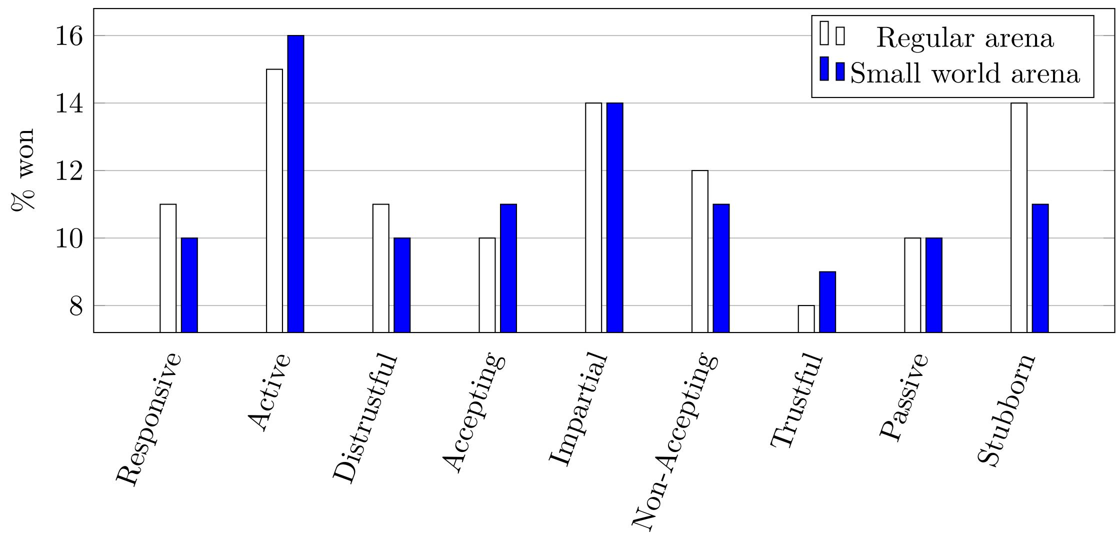


Figure 4.4: Most represented emotional characteristic excluding draws, as a percentage of the number of games played. Grouped by arena.

10 minute simulations. Draws were not counted. The results were not as hypothesised; rather than Trustful being the most successful agent, which was the case in Lloyd-Kelly et al. [70], Active was the most dominant agent in the simulations conducted. Not only was Trustful not the most successful, but it was actually the worst performing of all the characteristics.

The reason behind the Trustful characteristic's failure comes down to the fact that it takes a long time to settle into its mutual outcomes. Trustful will take up to three defections before it changes the opponent to mutual cooperation and with the average number of interactions between two specific agents only being 3.37 (Standard Deviation 10.71) then the agent is unlikely to settle into mutual cooperation during the run. The admiration threshold will trigger in the Trustful agents, making them vulnerable to being replaced with a different characteristic as they will have a low average payoff because they are being taking advantage of.

Active's success is due to its ability to taking advantage of the forgiving strategies, leading to a high average payoff in very few interactions. When Active is cooperating and is taken advantage of, it immediately switches to defection, preventing the high average payoff being affected. With the high average payoff from very few interactions, Active is consistently triggering other agents' admiration thresholds, allowing Active to spread.

Table 4.8: Average scores (Standard Deviation) for all agents across all scenarios for varying number of agents.

Number of Agents	Regular Arena	Small World Arena
9	1.64 (1.41)	1.24 (1.43)
36	2.04 (0.46)	2.01 (0.57)
72	2.03 (0.29)	2.11 (0.29)
144	2.02 (0.17)	2.08 (0.18)

Responsive is affected by responding to cooperation too quickly as it is unable to boost its score quickly since it tries to avoid taking advantage of other agents by reciprocating cooperation. While it may appear that Distrustful should therefore do well, as it responds quickly to defection and takes advantage like the Active characteristic, the aggressiveness of the Distrustful characteristic means that almost all other characteristics will settle on mutual defection quickly, while Active is able to maintain some mutual cooperation allowing it to raise its average payoff more.

4.4.5 Density Effects

The thesis will now look at whether varying the number of agents affects the average payoff, and if there are any differences in the average payoff between the regular and small world arenas. Table 4.8 shows the average payoff with the standard deviation. The standard deviation falls when the density of agents is increased. The reason behind the standard deviation of the average payoff falling when the number of agents is increased is fairly intuitive; increasing the density means that the agents have less time between individual interactions, increasing their overall number of interactions. Table 4.9 gives the average number of interactions with the standard deviation for each density of agents. In very low densities the agents are only going to get one or two interactions against an opponent, which prevents them settling on an outcome and since there are so few games there is a large disparity in the results as very few agents can respond to their opponent's actions.

When both the density and the number of interactions increases, the variance in average scores becomes much closer, however the average payoff itself is slowly falling. When the number of interactions is very high the agents will settle into their mutual outcomes with the majority being mutual cooperation or defection, whereas in slightly lower densities they have not settled, leading to a majority of asymmetrical outcomes. When half of the agents

Table 4.9: Average number of interactions (Standard Deviation) between specific agents for varying number of agents.

Number of Agents	Average Number of Interactions
9	1.56 (0.83)
36	1.92 (1.93)
72	2.15 (1.73)
144	3.13 (15.08)

Table 4.10: Average payoff (Standard Deviation) for an agent per interaction based on distance travelled in a small world arena.

Distance Travelled	72 Agents	144 Agents
High	2.11 (0.06)	2.13 (0.09)
Medium	2.05 (0.06)	2.10 (0.09)
Low	1.97 (0.11)	2.02 (0.07)

are in mutual cooperation and the other in mutual defection, the overall average will be 2, but the asymmetrical outcome average will be around 2.5. This difference in the theoretical average payoff shows why there is the slight dip in average scores when the agents are in higher densities.

4.4.6 Distance Travelled Effects

There is a noticeably larger variation when there is a significant number of agents. On observation of the agents in the arenas there were a few agents that were trapped in the corners of the arena unable to get past the agents that were blocking them in. These blocking agents were also unable to move, leading to a number of agents who only interacted with a few agents, but these interactions where repeated at every time step as the agents could not move away. From this there may be a notable difference in how the distance travelled would affect the average payoff. To achieve this the agents were ordered by distanced travelled and split them into three groups of the same size. The three groups are categorised as: the high movers who have moved 30 meters or more; medium who moved between 15 and 30 meters; and low who moved 15 meters or less. The average scores are shown in Table 4.10. The table shows that the more an agent moves, the higher the average score. When movement is high the number of repeated interactions decreases, leading to

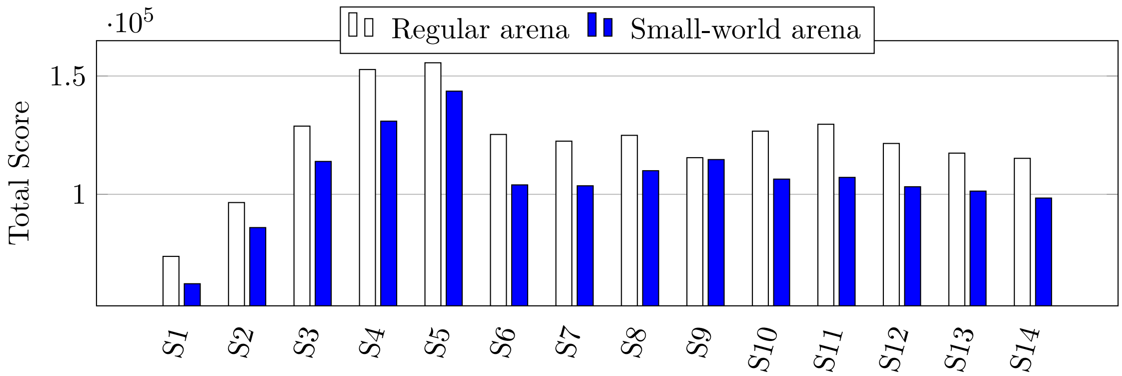


Figure 4.5: Total scores for each scenario.

these agents not settling on their mutual outcomes, which is the reverse for the low movers. When the number of interactions is low, the average score is slightly higher and this is again reflected by the distance moved.

4.4.7 Environmental Effects

The results in the previous section will now be considered in the context of what the environmental effects the two different arena structures had on the results. Figure 4.5, showed that increasing the number of initial cooperators, yields higher average payoffs. Differing the admiration thresholds does not have a significant effect on the total payoffs. However there is a difference in total score between the arenas; this is due to the fact that small world arenas have a larger surface area than the regular arena, and as shown previously, there are less interactions between specific agents in lower densities. Given that the density of agents is lowered in the small world arena, it should follow that the average score should also increase. This will be tested in Scenario 3, as it has equal distributions of initial actions and admiration thresholds, the average score (Standard Deviation) of the regular arena is 2.05 (0.12) and in the small world arena the result was 2.11 (0.12) providing more evidence that lower repeated interactions between agents will increase the average score.

Table 4.11 shows the average scores for scenarios 8, 11, and 14 which have differing admiration thresholds. From the table it can be seen that the regular arena is relatively stable whereas the small world arena shows a drop in scores from high to low thresholds. The difference is due to the percentage of unique interactions in each arena, which is shown

Table 4.11: Average payoff (Standard Deviation) per game for an agent based on distribution of admiration thresholds in both arenas.

Scenario	Regular	Small World
S8 (High)	2.09 (0.05)	2.26 (0.12)
S11 (Medium)	1.97 (0.16)	2.16 (0.03)
S14 (Low)	2.10 (0.11)	1.91 (0.02)

Table 4.12: Breakdown of interactions in both arenas across all runs.

Arena	Total Interactions	Unique Interactions	Unique Interactions%
Regular	417372	115261	28%
Small World	361682	115653	32%

in Table 4.12. In small world arenas the table shows that agents interact with individual agents less often. When the agents come to replicate using their admiration thresholds, the chance of an agent replicating into a characteristic which is not dominant is increased. This is due to the short term performance of a characteristic not reflecting the long term performance that can be achieved by the characteristic. Replicating characteristics that achieve success in the short term prevents the agents from achieving higher scores based on characteristics that have died out.

4.4.8 Conclusions and Alterations

This experiment has investigated the evolution of cooperation in mobile emotional agents. The initial experiment has shown that the distance travelled, the type of arena, and the density of the agents all have an effect on the success of the agents: these all affect the number of unique interactions. The arena type affects which strategies are viable, with the Stubborn characteristic being successful in a regular arena, but not as successful in a small world arena. However, strategies exist that do well regardless of the arena type, such as the Active characteristic.

In answer to the questions posed in the introduction of this experiment, it was shown that mobility and arena types do affect the simulated emotional agents, and as a result different emotional characteristics become more successful as compared to those of [71]. The Active characteristic being the most successful was an unexpected result as it shows that mobility has a large effect on the success of a characteristic. The experiments have shown that in regular arenas the total payoffs increased when there were more initial cooperators, as also shown by [95]. In contrast, in the small-world arena, Ranjbar et al. [95] found payoffs decreasing with the addition of cooperators, whereas in this experiment the payoffs have seen an increase. The experiment shows that the payoffs were affected by the arena structure. From this experiment further experimentation on how arena structure effects emotional agents will follow in the expended experiment.

4.5 Expanded Experiment

This next experiment will provide further evidence of the effects seen in the previous experiment. Knowing that the available space affects the result, the design of the arenas will now ensure that the arenas have the same amount of floor space. With this variable removed, any differences that are noted between the arenas will be due to structure rather than differences in floor space. To help validate that the differences are due to structure additionally two more arenas are introduced to be used alongside the regular and small world environments. Examples of all the arenas are shown in Figure 4.6.

The four environments to be constructed as arenas are:

Empty An empty arena

Regular An arena with a block in the middle

Small World An arena with a block in the middle, the block has paths that can be traversed. The paths will provide a shortcut to other parts of the arena.

Random An arena with blocks distributed randomly around the arena.

The empty arena is constructed to have no obstacles. The random arena is different for each run of the experiment, its shape is constructed from the regular arena. The inner obstacles are split into twenty equal sized blocks which are then placed randomly within the arena, each block is guaranteed to not overlap any previously placed block.

For each of these scenarios there will be a number of sub-scenarios using a different numbers of agents. The number of simulated mobile agents will range from 27 to 108, with each emotional characteristic being represented equally in each sub-scenario. The exact numbers for each density are given in Table 4.13. These sub-scenarios have been included

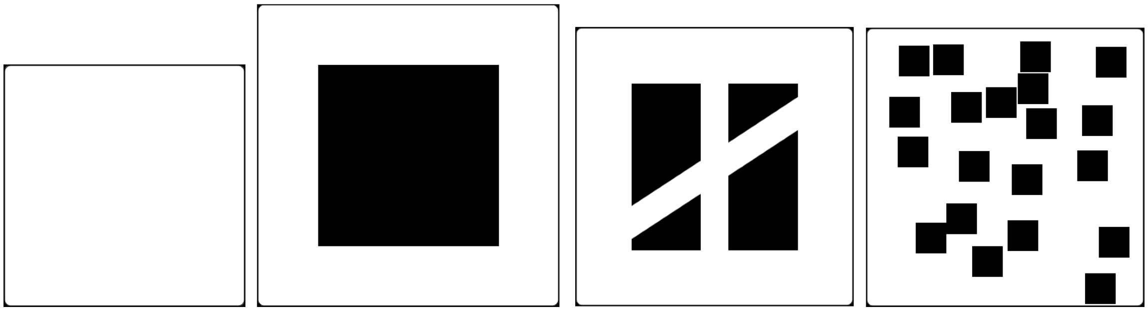


Figure 4.6: Examples of arenas used. The arenas have the same amount of traversable area. Arenas left to right are: Empty Arena, Regular Arena, Small World Arena, Random Arena.

Table 4.13: Altered sub-scenarios for each scenario presented in Table 4.3. The sub-scenarios vary the number of agents. Also shown is the number of each emotional characteristic.

Sub-scenario	Density Description	Number of Agents	Number of each emotional Characteristic
1	Very low	27	3
2	Low	56	6
3	Medium	81	9
4	High	108	12

as there is the expectation when the number of agents increases, the density of the agents will increase, since the arena is still the same size. The prediction is that the effects seen in the initial experiment should be replicated as each arena is the same shape. For the random and empty arenas the hypothesis is that the empty arena will show a more extreme version of the small world arena, while the random arena will give a more extreme version of the regular arena as it restricts the movement of agents more.

The number of agents have changed from the initial experiment as the hypothesis is that in the random arena the very low densities of the initial experiment of 9 agents will struggle to interact at all. The experiment has also lowered the very high densities as some of the arenas are smaller to account for the differing floor space; ensuring that all robots are able to fit into the arena and are able to move reasonably freely.

Having an equal distribution of emotional characteristics initially ensures that the test of characteristic strength is not based on having an initially higher representation. In the experiment each combination of scenario and sub-scenario is conducted 10 times. Each run will last for 10 minutes during which the agents move around and interact, which

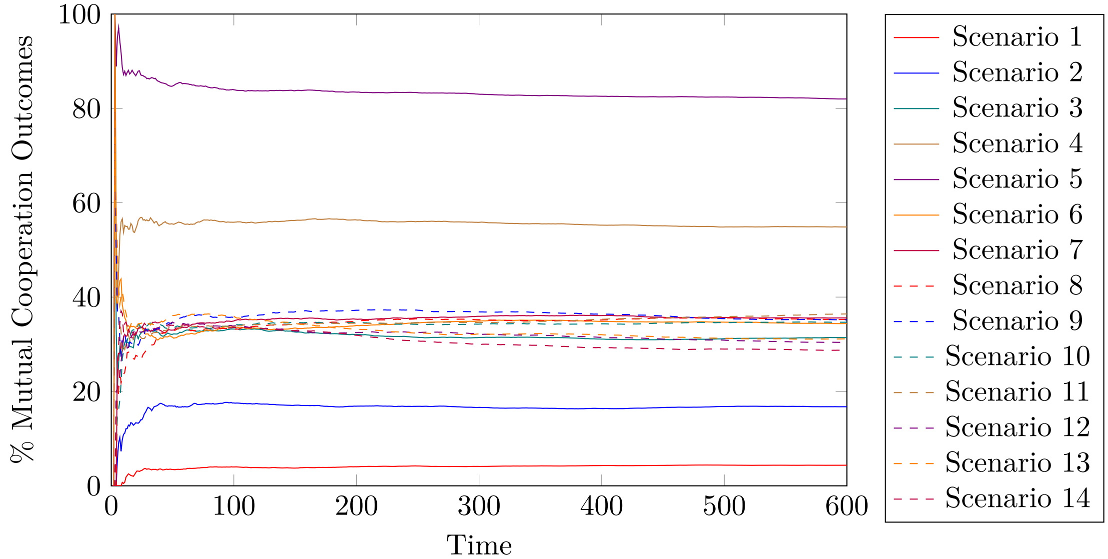


Figure 4.7: Level of cooperation by scenario in the expanded experiment.

allows sufficient interactions and replication to take place. The experiment will record data for each interaction including: agents involved, actions chosen, current number of games, current average, time initiated, and distance travelled. The experiment also records the number of each characteristic at the end of the run, as well as the final averages for each agent. This provides a good dataset to perform a deep analysis on the emotional agents.

The hypothesis is that the Active agent will be most dominant, as per the initial experiment. If the Active agent continues to be dominant in all arenas then the experiment can conclude that some strategies are more successful despite differences in arena or floor space.

In this section the results will be presented and discussed for the experiment detailed previously, firstly analysing the experiment by looking at the cooperation levels, followed by the successful characteristics, and finally the effects of agent density. The analysis will also take into consideration the effects that the arena has on each section.

4.5.1 Cooperation Levels

The results of the level of cooperation between the agents in the experiment as shown in Figure 4.7. The figure shows that the cooperation is stable with the level of cooperation achieved being in proportion to the starting level of initial cooperation. The reason that

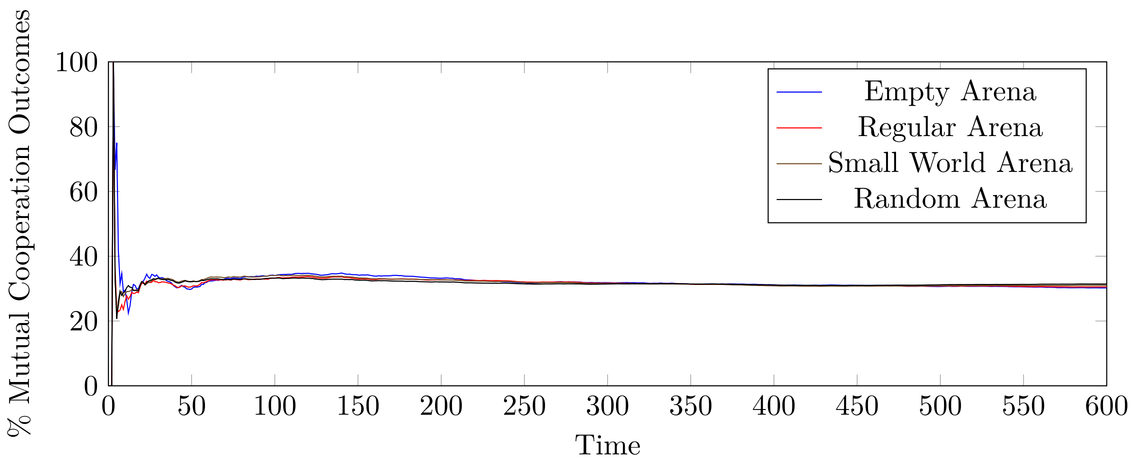


Figure 4.8: Level of cooperation per arena, when starting with equal levels of initial cooperation and defection.

cooperation does not change over time is that only agents which have an initial asymmetrical outcomes will change their action. When looking at Table 4.6 the results show that there are an equal amount of mutual cooperation and defection endings with a outcomes between two emotional agents that will repeat a pattern of asymmetrical outcomes. This is also the case of agents in mutual outcomes, as they will continue to be in either mutual cooperation or mutual defection.

The cooperation levels are also stable when analysing more closely the differences in arenas. Figure 4.8 shows the different cooperation levels for each arena in Scenario 3. Scenario 3 has an equal distribution of initial cooperators, defectors, and admiration levels. The figure shows that the arenas have some very minor variation but they all are around the same level of cooperation at the end of the run. This shows that the arena structure does not have a direct effect on cooperation levels between emotional agents. From this analysis the experiment can conclude that the differences in the results shown in Figure 4.7 are due to the initial settings of the experiment and the emotional characteristics of the agents. Next the analysis will look at which characteristics are successful, and compare the successful characteristics to the initial experiment.

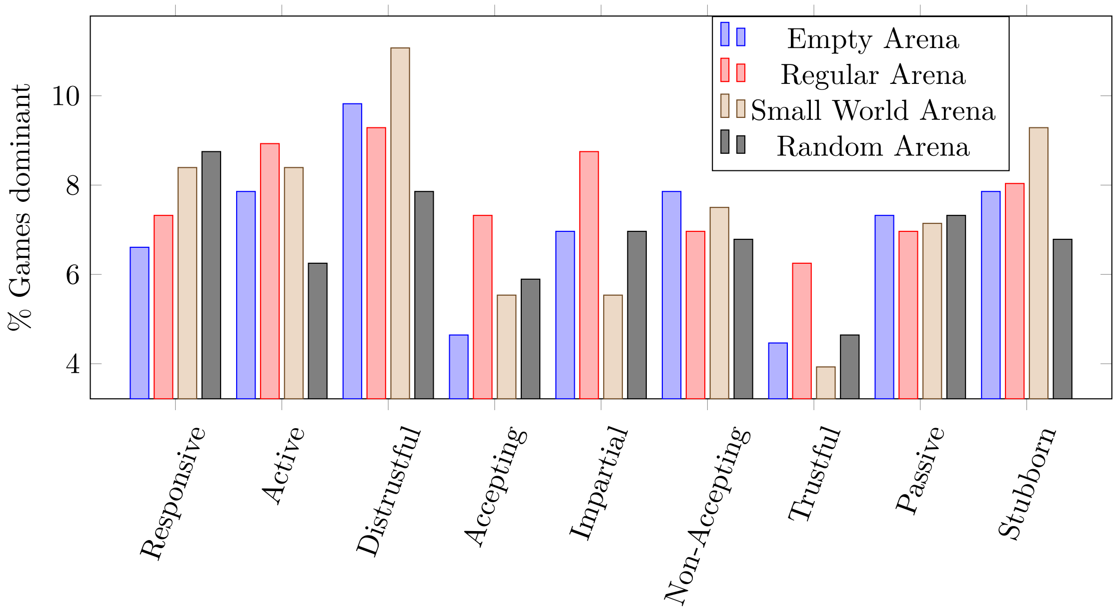


Figure 4.9: Most successful characteristic in the expanded experiment, for each arena.

4.5.2 Successful Characteristics

This section will investigate which characteristics are the most successful, where success is measured as how often a characteristic becomes dominant. Dominant characteristics are defined as having replicated so that they make up the majority of the agents. Lloyd-Kelly et al. showed that the Trustful agent was most successful in a static environment [71], however when mobility was added, the initial experiment showed Active as the most successful, due to the much larger range of agents played against and the few times that each agent played against each other. Figure 4.9 shows the results for the expanded experiment where, again as per the initial experiment, Trustful performs poorly. However Distrustful is the most successful in the empty, small world, and regular arenas and in the random arena Responsive was the most successful.

The unexpected results came from the regular arena and the small-world arena, with the successful characteristic being Distrustful. The initial experiment reported that Active and Stubborn were the most successful in the regular arena and Active and Impartial for the small world arena. The difference between the two most successful characteristics in the regular arena is very small in both this experiment and the initial experiment, with the

Table 4.14: Interaction distribution for each arena.

	Empty	Regular	Small World	Random
Total Interactions	391994	437983	537089	418825
Unique Interactions	141618	133569	112875	79211
Percentage Unique	36	31	21	19

Active and Stubborn characteristics being the most successful. This leads to the conclusion that Active and Stubborn are both dominant characteristics in the regular arena and the ordering comes down to random chance.

The small world arena does not have this same outcome with the two highest performers being Distrustful and Stubborn. The highest performers are Active and Impartial in the initial experiment. To see why this was the case, there is a need to look at the number of interactions per arena, which can be seen in Table 4.14. The small world arena has more total interactions and less unique interactions, and a lower percentage of unique interactions than the initial experiment where the values were 361682, 115653, and 32% respectively as per Table 4.12.

The random arena had the effect of separating the agents into groups, limiting the range of agents that could be played against. This causes the number of games with a particular agent to increases when compared to the other arenas. This in turn changes the dynamic of the game as agents are able to boost their scores with mutual cooperation and prevent losses with mutual defection, unlike more open arenas where this dynamic is reversed. This leads to the most successful agent being the agent which is able to place itself into mutual outcomes the quickest, which is the Responsive characteristic.

The empty arena allows for agents to play against the largest range of agents. The dynamic here is that an agent encounters an individual agent less often than in the random arena. The most successful agents respond to cooperation slowly; this allows the agent to sucker-punch its opponent without retaliation, raising its payoff quickly. Since these agents are defecting they are not open to being sucker-punched themselves. Distrustful becomes the most successful characteristic as it is able to minimise the amount of times it is on the receiving end of a sucker-punch, which lowers the payoff, as it responds to defection quickly.

There is a physical difference in the small world arenas in this experiment and the initial experiment. In this experiment, the arena design takes into account the difference in available floor space between the each of different types of arena structure. The initial

Table 4.15: Average payoffs (Standard Deviation) for an agent based on distance travelled.

Distance	84	108
high	2.07 (0.67)	2.06 (0.60)
medium	2.04 (0.70)	2.03 (0.58)
low	1.62 (1.21)	1.86 (0.94)

experiment did not take this into account so the small world arena has more space than the regular arena. This causes the agents in the small world arena to be more cramped as the width of the corridors is reduced. The reduced width forces the agents closer together causing them to have more interactions with the same agents. However, for the agents that do manage to move around the arena a lot, they will meet a wider range of characteristics, leading to their situation being more like an empty arena and the most successful characteristics reflect this.

The most successful characteristics can be attributed to the agents that moved the furthest as shown in the initial experiment. Inspecting the results shown in Table 4.15 for this expanded experiment, the same attributes and effects can be seen.

When the arenas become more closed, the payoffs achieved by taking advantage early become more important, especially if the agents are able to interact with other agents more quickly. This effect is most noticeable in the small world arena where the dominant characteristics take the most advantage of other characteristics, with Distrustful being the most dominant as it protects its payoff the most. The regular and small world arenas are similar, however in the expanded experiment the regular arena acts more open due to its larger corridors. Dominant characteristics in the regular arena react quickly to defection as previously noted, however this arena also allows consistent interactions with the same agent. Agents that are taken advantage of by their opponents can still become dominant if they also take advantage of their opponents. This is seen by the success of the Passive, Stubborn, and Responsive characteristics.

In the random arena, the agents are more limited in the range of characteristics they can interact with. This closed off arena allows the Active characteristic to become the most dominant by a wide margin. The advantage that can be achieved from defecting in this arena is reduced as the agent is likely to be punished since the chance that the agent meets the same agent again is heightened; however, a small advantage can be taken provided that the agent protects the payoff quickly by reacting to this punishment. This is seen by the

Table 4.16: Average Payoffs (Standard Deviation) for an agent based on number of robots in an arena.

No. Of Robots	Empty	Regular	Small World	Random
27	2.00 (1.13)	2.03 (1.04)	2.02 (1.03)	1.51 (1.40)
54	2.07 (0.79)	2.03 (0.79)	2.02 (0.86)	1.72 (1.18)
84	2.06 (0.66)	2.04 (0.68)	2.00 (0.78)	1.85 (1.05)
108	2.05 (0.57)	2.03 (0.58)	2.02 (0.69)	1.94 (0.82)

success of Responsive and Active.

4.5.3 Density Effects

Looking at the average scores of an agent in differing densities as shown in Table 4.16, there is a noticeably large variance in the average payoff. Increasing the density lowers the standard deviation. When the density increases and the number of interactions increases as well, the variance in average scores becomes less pronounced. When the number of interactions is very high the agents will settle into their mutual outcomes with the majority being mutual cooperation or defection, whereas in lower densities the mutual outcomes of the agents have not been achieved, leading to a majority of mixed outcomes. When half of the agents are in mutual cooperation and the other half in mutual defection, the overall average will be 2, but the mixed outcome average will be around 2.5, showing the slight dip in the average scores in higher densities.

Similarly in the random arena, the standard deviations become smaller and the average payoff moves towards an average of 2 over time. However due to the random arena causing the agents to separate into different groups which do not interact with each other, this causes the average payoff to drop significantly in lower densities. This can be also be seen in the very low densities reported in the initial experiment where the agents achieved an average payoff (Standard Deviation) of 1.64 (1.41) and 1.24 (1.43) for the regular and small world arenas with 9 agents respectively, as shown in Table 4.8.

Table 4.17 shows the average scores for scenarios 8, 11, and 14, which have differing admiration thresholds. In the empty, regular, and small arenas, the average payoff decreases over time, whereas the random arena is stable. The initial experiment attributed the differences in arena to the number of unique interactions, which is partly reinforced in the expanded experiment.

Table 4.17: Average payoff (Standard Deviation) of agents based on different arenas and distributions of admiration levels, highlighting the how average payoffs are related to the admiration level.

	Empty	Regular	Small World	Random
Scenario 8 (High)	2.13 (0.67)	2.10 (0.67)	2.08 (0.77)	1.85 (1.05)
Scenario 11 (Medium)	2.10 (0.66)	2.08 (0.67)	2.06 (0.77)	1.90 (1.01)
Scenario 14 (Low)	1.95 (0.66)	1.92 (0.68)	1.95 (0.72)	1.75 (1.02)

An effect noted in the initial experiment was that when agents interact with individual agents less often and they come to replicate using their admiration thresholds, there is an increased chance of them replicating into a characteristic which is not dominant. Agents choosing to replicate a non-dominant characteristic is due to the short term average payoff being larger than the long term average payoff that can be achieved by the dominant characteristic; this then prevents the agents achieving higher scores. The effect still holds in the expanded experiments because the number of unique interactions is around the same in the small world arena, and increased in the regular arena when compared to the previous work. Thus the results lead to the conclusion that if the agents interact with the majority of the agents in the arena, the average payoff will be highest for the high admiration thresholds. If the agents do not interact with the majority of agents, as is the case for the random arena, then the payoffs will be stable.

4.6 Conclusion

In this chapter three experiments have been conducted that had the aim of contributing to the main research question **RQ**. The contribution of this chapter includes experiments which have provided a suitable computational setting of emotional decision making in social dilemmas (**SRQ1** and **SRQ2**). The experiments also provided evidence that mobility affects the outcome of cooperation amongst agents (**SRQ3**).

The chapter provided a construction of an experimental setting that uses an arena, and a number of agents interacting. The agents are assigned an emotional characteristic from a list of possible implementations. This emotional characteristic is then used in the agents' decision making process. Emotional characteristics are defined from Lloyd-Kelly et al. [69], as the implementation is grounded in work from psychology literature, in addition to being computationally viable.

The first experiment showed that mobility has an effect on the outcomes achieved by emotional agents, changing not only the average payoffs, but the characteristics that are the most successful in terms of replication. The reason that these changes occur is due to the number of unique interactions that the arenas enforce, with the regular arena having fewer unique interactions as there was less room for the agents to move about in. The lack of space caused the agents to interact with their neighbours more often. The small world outcomes from the experiments in this chapter contradict the work of Ranjbar et al. [95] which had mobile agents in a regular and small world environments. The reason for this contradiction is in the design of the experiment. In Ranjbar et al. [95] mutual defection will cause the two mobile agents to crash, and they noted that in higher densities of agents, more unexpected crashes occur, whereas in the experiments conducted in this chapter, cooperation and defection do not affect the movement of the agents.

The second experiment was adjusted so as to find out whether the unique interaction effects were due to the density of the agents, or the structure of the arena. The second experiment was also expanded to include two more arena structures, the empty arena and the random arena. The outcome of this experiment showed that the unique interactions were linked to the structure of the environment, decreasing the randomness of the arena increased the number of unique interactions. The experiment also showed that the kind of characteristics that are successful need to be quick to respond to defection, while taking a small advantage. The most trusting characteristics still performed the worst.

The work described in this chapter yields two further lines of inquiry, The first is whether the mobility effects seen are generalisable to agents using other strategies, which is addressed in Chapter 6. The second aspect is whether this implementation of emotional agents is applicable to the wider literature on social dilemma agent strategies, which is addressed in Chapter 7. The next chapter will look at another implementation of a psychologically grounded agent strategy, one that incorporates an interpretation of mood.

Chapter 5

Modelling Mood as a Computational Decision Maker

This chapter contributes to the research question **SRQ1** by providing a model of mood. The chapter will also contribute to research questions **SRQ2**, **SRQ3** by showing how the developed model of mood is implemented in agents that have mobility. The experiments are conducted in a similar manner to the experiments in Chapter 4. The chapter opens with Section 5.1 which provides the definition of the generic mood model, which will be supplemented with an explanation of the reasoning and psychological grounding for the choices made in the creation of this model. Following this, in Section 5.2, a specific implementation of the mood model that will be used in the experiments is given, along with the psychology grounding needed to validate the implementation against the developed model. The first set of experiments in Section 5.3 looks at cooperation levels and how resilient the implemented model is against a number of pure defectors. The agents using the mood model are then implemented in multiple arenas in Section 5.4. The final set of experiments in Section 5.5 will look at how to implement the mood model using reinforcement learning to underpin the decision making process rather than the emotional model that was used in the previous experiments. Finally Section 5.6 concludes the chapter, by summarising the contributions of this chapter to the research question. The work in this chapter has been published in [23, 25, 24].

5.1 Generic Mood Model

In this section the definition of the mood model and the psychology justification for the choices made is given. This section focuses on the generic mood model and as such will avoid having any specific implementation details. The aim is to allow the mood model to potentially be used in other areas that are not specific to computer science. The start of this section justifies how mood is represented, and will also show how the mood model splits mood into a number of different levels. For each level of mood the details on how the decision making process should be affected is given, along with the psychological justification for each choice.

As noted in Chapter 2 Section 2.2, mood is defined as having three main categories, positive, negative, and neutral. This reflects how mood is often represented in the psychology literature [127, 56, 122]. When considering how to update the mood after an interaction, the current mood should always be considered as part of the process, as was shown by Aspinwall [4].

When considering negative moods, the model should not rely on heuristics, nor should the model be concerned with how long the decision takes. The decision that is made by the model should therefore be more thought through, and clearer in why a choice was made. Decision making while in a negative mood has been shown to be more considered [4], and the decision takes longer to make when compared to positive moods [57]. From this work it was shown that being in a negative mood would lead to a more "rational" choice [108, 57]. In addition when considering the affect the action has on mood, the impact the action has on what is learned should be greater when the agent is already in a negative mood. The psychological grounding for this decision is that negative moods have a stronger effect on memory and learning, as shown by Baumeister et al. [8].

Moving onto positive moods, they are the opposite of negative moods. The model must make use of the available heuristics as part of the overall decision making process and aim for the ideal outcome. When the mood increases the model should put less weight on the risks of an action and rely more heavily on the given heuristics to achieve an ideal outcome. While the decision making process shouldn't be rushed in positive moods, the heuristics should ensure that the decision making process is not slower than when in a negative mood. Positive moods using heuristics and aiming for a global ideal was shown by Lount [74]. Choosing riskier actions as mood becomes more positive was shown by Leahy [65] and Hertel et al. [57]. Hertel et al. [57] and Aspinwall [4] showed that positive moods lead to more heuristic decision making, which can be quicker than decision making in a negative mood.

For neutral moods, the model does not impose any restrictions on the decision making process, as the model is currently not in either a negative or positive mood. Therefore the mood model should not influence the decision making process.

To summarise the model: when in a negative mood the decision should become more "rational" and considered as the mood becomes lower. When in a positive mood the decision should become more heuristic and idealistic in a global sense and the mood becomes more positive. When in a neutral mood the model should have no effect. There are a number of aspects to the model which are intentionally vague, such as how the mood changes over time. The only requirement the model enforces is that the mood is required to be a part of the mood value update process. The model leaves specific implementation detail up to the designer of the system, with the only requirement being that no part of the implementation contradicts the psychology literature. The lack of detail in the implementation is to ensure that the model can be applied to a wide range of applications. The next section will demonstrate the main implementation that will used through this thesis.

5.2 Mood Model Implementation

Now that the generic mood model has been defined, the implementation of the mood model that will be used in the experiments will be given, along with the psychological backing for the choices made. As per the generic mood model, the implementation will split the mood into three parts: negative, neutral, and positive. The implementation of the mood model will only affect the decision that an agent makes. The focus of the experiments is to analyse how the mood model affects what decisions are made, rather than simulating how mood can affect the agent physically. The implementation described focuses on how the model will affect decision making for an agent making decisions in a social dilemma, notably the Prisoner's Dilemma.

The generic mood model makes references to mania, and depression as examples of extreme positive and negative moods respectively. In this implementation positive and negative moods will be further split into two more categories, leading to the model having 5 possible mood states. These are; very positive, positive, neutral, negative, and very negative.

Starting with neutral moods, the mood model will not affect the agents' decision making.

The mood will affect how agents react to unknown opponents since they do not have any emotional attachment to them. When the mood levels are extreme they will override the current emotional decision. Having the implementation distinguish between opponents was done to represent that mood levels in humans do not necessarily reflect cooperation as a whole, but affect the choice made [74]. Therefore the implementation will use the underlying decision making process, which will be the emotional implementation as per Chapter 4 with each of the emotional characteristics shown in Table 4.1.

For negative moods the generic model requires that decisions should be "rational". Since the agent will be playing the Prisoner's Dilemma, negative moods will lead to defection, as this is the Nash equilibrium and can be considered the more rational decision. Negative moods are split into negative and very negative. The difference will be that agents in negative moods will defect with any new opponents, while very negative mood levels will lead to defection against all opponents.

Finally with positive moods the generic model requires that the decision made is idealistic without regards for issues to an individual. In the Prisoner's Dilemma the ideal outcome for a society of agents is mutual cooperation, therefore positive moods to lead to cooperation. Similar to negative moods, positive moods are split into positive and very positive. Agents in positive moods will cooperate with any new opponents, while agents in very positive moods will cooperate with all opponents.

Now that the decision making process is defined at each mood level, how the mood will be represented and how the mood levels relate to the representation needs to be given. Mood will be represented as a $\mathbb{R}$ strictly between the values of 0 and 100. This representation of a mood value allows for easy understanding with negative moods being lower numbers and higher numbers being more positive moods. Keeping the mood as a number also keeps the computation simple.

For a full implementation each mood state is defined as follows: a mood of below 10 is characterised as extremely negative, below 30 as negative, higher than 70 as positive and above 90 as extremely positive, and between 30 and 70 as neutral. These numbers were chosen to ensure that the mood level can move between each of the states reasonably quickly, while also being intuitive to recognise what the move level represents. The individual numbers are arbitrary, and the results will be affected when these change along with the payoff matrix. Equation 5.1 shows how the agent chooses an action based on the mood model with the simulated emotions. An initial action is the action an agent would take if the mood model or the emotional model is unable to define what the action should be, which should only happen with the first few interactions the agent makes. Definition 4 gives the algorithm for what action an agent using this model of mood will choose.

Definition 4. Let Ag be the set of all agents, with i and $j \in Ag$. Let t denote time. Let m_i^t return the mood of agent i at time t, in the range $]0, 100[$. Let $\eta_{i,j}$ return the number of interactions agent i as with agent j. Let I_i return the initial action of agent i. Let $E_{i,j}^t$ return the action that agent i would take against agent j based on i's simulated emotions, at time t.

$$Ac_{i,j}^t = \begin{cases} COOP, & \text{If } m_i^t > 90 \text{ or } (m_i^t > 70 \text{ and } \eta_{i,j} = 0) \\ DEFECT, & \text{If } m_i^t < 10 \text{ or } (m_i^t < 30 \text{ and } \eta_{i,j} = 0) \\ E_{i,j}^t, & \text{If } 30 <= m_i^t >= 70 \text{ and } \eta_{i,j} \neq 0 \\ I_i, & \text{Otherwise} \end{cases} \tag{5.1}$$

Now that each mood level is defined, how the mood value will change needs to be given. An agent's mood value will go up or down based on the difference between the payoff received and its average payoff, as this represents how well the agent thinks it has done in that game [45]. Then additionally the mood value will go up or down based on how the agent feels towards inequity between the average payoffs. The inequity aversion model *Homo Egualis* is used to represent inequity as a value [45]. The model requires an α and β, where α represents how much an agent cares when it is doing badly and β represents how much an agent cares when its opponent is doing badly. α and β need to be between [0,1]. To represent an idealistic situation where all agents care about their success just as much as their opponents' success $\alpha = \beta$. To distinguish between the α of the Homo Egualis model and the α which represents the learning rate in machine learning, the Homo Egualis α will be represented by Ξ in the remainder of the thesis.

The amount the agent cares is represented by applying the mood to the α value, such that higher moods give a lower α. This results in mood changes being larger when the mood is low. If the mood is low then the agent "thinks" it is doing poorly in the arena when compared to other agents, which is to represent the property that humans can care more about equality when doing poorly in society [45].

Definition 5. Let Ag be the set of all agents, with i and $j \in Ag$. Let t denote time. Let r_i^t return the payoff of agent i at time t. Let m_i^t return the mood of agent i at time t, in the range $]0, 100[$. Let μ_i^t denote the average payoff for agent i up to time t. Let F_i^t return the opponent of agent i at time t.

$$\begin{aligned} \Xi_i^t &= (100 - m_i^{t-1})/100 \\ \Omega_{i,j}^t &= \mu_i^t - \Xi_i^t \cdot \max(\mu_j^t - \mu_i^t, 0) - \Xi_i^t \cdot \max(\mu_i^t - \mu_j^t, 0) \\ m_i^t &= m_i^{t-1} + (r_i^t - \mu_i^{t-1}) + \Omega_{i,j}^{t-1} \text{ where } j = F_i^t \end{aligned} \tag{5.2}$$

Equation 5.2 shows how the Ξ value is calculated from the current mood of an agent; which places the mood value in the range of $]0, 1[$ so it can be used as the Ξ. For example a mood value of 75 will return an Ξ of 0.25. Ω in Equation 5.2 is the simplified version of the Homo Egualis function [51], as there are only two agents in a single interaction and $\Xi = \beta$. The equation gives us a numerical representation of inequity that the agent has for that interaction. Equation 5.2 shows the overall implementation of mood using the previous mood value, the average payoff, the received payoff, and the Homo Egualis function to update the mood value after an interaction with another agent. m gives us the current mood value of an agent. The mood will increase or decrease depending on the difference in the received payoff and the average payoff, meaning that the mood will increase when the agent is doing better than expected and decrease when it is doing worse than expected. The inclusion of Ω adjusts the amount the mood value changes based on how fair the agent thinks the result was.

An example of the mood value going down; Assume the mood value is 70. There is an interaction with an opponent and my payoff (r) is 0 and the opponents is 5. Assume my average payoff is currently $2(\mu_i)$. At this point the mood will be going down by 2, now the Homo Egualis equation will be calculated. $\Xi = 0.3$, assume the opponent has an average of 1. $2 - 0 - (0.3 * 1)$ This will lead to $\Omega = 1.7$. Therefore the mood only drops by 0.3, which reflects that people in higher moods care less for equality, with lower moods makes the mood drop further.

5.3 Mood Experiment

The model of mood has been defined, along with an agent implementation of the model for social dilemmas. The model of mood will be tested in a similar manner to the emotional agents in Chapter 4. The aim of this experiment is to show that these moody agents are able to sustain cooperation in self-play and can survive when faced with an invasion of defectors. The experiments will take place in the 5m^2 regular arena shown in Figure 4.6. Agents that

Table 5.1: Mood experiment scenarios showing as a percentage the different distributions of starting mood levels for the agents.

Scenario	Low Mood	Neutral Mood	High Mood
1	100	0	0
2	0	100	0
3	0	0	100
4	33	33	33
5	70	15	15
6	15	70	15
7	15	15	70

use the model of mood, (referred to as moody agents) will have equal distributions of initial actions and emotional characteristics.

5.3.1 Experiment Scenarios

The first experiment will explore how the evolution of cooperation is affected by differing initial mood levels. The initial level of mood will be categorised into three types, low, neutral and high where low has a mood level of 30, neutral is 50 and high is 70. There will be seven scenarios each with a different distribution of these levels among the agents which can be seen in Table 5.1.

Each of these scenarios will be run against a number of sub-scenarios. The sub-scenarios define how many agents will be in the arena, with a range from 45 to 144 agents, the details of the scenarios can be seen in Table 5.2. The number of agents is different to the previous experiments, as there is only one arena the number of agents can be increased. The increased number of agents ensures the prominence of the density effects for the mood experiment. The aim of the varying starting proportions of initial mood levels is to see how different moods affect cooperation and explore if the mood model allows cooperation to increase in the society of agents over time.

5.3.2 Results

Figure 5.1 shows us the percentage of cooperation between the agents after the given number of interactions for each scenario, with an extra scenario which excluded the mood model and only used the emotional strategy. The results given are quite intuitive: cooperation evolves

Table 5.2: Mood experiment sub-scenarios, showing the number of agents that will be simulated for each scenario. The number of emotional characteristics that will be represented is also given.

Sub-scenario	Number of agents	Number of each emotional Characteristic
1 - Low density	45	5
2 - Medium density	81	9
3 - High density	117	13
4 - Very High density	144	16

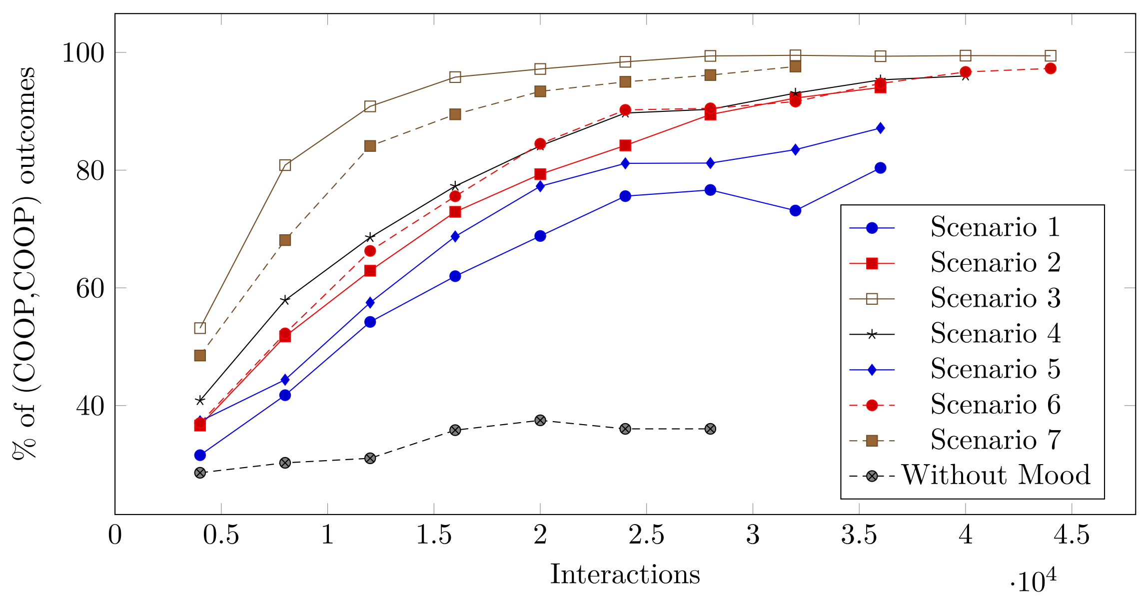


Figure 5.1: Percentage mutual cooperative outcomes over all runs for each scenario.

throughout the agents, and the speed at which this is achieved is directly proportional to the average level of mood. The fastest is the scenario with 100% of agents starting with positive mood levels and the slowest is the scenario with 100% of agents having negative mood levels. The increase in cooperation can be attributed to the mood model: looking at Table 4.16 in Chapter 4 shows how cooperation is stable over time. The only notable difference between the emotional agents from Chapter 4 and the moody agents in this chapter is the addition of the mood model.

The results show us that the mood model can support the evolution of cooperation over time and sustain cooperation; the result was expected as when cooperation is high the mood moves very little. Given two agents interacting in the Prisoner's Dilemma, with one being in a positive mood and one being in a negative mood, the negative mood will rise faster than the positive mood can go down, which is a property of the implementation of the Homo Egualis equation. Given that the mood levels are increasing overall, this leads to more agents in a cooperative state which in turn raises cooperation. The increasing mood values affecting cooperation are most apparent in the later stages of the simulation when the agents start with negative moods, as the agents which are cooperating meet a group of agents which are not cooperating. This led to a dip in cooperation followed by the continuing rise of cooperation when a large amount of agents with opposing moods meet.

To justify the claim that the speed at which cooperation is achieved is proportional to the starting level of mood, the average mood values against the number of mutual cooperative actions has been plotted in Figure 5.2. The results are plotted for Scenario 1 as this is where the effect is most pronounced; cooperation between agents does not rise as quickly since agents which are cooperating meet larger groups of defecting agents, which slows the increase in cooperation but increases the mood values. The two groups will later start cooperating, due to the increasing mood overall, which is shown through the decreasing standard deviation. Therefore the average level of mood reflects the average level of cooperation, and the higher the starting level of mood the faster the increase in cooperation is achieved.

5.3.3 Resilience Experiments and Results

The next experiment is to test the resilience of the cooperation that evolves over time. To test the resilience of the moody agents, pure defectors will be introduced into the arena at the beginning of the experiment; the pure defectors cannot replicate themselves but the

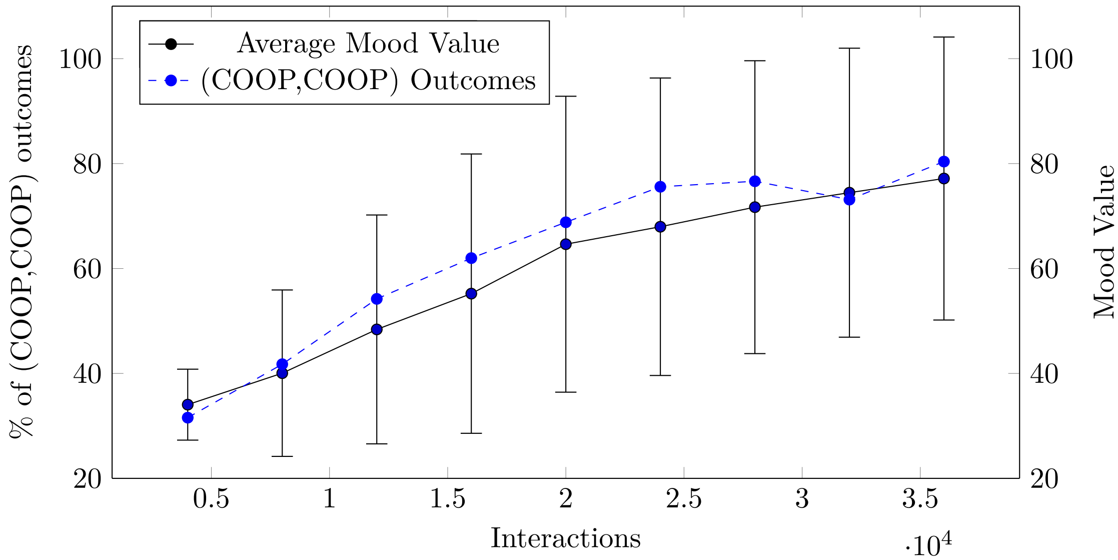


Figure 5.2: Average mood value with standard deviation against percentage of mutual cooperative outcomes in scenario 1.

Table 5.3: Resilience experiment scenarios.

Scenario	Mood Level	Number of Defectors	Total Agents
1	Low	43	106
2	Low	63	126
3	Low	83	146
4	Neutral	43	106
5	Neutral	63	126
6	Neutral	83	146
7	High	43	106
8	High	63	126
9	High	83	146

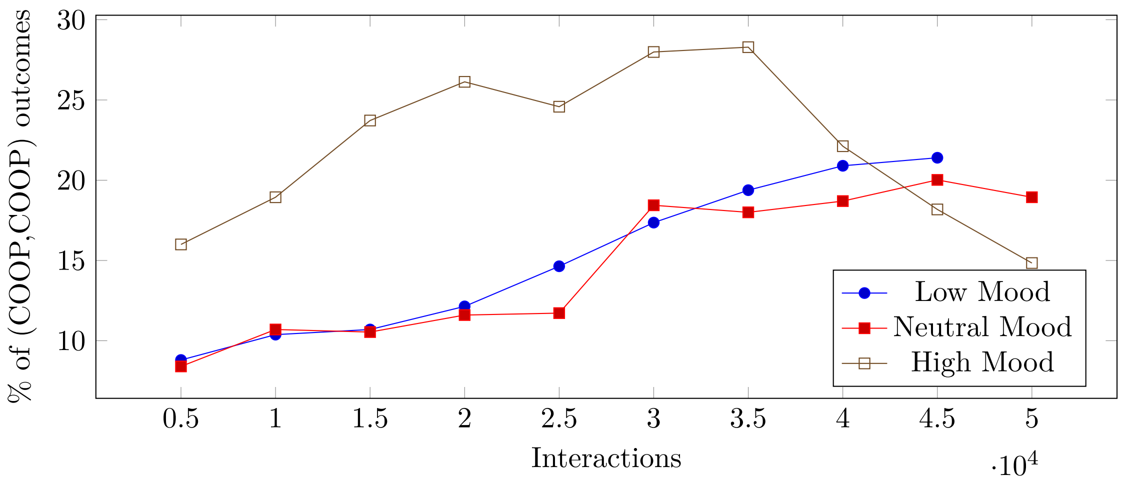


Figure 5.3: Percentage of mutual cooperation outcomes for each initial mood level.

emotional agents may take on the role of a pure defector due to their admiration emotion. Each scenario will have 63 agents whose initial mood is dictated by the scenario: the moods are categorised as high (70), neutral (50) and low (30). The numbers of pure defectors are 43 (minority defectors), 63 (equal defectors and emotional agents) and 83 (majority defectors). The details of each scenario are shown in Table 5.3. The aim of the experiment is to show the resilience the moody agents have to the invading defectors.

Figure 5.3 shows that when the mood is low, the emotional agents as a group are more resilient to an invading population of pure defectors. In positive moods cooperation between the emotional agents rises quickly, this in turn raises the mood of the agents as well. It therefore does not take long for the mood to increase to the point where the agents can be considered pure cooperators due to the mood level being very high. When this happens and agents are faced with the pure defectors, the only outcome between an emotional agent and a pure defector can be an asymmetrical outcome, with the pure defector defecting. The pure defector taking advantage causes the average score of the defectors to increase and the emotional agents' average to decrease. These changes in average payoffs will be affected rapidly because of the payoff difference. When replication occurs in the emotional agents they choose to become pure defectors because of the payoff difference, which leads to the collapse of cooperation as pure defectors become more prevalent.

In contrast when the emotional agents are in a negative mood, it takes longer for them to get their moods to the level where they are indistinguishable from pure cooperators;

Table 5.4: Average score (Standard Deviation) of the pure defectors, and the increase in payoff the pure defectors received when compared to the average payoff of the moody agents. Shown for each starting mood level .

Starting Mood Level	Average of Pure Defector	Payoff Increase
Low	1.45 (0.49)	0.04
Neutral	1.82 (0.69)	0.22
High	2.17 (0.85)	0.50

this allows them to protect themselves from the pure defectors by using their emotional choice, which switches their action to defection for that particular opponent. Actions driven by emotions rather than mood are bounded to a particular opponent, allowing the agents to evolve cooperation with other emotional agents without replicating into pure defectors since the defectors have a low average with the number of mutual defection outcomes they receive increasing over time.

These results show both expected and unexpected results. The expectation was that cooperation would continue to be stable over time as the simulated moods and emotions would adapt to the invasion force, as seen in the low and neutral starting moods. However the collapse of the positive mood was unexpected.

To explain why the positive mood levels have collapsed, it was shown that positive moods do not adapt quickly to the pure defectors and therefore are taken advantage of. The advantage taken then leads to the emotional agents becoming pure defectors as their average score is not high enough when compared to the pure defectors. To show why the moody agents that start in a positive mood have collapsed, Table 5.4 shows the increase in average payoff when compared to the moody agents. The increase in average payoff for the pure defectors when facing moody agents that start with positive mood levels more than double the increase when facing moody agents starting in neutral mood levels. The defectors are clearing taking advantage of the positive moods the most.

As the positive moods are being taken advantage of the most, the expectation is that the payoffs for the defectors should be the highest when faced with the highest mood. The average scores of the defectors are shown in Table 5.4 and clearly show that the defectors do the best when faced with positive moods, meaning that they will replicate the fastest in the positive moods. The neutral and negative moods do not collapse as they adapt to the newly replicated defectors through the use of their directed emotion strategy. The positive moods do not do this as when the mood is very high the emotional agents act as pure cooperators against the pure defectors.

The low and neutral moods did not collapse over the time that the simulation was run for. However if the experiment was conducted for longer periods of time, then the low and neutral moods may collapse since the agents' mood levels may reach the very high mood levels, which overrides the adaptation the agents created when facing the pure defectors.

5.4 Mood in Multiple Arenas

The first set of experiments that have been conducted have been simulated in one arena. The background chapters have shown that multiple arenas need to be identified and tested to ensure that the results are robust. In the next section the experiments that were conducted in the previous section will be conducted again, with the number of arenas increased. The new arenas will be the same as per the previous chapter. The aim is to show that the results of the previous experiment are valid for different arenas.

5.4.1 Experiment Outline

The experiment was conducted with the same scenarios, as shown in Table 5.1, however instead of the sub-scenarios from the previous section, the range of robots was 27 to 108 agents. The change was to keep the number of agents consistent between the mood and emotion experiments for accurate comparison. The details of the number of robots can be seen in Table 4.13. The admiration threshold for each agent was set to three (High) as per the previous experiment. The experiment was conducted in four arenas, shown in Figure 4.6. The hypothesis is that there will be little difference in the level of cooperation from the previous experiment. Additionally the mood will stop individual characteristics becoming dominant as the mood evens out the differences in average payoffs.

5.4.2 Results

Figure 5.4 shows us the percentage of cooperation between each minute for each scenario in the mood experiment. The results given are quite intuitive; cooperation evolves throughout the agent society, and the speed at which this is achieved is directly proportional to the average level of mood. The fastest is the scenario with 100% of agents starting with positive mood levels and the lowest is the scenario with 100% of agents having negative mood levels.

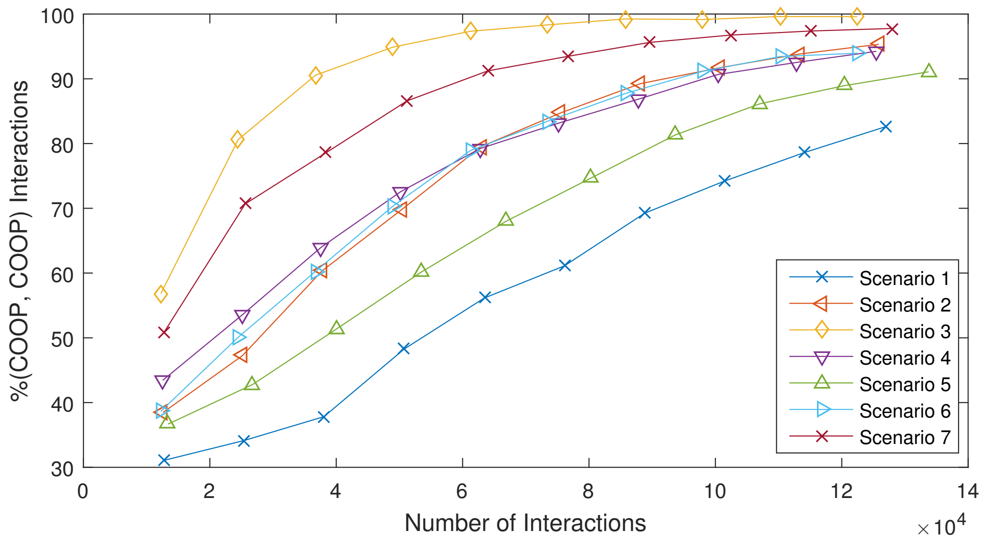


Figure 5.4: Average level of cooperation per scenario.

Cooperation raising over time can be attributed to the mood model as when compared to Figure 4.7 the cooperation only rises with the addition of mood, reflecting the results in the previous experiment, shown in Figure 5.1.

These results show us that the mood model can support the evolution of cooperation over time and sustain cooperation; this was an expected result as when cooperation is high the mood moves very little. The results of the experiment again show that when two agents play the game, with one being in a positive mood and one being in a negative mood, the negative mood will rise faster than the positive mood can go down which is a property of the implementation of the Homo Egualis equation. Having the average mood level increasing leads to more agents in a cooperative state, raising cooperation for the society. This effect is most apparent in scenarios where the agents start with negative moods, as there is a dip in cooperation followed by the continuing rise of cooperation when a large amount of agents with opposing moods meet.

To justify the claim that the speed at which cooperation is achieved is proportional to the starting level of mood, the average mood value is plotted against the number of mutual cooperation outcomes, as can be seen in Figure 5.5. The results are plotted against Scenario 1 as this is where the effect is most pronounced; When the cooperation between agents falls, the average mood level still rises. As cooperation rises the standard deviation

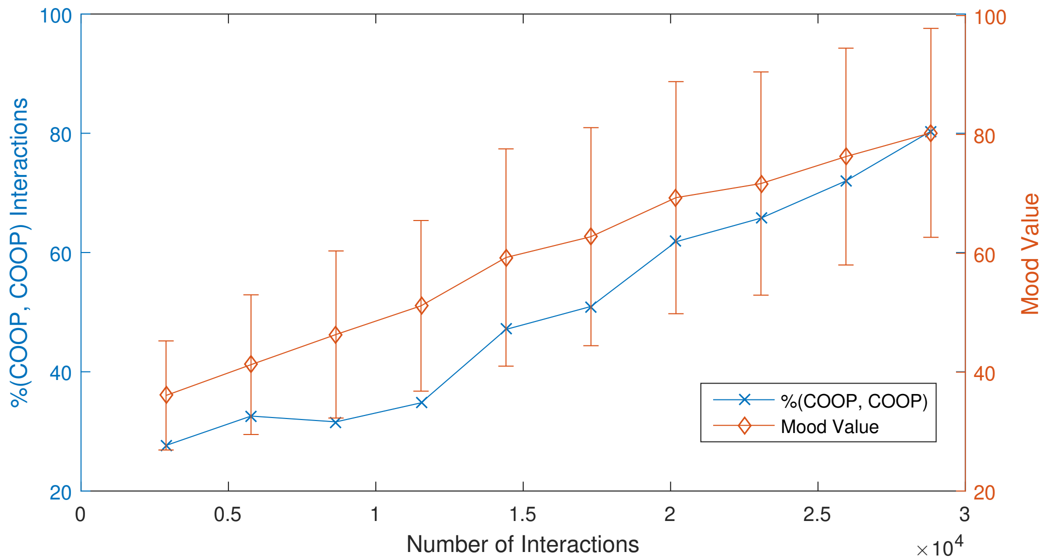


Figure 5.5: Level of cooperation for the regular arena in Scenario 1 against the average level of mood.

of mood levels gets wider but as the standard deviation gets smaller the cooperation still rises, showing us that the negative moods are rising more quickly than positive moods are lowering. The figure therefore shows us that the mood reflects the level of cooperation, and the higher the starting level of mood the faster cooperation is achieved.

There is little difference in the regular, small world, and empty arenas in terms of cooperation, as shown in Figure 5.6. However there is a very slight increase in the cooperation rate for the random arena; in that arena the structure has a tendency to constrain agents in small groups so the number of agents interacted with is slightly smaller.

To summarise, the experiment has shown that the previous work is valid across the given arenas and that the arena only has very minor effects on the level of cooperation. By grouping the agents, the level of cooperation increases faster.

Let us now consider which characteristics are most dominant in the mood experiment, as shown in Figure 5.7. The most notable difference is that in the mood experiments there are fewer games where there is a dominant characteristic. This was expected as the mood makes previous games affect the current game regardless of opponent, so the effect of the characteristic is reduced. The unexpected aspect is that there is a clear number of characteristics which are dominant. For the small world arena the Distrustful characteristic

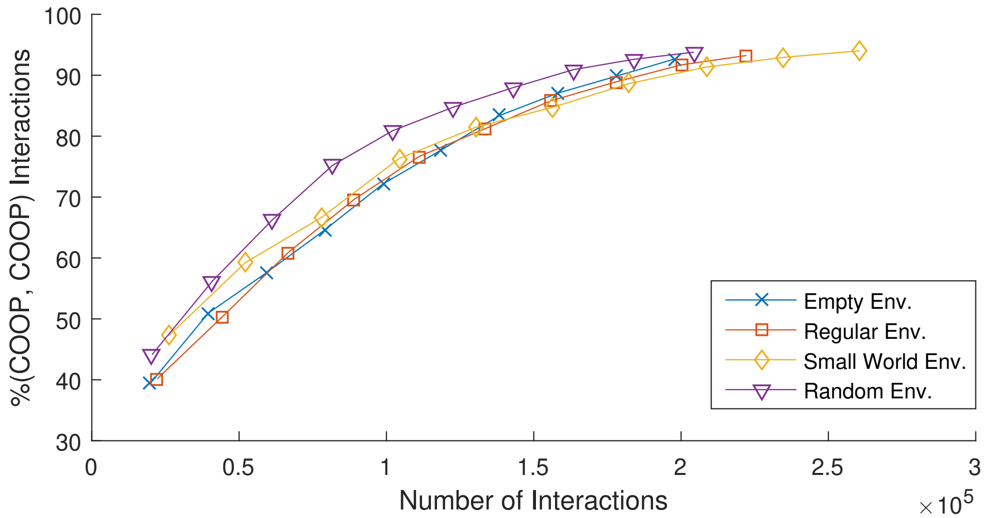


Figure 5.6: Level of cooperation for each arena.

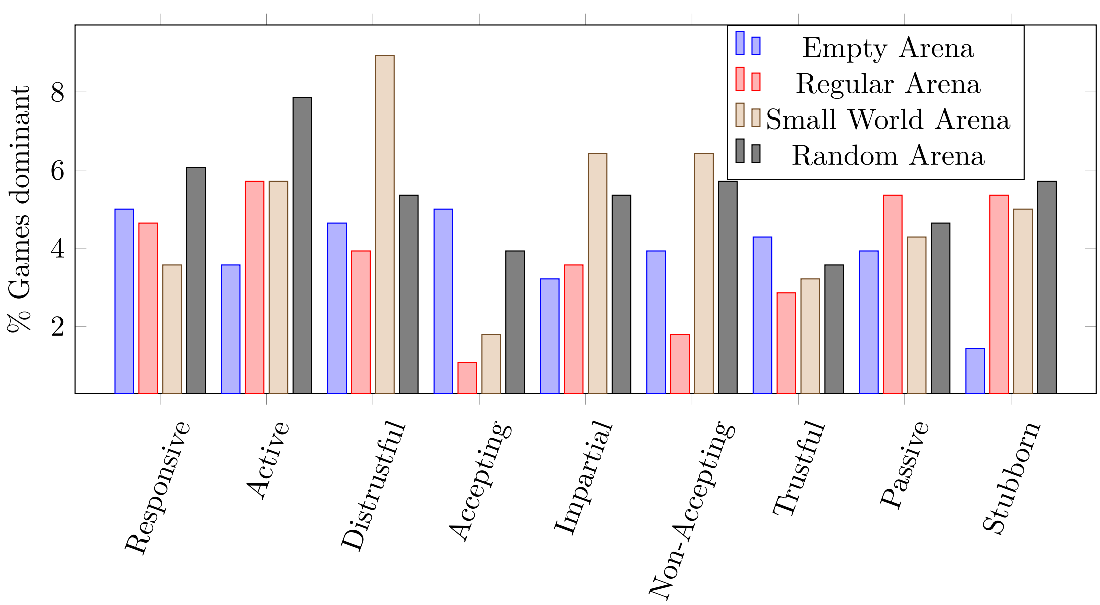


Figure 5.7: Most dominant characteristics.

is dominant, while in the random arena the Active characteristic is dominant.

For each arena different characteristics are dominant. In the more open arenas the games are closer with the characteristics becoming dominant that achieve cooperation quickly while protecting themselves from being taken advantage of. The reason that the agents need to protect themselves is due to the low number of times that agents meet with the same agent. The low number of interactions with a specific opponent requires the agent to protect its payoff quickly as it is unlikely to be able to punish this behaviour or force the cooperation to happen. Later in the experiment when the mood effects take place and cooperation is enforced, the difference comes at the beginning, where agents that protected their payoff do better, whether they took advantage of cooperators or cooperated to raise their payoffs.

When the arenas become more closed, the payoffs achieved by taking advantage early become more important, especially if the agents are able to move and interact with a wider range of agents more quickly. In the small world arena, the dominant characteristics take the most advantage of other characteristics with Distrustful being the most dominant as it protects its payoff the most, while punishing quickly. While the regular and small world arenas are similar, in the experiment the regular arena acts more open due to its larger corridors. Larger corridors make the successful characteristics closer in the number of games dominant, as in the empty arena. Dominant characteristics in the regular arena react quickly to defection as previously noted, however the regular arena also allows consistent interactions with the same agent. Agents that are taken advantage of by their opponents can still become dominant if they also take advantage of their opponents. This is seen by the success of the Passive, Stubborn, and Responsive characteristics.

In the random arena, the agents are more limited in the range of characteristics they can interact with. The random arena is effectively a closed off arena which allows the Active characteristic to become the most dominant by a wide margin. The advantage that can be achieved from defecting in the arena is reduced as the agent is likely to be punished since the chance that the agent meets the same agent again is heightened, however a small advantage can be taken provided that the agent protects the payoff quickly by reacting to this punishment.

As in the previous experiment, the distance travelled by an agent has an effect on the expected average payoff. The results are shown in Table 5.5. The results show the same effects as those in Table 4.15, with the larger the distance travelled the higher the payoff. The scores are higher overall for the moody agents due to the higher level of cooperation.

Table 5.5: Average payoffs (Standard Deviation) for an agent based on distance in the mood experiment.

Distance	84	108
high	2.64 (0.55)	2.71 (0.46)
medium	2.63 (0.58)	2.70 (0.43)
low	1.94 (1.32)	2.38 (1.03)

To summarise, the structure of an arena has an effect on which characteristics are the most successful for moody agents. The introduction of the mood model has allowed the agents to increase the level of cooperation over time in all the arenas.

5.4.3 Resilience Experiment and Results

The resilience experiment will be repeated to ensure that the results are applicable to all arenas. The hypothesis for this experiment is that the results will be similar to the previous experiment, with positive moods performing well, then collapsing, and negative moods being the most stable. Additionally negative moods will lose the least amount of agents to the defectors as their average scores were reported to be closest to the defectors which should prevent the replication happening. Under the same reasoning positive moods should lose the most agents.

Let us first consider how cooperation has been affected by the addition of pure defectors, shown in Figure 5.8. Similar effects can be seen when compared to the previous experiment, shown in Table 5.3. There the positive mood level rises then collapses as the defectors replicate and collapse, while for neutral and negative moods cooperation rises despite the addition of pure defectors.

The drop in cooperation varies by arena, with more open arenas showing less of a drop and more closed off arenas showing more of a drop. The random arena and the empty arena are plotted in Figure 5.8. In the random arena, the effect is most pronounced while in the empty arena the effect is least pronounced. The difference between the arenas is due to the different chances of meeting defectors. In the empty arena an agent will meet more agents, making the chance of meeting a pure defector with a higher average payoff less likely. However in the random arena where the agents are split into small groups the chance of meeting a pure defector with a higher payoff is increased.

The previous resilience experiment showed how positive mood levels collapsed due to

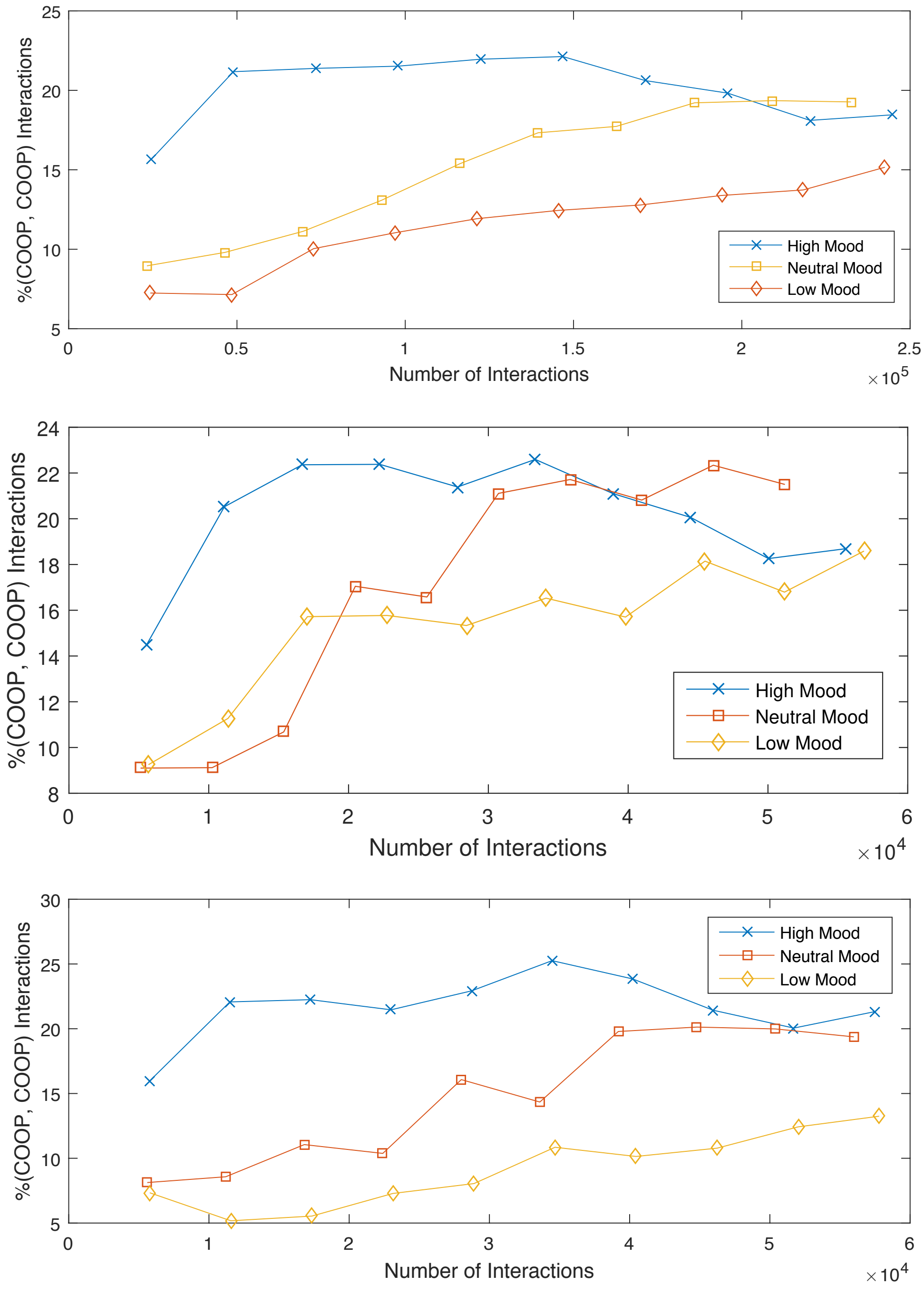


Figure 5.8: Levels of cooperation in the resilience experiment based on starting mood for all arenas (top), random arena only (middle), and empty arena only (bottom).

Table 5.6: Average payoffs (Standard Deviation) for agents by strategy in the resilience experiment.

	Low Mood	Neutral Mood	High Mood
Defectors	1.51 (0.49)	1.86 (0.64)	2.22 (0.81)
Emotional	1.45 (0.51)	1.57 (0.57)	1.64 (0.60)
Difference	0.06	0.29	0.58

Table 5.7: Average increase (Standard Deviation) in pure defectors for each mood level in the resilience experiment.

	Low Mood	Neutral Mood	High Mood
Increase	10 (10.22)	5 (2.89)	7 (3.23)

pure defectors taking advantage of them. Comparing the results to the previous experiment shows that the results are extremely similar. High moods do not adapt quickly to the pure defectors and therefore are taken advantage of. The advantage taken then leads to the emotional agents becoming pure defectors as their average score is not high enough when compared to the pure defectors. The difference between average score of the defectors and the average score of the emotional agents for each starting level of mood is shown in Table 5.6. This shows that the positive mood difference is more than double the neutral mood difference. The defectors are clearly taking advantage of the positive moods the most.

As the positive moods are being taken advantage of the most, the expectation is that the payoffs for the defectors should be the highest when faced with the highest mood. The average scores of the defectors are also shown in Table 5.6 which clearly shows that the defectors do the best when faced with positive moods, meaning that they will replicate the fastest in the positive mood scenarios. The neutral and negative moods do not collapse as they adapt to the newly replicated defectors through the use of their directed emotion strategy. The positive moods do not do this as when the mood is very high they act as pure cooperators.

Looking at the increase in defectors for each mood level, the expectation is that negative moods have the smallest increase and positive moods have the highest. The results from the experiment, as shown in Table 5.7, show an unexpected outcome: the highest increase in defectors is in the negative mood levels, while neutral and positive mood showed expected results. The reason why negative mood levels perform the worst is shown by the standard deviation. Low mood levels act closer to pure defectors which enables them to keep the payoffs of pure defectors low as they always defect so both agents will attain an average of 1. The difference comes when the negative moods attempt a cooperative action; if the negative mood agent attempts a cooperate action with a pure defector it raises the pure defector's average higher than the majority of the low agents. When the replication happens the pure defectors will always replicate, causing the high increase in pure defectors. However if the negative mood agents attempt a cooperative action with the other emotional agents such that the emotional agents start cooperating, their high average prevents pure defectors from replicating as they cannot get the advantage from any of the other agents. The result is a high average increase and high standard deviation for agents starting in low moods.

The neutral moods achieve the smallest average increase in pure defectors when compared to the other starting mood. The agents starting in a neutral mood achieve this as they are quick to adapt to the pure defectors as these agents are using the emotional aspect of their decision making. Using the underlying emotional decision making process is an advantage in getting cooperation between other moody agents since by using their emotional aspect of decision making they are more responsive to cooperation. This allows the neutral mood agents to increase their payoff between each other, which the pure defectors can not do and since the neutral moods have adapted to the pure defectors, they do not replicate as often. The positive moods act similarly to pure cooperators allowing the pure defectors to take advantage quickly as mentioned, leading to the higher increase in defectors when compared to the agents starting in neutral moods.

To summarise, high starting moods create cooperation the fastest but collapse over time. The rate of collapse differs depending on the arena, with more open arenas showing slower rates of cooperation collapsing. Agents starting in a low mood are both the best and the worst at protecting themselves against the invading strategy. In terms of the best, the low mood agents tries to cooperate with another low mood agent, which causes the low mood agent to minimise the number of new pure defectors. In terms of the worst, the low mood agent tries to cooperate with a pure defector, which causes the pure defectors to replicate rapidly.

5.5 Mood within Reinforcement Learning

This section shows how the mood model can be implemented without the underlying OCC model. To achieve this the mood model will be implemented with another decision making process and will show improvements when compared to the decision making process on its own. The experiment will use the reinforcement learning algorithm SARSA as the underlying decision making process. SARSA and the mood model will be tested in two social dilemmas; the Prisoner's Dilemma and the Stag Hunt. Success will be measured by the level of cooperation. The payoff matrix that is used for the Stag Hunt is given in Table 3.2.

5.5.1 Mood Model Integration

To integrate the mood model that has been defined at the start of this chapter into the SARSA algorithm (Definition 1 which is given in Chapter 3) the way in which the estimation of future rewards is computed needs to be changed. The estimation that will be calculated will take into account how high moods use more heuristic reasoning, while low moods will use a more rational approach. SARSA uses $Q(s_{t+1}, a_{t+1})$, the value of the next state-action pair, as an estimate. The mood model implementation will estimate Ψ based on mood, yielding the update rule:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma\Psi - Q(s_t, a_t)]. \tag{5.3}$$

Ψ returns the mean payoff received from an opponent over time to estimate what the future rewards will be from that opponent. The length of time that Ψ looks at is affected by the level of mood. Higher moods will look at fewer interactions, reflecting that high moods use a more instinctive response that is calculated quickly [57]. Lower moods will increase the number of interactions to look at, reflecting that lower moods lead to a more considered approach [57]. The maximum number of interactions that Ψ will look at is 20, to ensure that individual outcomes have an effect on the mean. The computation is formalised in Definition 6. For example if the mood level is 25 (low mood), then Ψ will calculate the mean of the previous 75% outcomes with that agent.

Definition 6 (Estimation of Future Rewards)**.** Let Mem_i^a be the set of rewards obtained by agent i when using action a where $|Mem_i^a|$ is at maximum 20, and $Mem_i^a(0)$ returns the most recent reward. Let m_i return the mood of agent i.

$$\begin{aligned} \Xi_i &= (100 - m_i)/100 \\ \zeta &= ceil(|Mem_i^a|/\Xi_i) \\ \Psi &= \left(n \sum_0^{\zeta} Mem_i^a(n) \right) / \zeta \end{aligned} \tag{5.4}$$

To choose which action to take, the mood model will use the ϵ-greedy method. This method selects the action with the highest Q value with probability $1 - \epsilon$, and a random other action with probability ϵ. In the implementation $\epsilon = 0.1$ initially. The mood value can also affect what the epsilon will be. Definition 7 gives the value of ϵ. The value of MA (Mood Affect) is defined in the experiment setup.

Definition 7 (ϵ calculation)**.** Let m return the mood value. Let Ac return the action that the agent would take. Let C be cooperation and D be defection. Let MA return the value of Mood Affect.

$$\epsilon = \begin{cases} 0.1 + MA, \text{ If } (m < 30 \text{ AND } Ac == C) \text{ OR } (m > 70 \text{ AND } Ac == D) \\ 0.1, \text{Otherwise} \end{cases} \tag{5.5}$$

In neutral moods there is no change in the value of epsilon. If the agent is in a bad mood ($m < 30$) and has chosen to cooperate there is an increase to the ϵ which reflects how humans are more likely to defect in these kinds of social dilemmas [53]. When the agent is in a good mood ($m > 70$) then the ϵ will increase as well, if the agent has initially chosen to defect, reflecting how people in good moods are more likely to choose an idealist option, even if that choice is risky, as discussed in [57, 65].

Finally, an important choice when applying reinforcement learning is how to initialise the state-action values $Q(s, a)$. In the implementation of the mood model with reinforcement learning the first reward that the agent receives for that state action pair will be set as the initial Q value, as this best reflects how people learn about new experiences. For example, [109] show that resetting the initial values to new data, Q-learning can predict how people choose between a risky option and a safe option.

There are some similarities between this implementation of mood model and reward shaping. Reward shaping supplies an additional reward to the reward function that would normally be received for a particular action [85]. For the SARSA algorithm the update rule will be updated to the following.

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + F(s_t, a_t, s_{t+1}) + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] \tag{5.6}$$

F returns a separate reward based on the state action pair and the next state. The reward is defined by the designer of the system. This technique can allow reinforcement learning to scale up to more complex problems [32, 92]. The main difference between reward shaping and the mood model is that the shaping reward function supplies additional rewards, whereas the mood model directly affects decision making and the estimation of future rewards.

5.5.2 Experiment Setup

The experiment will be conducted using the stage library [119] simulating 70 agents within the $5m^2$ regular arena, shown in Figure 4.1. For both SARSA and the moody agent $\alpha = 0.1$ and $\gamma = 0.95$.

The agents will engage in a random walk around the arena. When an agent has line of sight and is sufficiently close to another agent then those two agents will engage in an iteration of either the Prisoner's Dilemma or the Stag Hunt game depending on the scenario. Then they will both continue moving around the arena. This will continue for 10 minutes, then the arena will reset (to prevent agents getting stuck in corners of the arena) with the agents retaining their memory and Q values. This is repeated until they converge. Convergence is said to have occurred when the average proportions of outcomes from the last 5 × 10 minute runs are within 0.005 of the average proportions of the last 25 × 10 minute runs.

The experiment will be conducted over a number of different scenarios, which include variations on how much the mood value affects the choices made by the agents (which is indicated as MA) and can be seen in Table 5.8.

The experiment will start with setting MA to 0, to compare the differences between the mood model implementation and SARSA without the noise of the MA affecting decision making. The next set of scenarios increase the value of MA to see how much of mood needs to affect the decision making process so that the agents converge to cooperation. In Scenario 6, the value of MA is a variable amount v which is dependant on the mood, the exact calculation of v is shown in Definition 8. Finally Scenario 7 will show the results for SARSA.

Table 5.8: Scenarios with different values of MA.

Scenario	MA
1	0
2	0.2
3	0.4
4	0.6
5	0.8
6	v
7	SARSA

Table 5.9: State definitions

Name	$St =$	Description
Stateless	$\emptyset$	No state information used
Agent State	Ag	Agent as the state
Mood State	$Ag \times MV$	Agent and Mood level as the state

Definition 8 (Variable MA Value)**.** Let m_i^t return the mood of agent i at time t.

$$v = \begin{cases} (m_i^t - 50)/100 & \text{if } m_i^t > 70 \\ (50 - m_i^t)/100 & \text{if } m_i^t < 30 \\ 0.1 & \text{otherwise.} \end{cases} \tag{5.7}$$

For each of these scenarios compare three different definitions of the state space. The first is *Stateless*, where the agents have no knowledge of the arena or who their current opponent is when an interaction occurs. The second is *Agent State*, where the agents can distinguish between the opponents they are interacting with. Finally, the *Mood State* is where the agents additionally observe the mood their opponent is in. The definitions are given in Table 5.9 where Ag is the set of agents and $MV = \{High, Neutral, Low\}$ is the set of mood representations. Given the mood m_i of agent i, MV_i is *High* when $m_i > 70$, *Low* when $m_i < 30$, and *Neutral* otherwise.

The prediction is that the memory-only experiment will show a small amount of variation when compared to the SARSA algorithm as the only difference between them is how they predict future rewards ($H1$), leading to the same actions being selected in both cases.

In both of these cases the prediction is that both types of social dilemma converge to defection in a stateless scenario as the opponents faced will be randomised, preventing any cooperation from being sustained. When the agents are able to distinguish between opponents, the prediction is that some small levels of cooperation will appear that will then fade into defection as the decision making agent may randomly switch due to the ϵ greedy action choice ($H2$).

In regards to the other scenarios, the hypothesis is that high levels of MA will give higher levels of cooperation ($H3$); This prediction is based on how the mood model has increasing levels of cooperation over time in the previous experiments detailed in the previous sections. The levels of mood will increase with mutual defection, however when the agent starts to cooperate the mood will fall, causing the other agents' mood to go up. The rate at which the mood decreases in the original agent will be less than the rate at which it rises in the opponent, ending with the overall mood increasing, which in turn will lead to the agents using the higher ϵ when choosing to defect. $H1$, $H2$, and $H3$ refer to the hypotheses for this experiment.

5.5.3 Results

Understanding whether the introduction of this mood model has been a success, success first need to be defined. Success is defined in two ways. Whether mutual cooperation can be created and sustained is the first way. The second definition of success is to analyse the average payoffs of the agents, and determine whether moody agents are have a higher average level of cooperation when compared to SARSA.

Tables 5.10, 5.11, and 5.12 show the proportions of the different outcomes the agents converged to, along with their 99% confidence values for the *Stateless*, *Agent State*, and *Mood State* scenarios respectively.

SARSA shows a strong preference for cooperation in prior work [118]; in this experiment it shows a strong preference for defection. The reason for this disparity is due to a change from a two agent setting to a larger setting, which is reflected in the cooperation increasing when the agents are able to distinguish between different opponents. While change in the number of opponents represented by a state explains some of the differences, there are effects from allowing the agents to move and the resulting inconsistency in the number of interactions. Introducing movement into the experiment introduces randomness into when any two individual agents may interact. The randomness does not allow accurate

Table 5.10: Proportions of outcomes converged to with 99% confidence intervals, for each scenario and social dilemma, when no state information is used (*Stateless*).

Scenario	Game	Coop	Defect	Non Mutual
1	PD	0.084±0.007	0.486±0.015	0.431±0.011
1	SH	0.114±0.007	0.436±0.015	0.450±0.011
2	PD	0.144±0.009	0.386±0.014	0.470±0.010
2	SH	0.157±0.009	0.369±0.013	0.474±0.009
3	PD	0.297±0.010	0.197±0.008	0.506±0.008
3	SH	0.313±0.011	0.207±0.010	0.481±0.008
4	PD	0.499±0.010	0.089±0.004	0.412±0.008
4	SH	0.804±0.006	0.010±0.001	0.186±0.006
5	PD	0.789±0.005	0.014±0.001	0.197±0.005
5	SH	0.809±0.004	0.010±0.001	0.181±0.004
6	PD	0.366±0.008	0.149±0.006	0.484±0.006
6	SH	0.634±0.026	0.060±0.009	0.306±0.019
7	PD	0.017±0.002	0.765±0.007	0.218±0.007
7	SH	0.028±0.006	0.735±0.018	0.236±0.012

Table 5.11: Proportions of outcomes converged to with 99% confidence intervals, for each scenario and social dilemma, using the *Agent State*.

Scenario	Game	Coop	Defect	Non Mutual
1	PD	0.211±0.007	0.498±0.011	0.290±0.011
1	SH	0.216±0.007	0.499±0.010	0.285±0.009
2	PD	0.247±0.009	0.393±0.010	0.361±0.008
2	SH	0.230±0.007	0.384±0.011	0.387±0.010
3	PD	0.321±0.008	0.243±0.007	0.436±0.007
3	SH	0.338±0.008	0.222±0.007	0.440±0.008
4	PD	0.427±0.009	0.143±0.006	0.431±0.007
4	SH	0.463±0.010	0.123±0.006	0.414±0.009
5	PD	0.632±0.012	0.060±0.005	0.308±0.009
5	SH	0.632±0.011	0.058±0.004	0.310±0.009
6	PD	0.361±0.009	0.193±0.007	0.446±0.007
6	SH	0.376±0.010	0.190±0.007	0.433±0.007
7	PD	0.211±0.007	0.497±0.011	0.293±0.009
7	SH	0.220±0.007	0.494±0.011	0.286±0.010

Table 5.12: Proportions of outcomes converged to with 99% confidence intervals, for each scenario and social dilemma, using the *Mood State*.

State	Game	Coop	Defect	Non Mutual
1	PD	0.213±0.006	0.484±0.012	0.303±0.010
1	SH	0.216±0.008	0.481±0.011	0.302±0.009
2	PD	0.246±0.007	0.379±0.008	0.375±0.008
2	SH	0.236±0.007	0.384±0.010	0.380±0.008
3	PD	0.314±0.008	0.244±0.006	0.442±0.008
3	SH	0.319±0.008	0.234±0.005	0.447±0.007
4	PD	0.454±0.009	0.135±0.007	0.411±0.007
4	SH	0.481±0.011	0.110±0.006	0.409±0.009
5	PD	0.623±0.011	0.066±0.005	0.311±0.007
5	SH	0.627±0.012	0.061±0.005	0.312±0.009
6	PD	0.365±0.008	0.194±0.005	0.441±0.008
6	SH	0.371±0.009	0.189±0.006	0.440±0.008
7	PD	0.211±0.007	0.483±0.012	0.306±0.010
7	SH	0.213±0.007	0.495±0.011	0.292±0.010

predictions on who the next opponent will be, or whether any two particular agents will converge in their pairwise interactions.

By comparing the SARSA (Scenario 7) outcomes to the memory only outcomes (Scenario 1), there is a small improvement to the memory only outcomes when the agents are anonymous, which is in contrast with the hypothesis H1. The difference in the improvements is down to how SARSA predicts future outcomes based on its current Q value, which takes into account all previous interactions, whereas the mood agents uses a limited memory of recent outcomes which allows them to adapt to the prevailing action consensus quicker than the SARSA agents.

Similarly, for the scenarios tested, the addition of states allows cooperation to increase, while also introducing larger amounts of non mutual actions and reducing mutual defection. This is in contrast to the hypothesis $H2$ which stated that any cooperation created would fall. However there was an exception with larger values of MA, where cooperation decreases with the addition of states. The addition of states increases the instability of the system, if the value of MA is high then the chance of an agent defecting is reduced to the point where cooperation spreads more effectively than defection. The reduction of information allows this cooperation to spread as agents try to converge on the group of agents as a whole rather than on an individual level. When MA values are low, then the additional information helps to prevent the spread of defection.

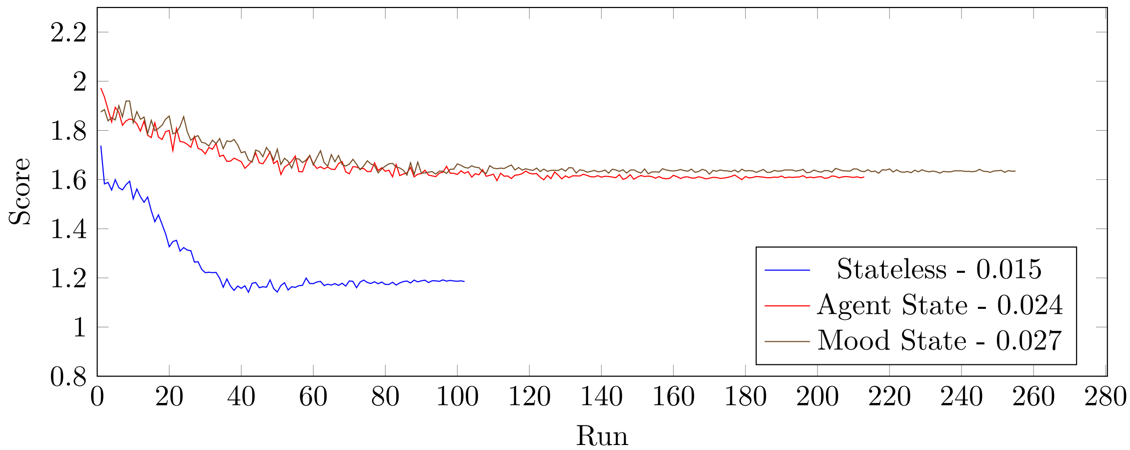
Hypothesis $H3$ was confirmed as higher levels of MA increase the level of cooperation. The differences between the level of cooperation in the Stag Hunt and Prisoner's Dilemma show that lower levels of MA are required to yield a high proportion of cooperation in the Stag Hunt. The difference is due to the payoff structure (Tables 3.1 and 3.2), as in the Stag Hunt mutual cooperation gives a higher payoff than the non mutual action does for the defector. The value of MA gives a higher guarantee that cooperation will be mutual so it is in the interest of the agent to cooperate, which is reflected in its Q values, whereas in the Prisoner's Dilemma the temptation to defect is still there as the individual payoff for defecting is higher than mutual cooperation.

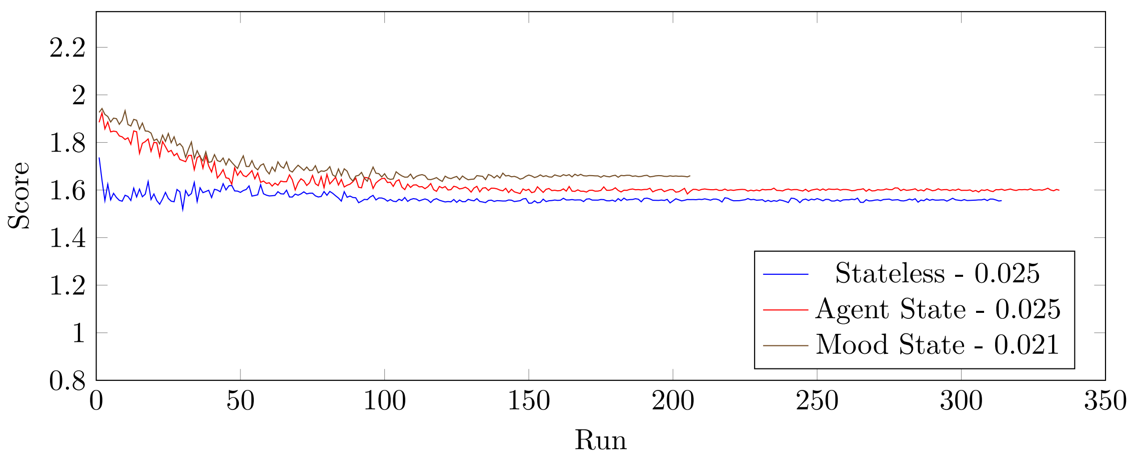
Finally the moody agents can achieve high levels of cooperation in the Stag Hunt and a majority of cooperation in the Prisoner's Dilemma. Next is the analysis of the second way to measure success, which is through the payoffs received by the agents. Figure 5.9 shows the average score of an agent through each run, for each type of state definition. Average scores were chosen rather than total scores as the number of interactions per run is not consistent across agents. The number of runs differs due to the time needed to converge being different for each state.

From these three figures the most successful agents were in the mood scenario, which is reflected by the higher levels of cooperation as noted previously. There are only minor differences between the different types of state. An exception is shown in the stateless scenario for SARSA, where the average score is much lower when compared to the agent and mood states, which is reflected by the lower cooperation levels.

When the social dilemma is changed to the Stag Hunt as shown in Figure 5.10 the same scenarios are the most successful in terms of average score. Variable mood shows an exception, here stateless is the most successful by a large margin. To see why this is the case, the payoff matrix of the games, how actions are chosen, and how the actions converge for the moody agents, need to be analysed.

The main difference in individual payoffs between the Stag Hunt and the Prisoner's Dilemma is that in the Stag Hunt the payoff for defecting against a cooperating agent is lower than the individual payoff for mutual cooperation. In the Prisoner's Dilemma this individual payoff is higher when defecting against a cooperating opponent. A perfectly rational agent will therefore choose to cooperate in the Stag Hunt if they know the other opponent is cooperating. However in the scenario choosing cooperation is not guaranteed by





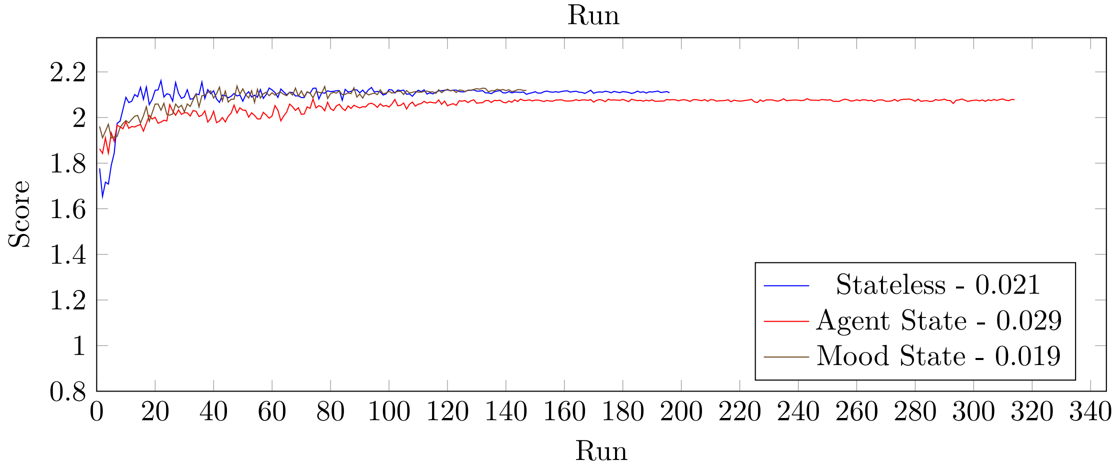


Figure 5.9: Average scores for SARSA (top), memory only (middle), variable mood (bottom) in the Prisoner's Dilemma over each run, with 99% confidence interval of final average score.

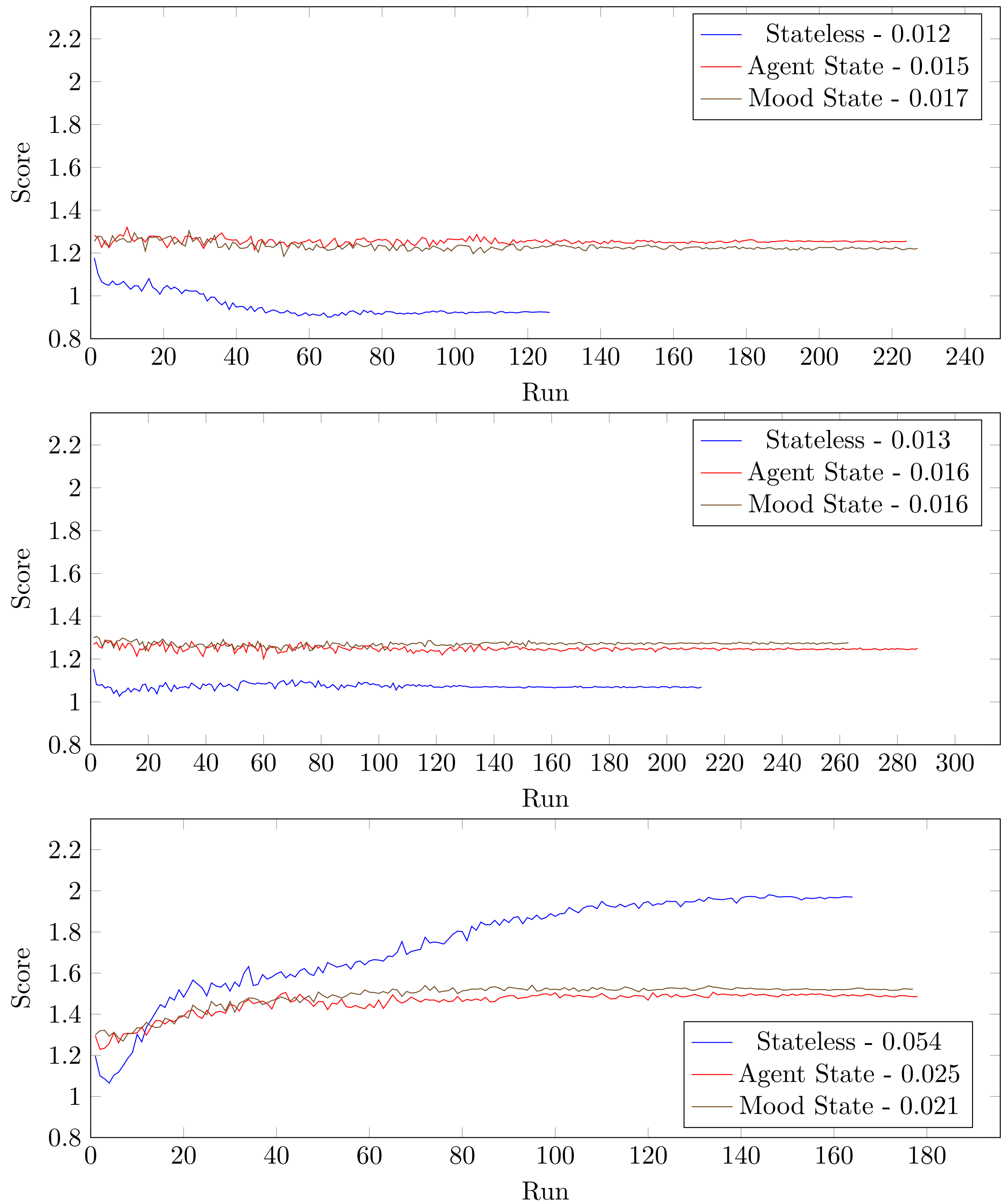


Figure 5.10: Average scores for SARSA (top), memory only (middle), and variable mood (bottom) in the Stag Hunt over each run, with 99% confidence interval of final average score.

the ϵ-greedy choice; when an agent inadvertently defects it receives the temptation payoff. In the Prisoner's Dilemma the agent will then continue defecting which leads to the drop in average score as more agents choose defection, as their Q values reflect the higher value that defection brings. In the Stag Hunt, agents will continue to choose cooperation since the Q value for defection does not rise higher than the Q value for cooperation.

When the agents can differentiate between opponents, the agent that receives the defect payoff will respond to the defecting agent at an individual level, rather than defecting against all other opponents. This prevents the spread of defection in the Prisoner's Dilemma, as shown by the average scores being higher in the agent and mood states when compared to the stateless in the memory only and SARSA scenarios. However in the Stag Hunt the cooperation has not spread for the same reason: cooperation needs to be created on an individual level first, which is shown by the weaker agent state and mood state average payoffs.

When using the mood scenario, the stateless scenarios perform just as well as the agent and mood states in contrast to the memory only and SARSA scenarios. The reason that defection does not spread is due to the effect mood has on action selection. There is an initial drop in average payoff as defection spreads, however as the mood is going up due to the agents receiving some payoff, the chances of an agent inadvertently choosing cooperation increases, which also causes more mutual cooperation, raising the moods of those agents even more making them even more unlikely to chose defection. The result is that the moody agents are able to break continued mutual defection and can replace mutual defection with sustained mutual cooperation over time.

5.6 Conclusion

To summarise, there is a proposed model of mood that can be used either independently or in conjunction with emotions. There is also a construction of this model using psychological research, along with identified general cases of outcomes that can be expected. The implementation was then applied to an experiment with validation from psychology research for the choices made, which explores cooperation in a multi-agent setting.

The experiments conducted showed that a combination of mood and emotion can support positive levels of cooperation within an agent society. The results show that mood levels in the agents are related to the level of cooperation that is achieved as a group. By adding an invasion force of pure defectors the cooperation between the emotional agents collapses over time when the mood levels of the emotional agents are high. In contrast when the mood is not high, the cooperation over time is more stable since the agents do not give the benefit of doubt to the pure defectors, preventing the defectors from achieving a higher average.

A second implementation of the model of mood used a classic reinforcement learning algorithm rather than the OCC model. This was evaluated extensively in an experimental setting using the Prisoner's Dilemma and Stag Hunt scenarios. The results were compared to the results that were produced by a SARSA implementation to investigate whether there was any improvement. The experiment used a number of scenarios that varied the amount of information that was available to an agent in addition to varying the strength of the mood model. Improvement was measured in two different ways, namely proportion of cooperation and average reward received by an agent per interaction. In contrast to previous work the experiment investigated scenarios which allowed agents to be mobile, introducing uncertainty in the interactions. The study is of use to designers of agent societies, by showing how mobility and mood affect different strategies of the agents.

By incorporating mood in reinforcement learning there was an increase in the level of cooperation when compared to SARSA, a decrease in mutual defection, and an increase in asymmetrical actions. Additionally the experiment compared the averages payoff of both the SARSA and moody agents showing that using mood increased the average payoff when compared to SARSA. There were differences between the two social dilemmas in regards to how effective the mood model was. Higher levels of cooperation where shown in the Stag Hunt when compared to the Prisoner's Dilemma, due to the payoffs of the game. Additionally there was a notable reduction in cooperation SARSA had when compared to prior research; the reduction in cooperation is mainly due to the mobility aspect that was captured.

This chapter has contributed to answering the research question posed. Concerning **SRQ1** which questioned how the implemented simulated emotions and mood has been answered in terms of mood, with the generic mood model given and the two different implementations given. The model and the implementations have been justified using the psychology literature to ensure validity. **SRQ2**, which asks how to capture cooperation for these agents, has been addressed by the experiments conducted, showing how success can be measured by both levels of cooperation and average payoffs.

SRQ3 asks how mobility affects the agents has also been addressed, by showing the differences the structure has on the moody agents, namely that the effects shown in the previous chapter are valid when the decision making process of the agents has changed. There is also a contribution by showing that mobility has an effect on SARSA which puts pressure on the level of cooperation.

For the broader research question of the thesis, this chapter has shown that agents that use the model of mood are able to evolve cooperation and can sustain the cooperation over time, showing that moody agents can be cooperative.

Chapter 6

Mobility and Interaction Topologies

The previous chapter has presented an implementation and analysis of simulated emotional and moody agents. The emotional and moody agents have been shown to be affected by the structure of the arenas when engaging in self-play in the Prisoner's Dilemma. The aim of this chapter is to generalise these results across both strategies and social dilemmas. The results will contribute towards **SRQ3**, which focuses on analysing mobility effects that occur when agents are able to move.

The chapter will start with the experiment setup in Section 6.1, with the results and analysis in Section 6.2 and the conclusion of the chapter in Section 6.3. This chapter gives a greater understanding of how the environment topology, along with mobility, affect social dilemma strategies. Throughout the chapter there will be references to static and mobile agents. A static agent refers to an agent which has a fixed number of opponents and does not move throughout the environment, which is modelled as a static network. A mobile agent refers to an agent that moves throughout the environment, modelled as an arena, and whose opponents will change over time. The work in this chapter has been published in [26].

6.1 Experiment Setup

The aim of this section is to give the outline of the experiment that will be conducted throughout the chapter. The experiment will have some similarities to the experiments from the previous chapter, but the aims differ significantly. The main aim of the experiment is to generalise the mobile and structural effects seen in the previous chapters.

This section will start by defining the strategies that will be used in the experiment. Next the arenas and network will be defined and how the networks will be constructed, along with the scenarios that the experiment will use. Finally the hypothesis of the experiment will be given.

6.1.1 Strategies

In this experiment there will be a number of different strategies that will be used. These can be split into two distinct types of strategies. The first type to define is the fixed strategies, which are deterministic strategies. The second type to define is the adaptive strategies which use reinforcement learning as part of their strategy. The fixed strategies that will be used in the experiment are:

Tit-For-Tat (TFT) Initially cooperates, then copies the opponent's last action [5];

Win-Stay Lose-Shift (WSLS) Initially cooperates, repeating the current action as long as it receives the highest payoff possible [86];

Random Cooperates or defects with equal chance;

Always Cooperate (ALL COOP) Always cooperates;

Always Defect (ALL DEFECT) Always defects;

Emotional - Active Shown in Chapter 4 to be the most effective emotional strategy in an arena. The strategy will switch to cooperation when the opponent has cooperated twice in a row, and switch to defection when the opponent defects. The strategy was named E2 in previous works [71, 70];

Emotional - Trustful Identified as the most effective emotional strategy in a network [70, 71], named E7 in that work. The strategy will switch to cooperation when the opponent cooperates, and switch to defection when the opponent defects three times in a row.

For the experiment, the state space for the reinforcement learning strategies is the set of opponents, with a given state being an agent. The two adaptive strategies used in this experiment are:

SARSA This is an on-policy reinforcement learning algorithm [115]. Definition 1 given in Chapter 3 shows the Q-Value update algorithm.

Moody This is also a reinforcement learning algorithm that uses a model of mood in both its action selection and estimation of future rewards as shown in Chapter 5. The definition is given in Section 5.2.

For the emotional strategies the initial action will be split equally between cooperation and defection. For the adaptive strategies the initial action will have an equal chance of cooperation or defection. The adaptive strategies' initial Q value for an action will be set to the first payoff an agent receives for that state action pair. The reason to set the initial Q value to the outcome is because Shteingart et al. [109] show that this best reflects how people learn, when the action to take is a choice between a risky or safe option. The generic mood model as given in Section 5.1 requires that reflecting psychology is a requirement for the moody strategy and applying the same Q value initialisation to SARSA allows a meaningful comparison between them.

In the experiment the action selection strategy of the adaptive action will be the ϵ-greedy method. The ϵ-greedy method selects a random action with probability ϵ, and the action with the highest Q-Value with probability $1 - \epsilon$. $\epsilon = 0.1$ for both adaptive agents. The learning rate is given as $\alpha = 0.1$ and the discount rate is given as $\gamma = 0.95$.

For the moody agent, there is an alteration to the previously defined algorithm. The aim of this alteration is to make the mood level reduce quicker when a poor outcome is obtained at high mood levels. This is achieved by altering the mood by the difference between the reward and the perceived average. Previously the reward was altered by the difference between the perceived average and the actual average. The new mood update step is shown in Definition 9, which shows how the mood level goes up by the size of the payoff, including any adjustments that the Homo Egualis equation [45] makes to the agent's perception of the reward.

Definition 9. Let Ag be the set of agents where $i, j \in Ag$ where j is the opponent, m_i be the mood of agent i, and t denotes time. r_i^t denotes the payoff of agent i at time t. $\Omega_{i,j}^t$ denotes the Homo Egualis equation [45] for agents' i and j at time t.

$$m_i^t = m_i^{t-1} + (r_i^{t-1} - \Omega_{i,j}^{t-1}) \tag{6.1}$$

The agents will be able to differentiate between opponents, each strategy will be applied to each agent individually and for the adaptive agents the opponent will represent the state.

6.1.2 Networks and Arenas

There will be a number of different interaction topologies in this experiment, where the topologies are modelled as networks. Each node in a given network represents an agent with the edges of the network allowing interactions to take place between the two connected agent nodes. In order to make a meaningful analysis between mobile arenas and networked interactions, the degree of the constructed network will be the effective degree of the mobile arena. The effective degree of a mobile arena is the average number of unique opponent that the agents face in the arena.

The networks that will be used have different structural properties, the structures that will be used in this experiment are:

Small world [125] Networks with high clustering and small characteristic path length, intuitively this is a network where the nodes have very few neighbours but the distance between any two given nodes is also small.

Fully connected Every node is connected to every other node.

Random The edges of the network are randomly distributed.

Regular All nodes in the network have the same degree. In this experiment the regular networks are random regular networks, where the connections to specific nodes may differ, but the degree of each node will be the same, allowing a meaningful comparison with the regular arena.

Examples of the networks are shown in Figure 6.1. A fully connected network is equivalent to an empty environment (no blocks), as the agents have an equal chance of meeting any other agent.

In the random environment 20 blocks are placed randomly around the environment, leaving 36% of the environment available for movement. For the static equivalent the random network will be constructed using the Erdős-Rényi method [39] to generate a random network with a component of one, where each edge has a 36% chance of being generated. The component of one ensures that there exists a path between all nodes.

To calculate the degree needed to generate the random regular network, the average number of unique opponents faced by an agent in the mobile arena is obtained. To ensure

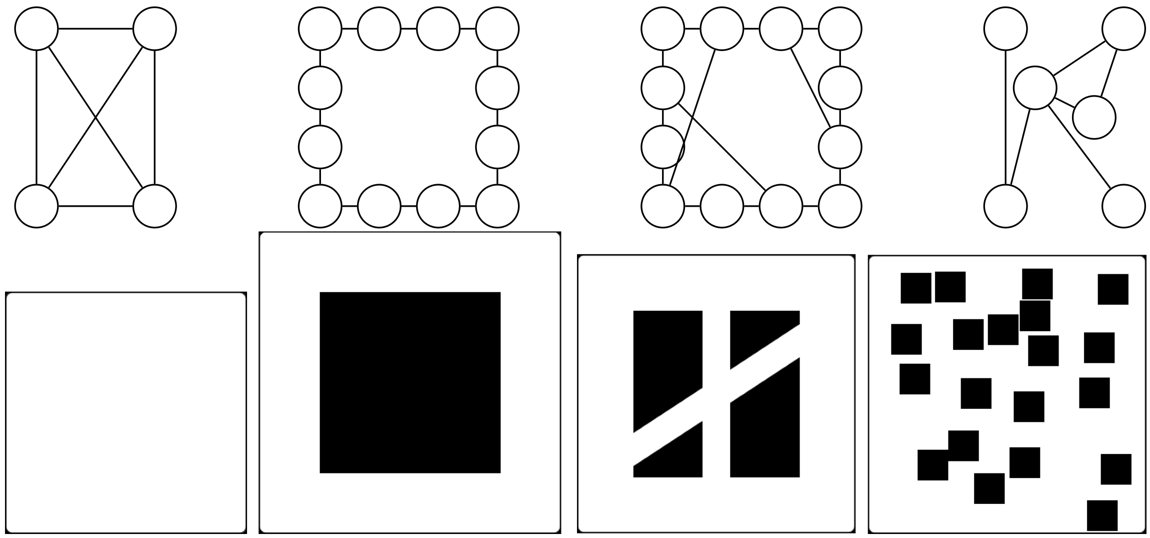


Figure 6.1: Fully connected, Regular, Small World, and Random environments (left to right). Example networks (top) and corresponding arenas (bottom)

that this number is an accurate representation, the number of unique opponents will exclude any opponents who are faced once only. The algorithm that is used to generate the random regular network is described by Kim and Vu [64]. The algorithm generates a random regular network in a relatively quick time based on the Steger and Wormald [113] algorithm. Additionally the experiment will keep constructing the network until the graph has a component of one.

Using the method for obtaining the degree as before, constructing the random small world network will use the Watts-Strogatz method [125]. This is constructed as a ring network where the edges in the graph have a rewiring probability of 40%. That is, each connection has a 40% chance of changing to a different agent. Examples of each of the static networks are given in Figure 6.1, where a 0% rewiring probability is a ring network and a 100% rewiring probability is a random network, with any number between producing a small world network.

6.1.3 Simulation Outline

The experiment will use 108 agents, with each agent using a single strategy and each strategy will be represented equally. The agents' initial positions are randomised for each run, in both the network and the mobile arenas. In the arenas, shown in Figure 6.1, the agents move randomly around the environment. The agents have basic obstacle avoidance and generate a random heading between -45 degrees and 45 degrees each second to allow a random walk. An interaction occurs when two agents are facing each other and are closer than 20cm. Interactions may happen after each second, after which they are given two seconds where they may not interact. This prevents agents having more than one interaction while they are passing each other. The arenas will be simulated using e-pucks [83] in Stage [119]. The mobile agents will interact for 20 minutes, then the positions will be re-initialised while the agents will retain any knowledge they have accumulated, to allow agents sufficient chance to meet a range of opponents. The simulation will be stopped once the agents have converged. To calculate if convergence has occurred, the proportions of mutual cooperation, mutual defection, and non-mutual outcomes of the 30 most recent 20-minute runs is taken. These numbers will then be compared to the same numbers from the 10 most recent 20-minute runs. Convergence is said to have occurred if the difference between the proportions calculated is within 0.005 of each other. The simulation will be conducted 50 times in order to generate an accurate result on what proportions of actions the agents converge on.

For the networked experiments, the agents' positions are randomised in the network. As time has no true meaning in the network, an agent will interact with every neighbour the average number of times the mobile equivalent interacts with a specific agent. For example, if in the arena an agent will average 3 interactions with a specific opponent, then in the networked environment the agent will interact with each neighbouring node 3 times. This counts as a single run, convergence in the network simulation will be based on the same properties as the mobile arenas.

6.1.4 Hypotheses

The first hypothesis predicts that networks will be more successful in supporting cooperation when compared to the arenas; this is hypothesis 1 ($H1$). This prediction is due to agents being able to retaliate against exploitative agents in a reliable manner, ensuring that cooperative agents will both receive a high average payoff as a pair. In arenas there is no guarantee that agents will meet the same agent more than once, allowing exploitative agents to be successful, as the opponent is unable to retaliate. The inability to guarantee retaliation may lead to successful arena strategies being those which take advantage of agents met rarely while cooperating with agents met frequently. The TFT and Trustful

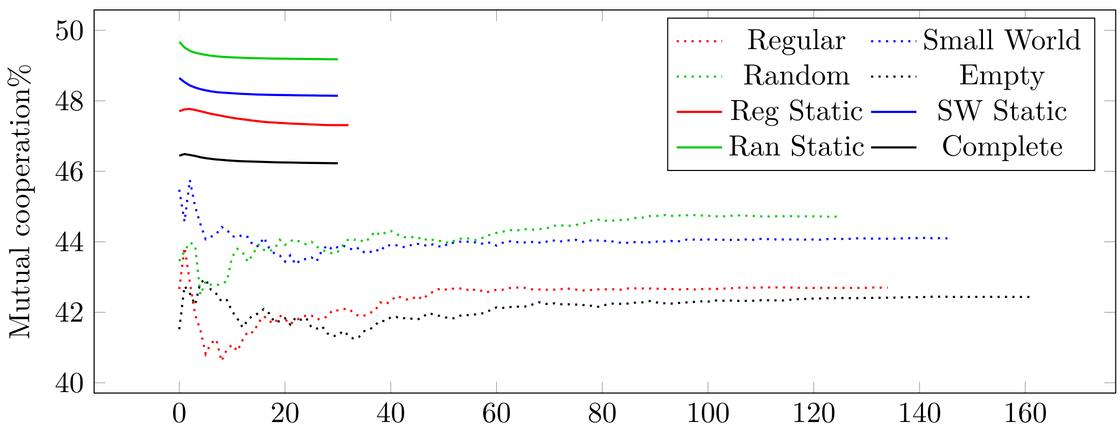


Figure 6.2: Proportion of mutual cooperation for each run, in each of the environments, for the agents playing the Prisoner's Dilemma.

strategies are effective in maintaining cooperation over time, and depending on their initial action can be effective in taking advantage in a one-shot interaction.

From what has been seen in the previous chapters there is the expectation that there will be differences between the environments in terms of which environments support the highest levels of payoff. The previous chapters have shown that high levels of mobility are a factor in supporting higher levels of average payoff in arenas with mobile agents. That work leads to the expectation that the empty environment will support the highest levels of average payoff; this is hypothesis 2 ($H2$). While the empty arena is represented by the fully connected network, the expectation is that the full connected network will achieve the lowest amount of cooperation and therefore payoff, as Lieberman et al. [68] has shown for networked interactions; this is hypothesis 3 ($H3$).

6.2 Results and Analysis

Examining the level of cooperation achieved by the society in each environment, Figure 6.2 show the results for the Prisoner's Dilemma and Figure 6.3 shows the results for the Stag Hunt. The level of cooperation is given as a percentage where 100% represents that every outcome was mutual cooperation for that particular run.

These figures support the hypothesis $H3$, as in both the Prisoner's Dilemma and Stag Hunt the network with the least amount of cooperation was the Fully Connected network. The level of cooperation increases with the level of randomness in the network construction,

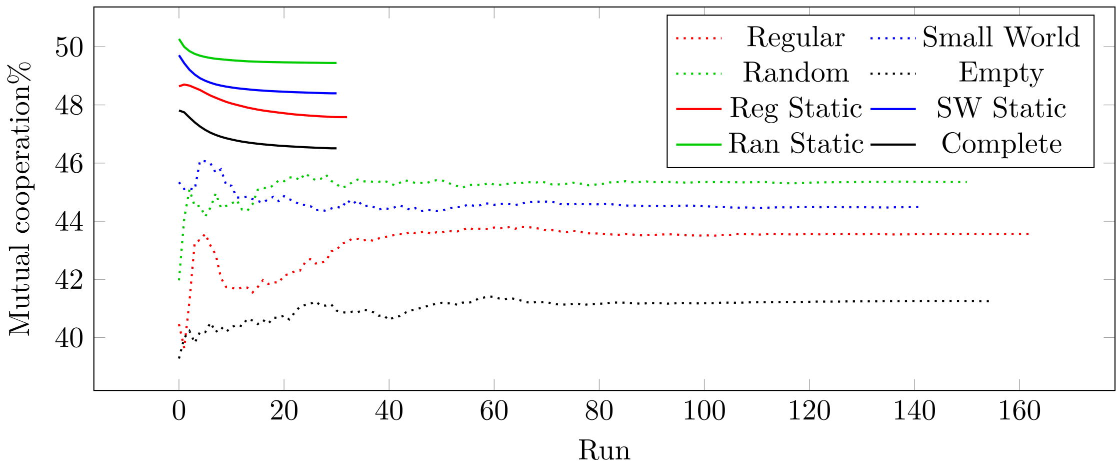


Figure 6.3: Proportion of mutual cooperation for each run, in each of the environments, for the agents playing the Stag Hunt.

further supporting the hypothesis $H3$, and the work reported in [68].

In addition, the arenas have a consistently lower level of cooperation and take longer to converge when compared to networked interactions ($H1$). To explore why this is the case, there is a need to look into the main difference between the two environments, namely the number of unique opponents the agents will face.

Figures 6.4, 6.5, 6.6, and 6.7 show the histogram of how many unique opponents an agent faced, for the empty, regular, small world, and random arenas respectively. Overlaying is a second histogram which excludes any opponent they interacted with exactly once. When an agent interacts with an opponent exactly once this will be termed a *singular interaction*.

When including the singular interactions, the range of agents faced does not line up with the expected distributions for the different arena structures. Across the networks there is the expectation of a singular peak in the histogram which would lean further right as the arena becomes more open. Excluding the singular interactions, these figures show that the shape of the environment affects the number of opponents as would be expected, with an agent in the empty environment interacting with the largest range of agents and the random environment facing the smallest range, in-line with expectations as this mimics the static networks. A significant factor can be seen in how the number of opponents decreases greatly when excluding singular interactions.

What the above means for the agents is that when mobility is introduced, there are effectively both the iterative social dilemmas and one-shot social dilemmas being played

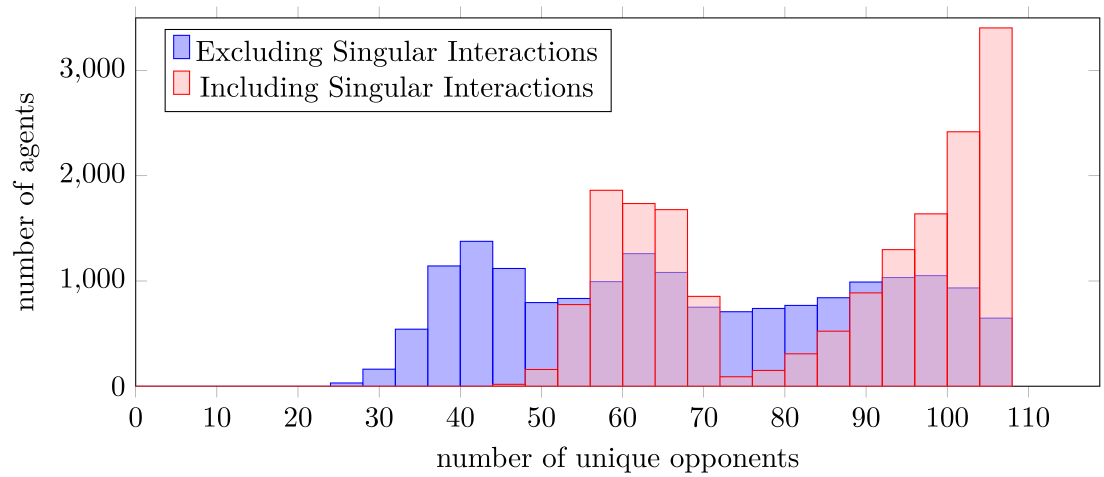


Figure 6.4: Histogram of unique opponents faced in the empty arena, including and excluding singular interactions.

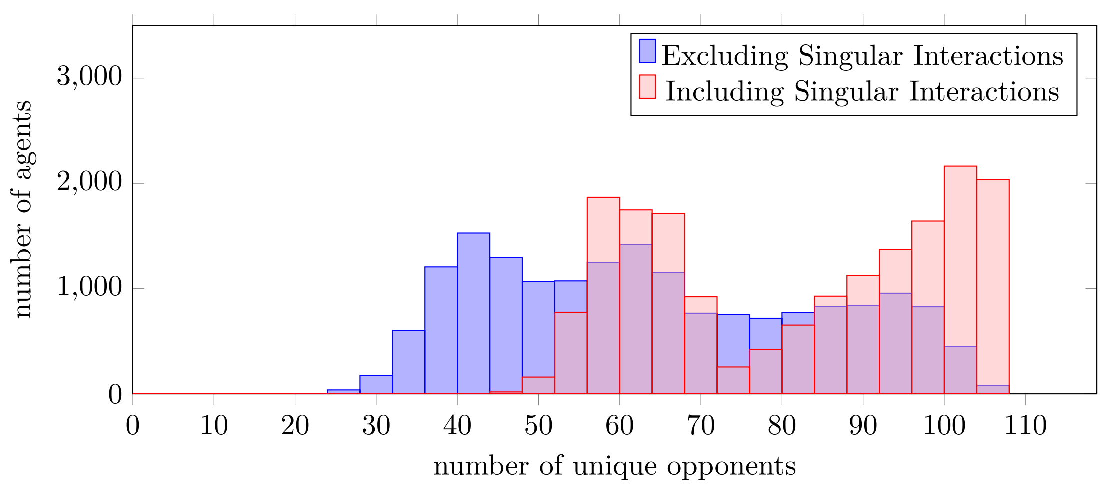


Figure 6.5: Histogram of unique opponents faced in the regular arena, including and excluding singular interactions.

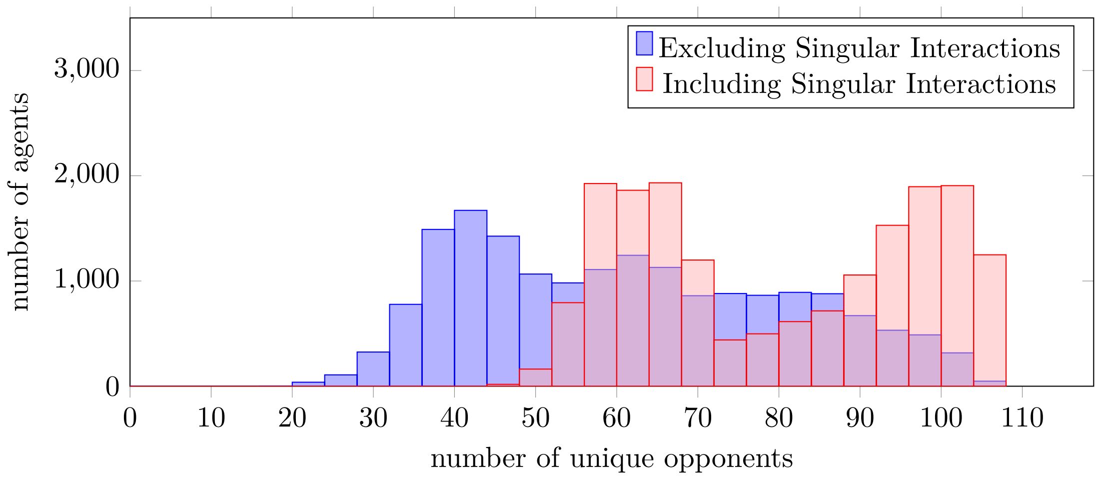


Figure 6.6: Histogram of unique opponents faced in the small world arena, including and excluding singular interactions.

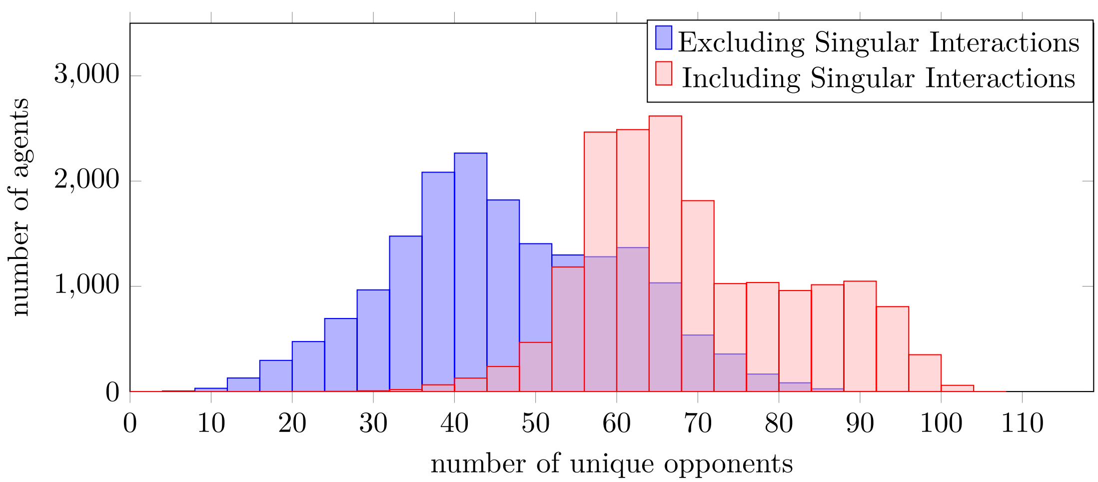


Figure 6.7: Histogram of unique opponents faced in the random arena, including and excluding singular interactions.

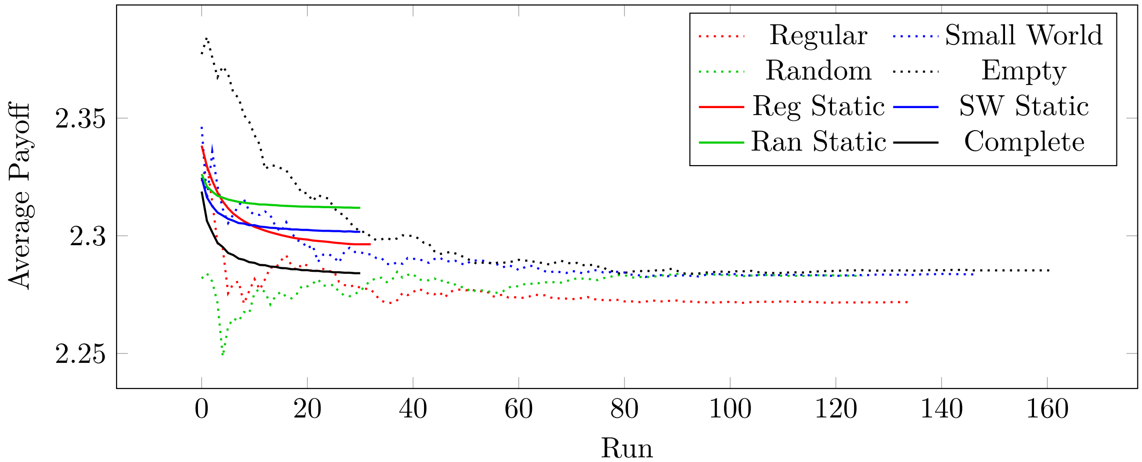


Figure 6.8: Average score for each environment in the Prisoner's Dilemma. Dotted lines show the mobile arenas, solid lines show the network environments.

at same time, with no knowledge on which is being played. Agents are able to take advantage of some opponents without retaliation, allowing defectors to go unpunished. Mutual cooperation is harder to sustain as cooperating agents can effectively be split up, as they may not interact with each other again. The average range of opponents is 89 (14.65 standard difference) in the mobile arenas, with an average of 76.5 (32 standard difference) in the static environments. The larger range of opponents in the mobile environment extends the convergence time, along with the irregular number of interactions with a specific opponent.

6.2.1 Payoffs

Figures 6.8 and 6.9 are used to analyse hypothesis $H2$ and $H3$. The figures show the average score of the society of agents in the Prisoner's Dilemma and the Stag Hunt respectively, for each environment. The results in these figures allow the analysis to reach an initial conclusion on hypothesises $H2$ and $H3$.

For hypothesis $H3$, Figures 6.8 and 6.9 showed that the fully connected network performed the worst with more randomised networks achieving higher scores in both the Prisoner's Dilemma and the Stag Hunt, validating hypothesis $H3$. When analysing Hypothesis $H2$, the original expectation was that the empty environment would perform best; this is supported for the Prisoner's Dilemma but not for the Stag Hunt. This

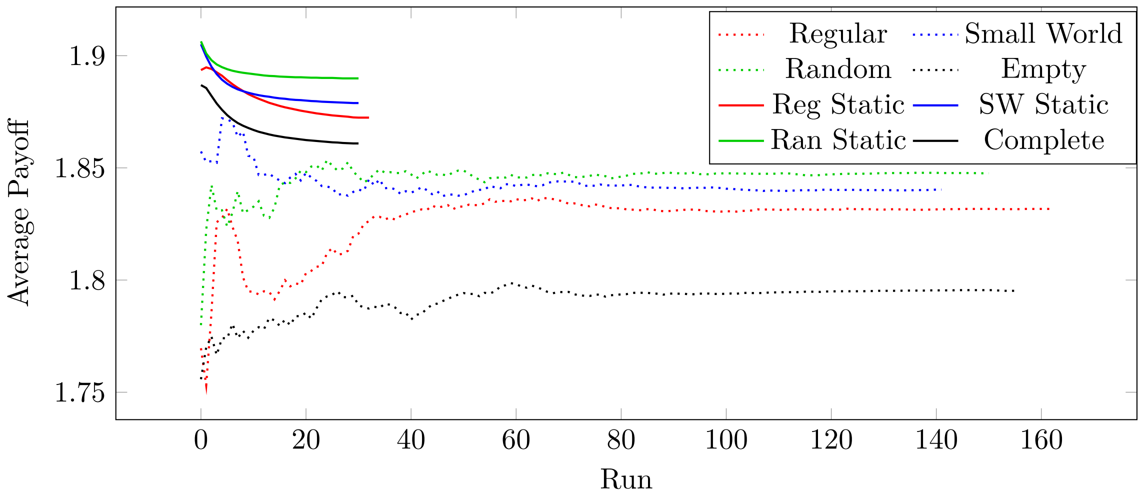


Figure 6.9: Average score for each environment in the Stag Hunt. Dotted lines show the mobile arenas, solid lines show the network environments.

leads to the conclusion that there does not seem to be a direct relation between the environment structure and the average payoff. Rather than focusing on whether there exists an environment which can affect the payoff directly, there will be more analysis on why two different social dilemmas yield different orderings (in terms of society payoff) for the different arenas, by looking deeper into the form of the games, and the effects on individual strategies.

The average scores of each strategy are shown in Tables 6.1, 6.2, 6.3, and 6.4. Tables 6.1 and 6.2 show the results for the Prisoner's Dilemma for the mobile and static environments respectively. The same is done for the Stag Hunt in Tables 6.3 and 6.4. The tables allow analysis of whether a majority of cooperation or a majority of defection is the most successful in the environments.

The tables show that in the Prisoner's Dilemma the fully cooperative agents perform the worst. The empty arena supports high levels of average payoff for strategies that choose defection more often. This defection is not seen in the Stag Hunt, which supports cooperative strategies. This is due to the temptation payoff being higher than the mutual cooperation payoff for an individual agent in the Prisoner's Dilemma, since the empty arena supports a large range of opponents, with a significant number of them being one-shot interactions. The agent that defects does not receive any retaliation allowing it to achieve high levels of payoff. The more cooperative agents do not receive a reduction in payoff

Table 6.1: Average score (percentage cooperation chosen) of each strategy in each of the mobile environments in the Prisoner's Dilemma. The most successful strategy is shown in bold.

	Environment			
Strategy	Empty	Small World	Regular	Random
WSLS	2.28 (80.4%)	2.33 (77.7%)	2.29 (79%)	2.37 (76.2%)
Random	2.32 (50%)	2.29 (50%)	2.30 (50%)	2.22 (50%)
All Coop	2.10 (100%)	2.14 (100%)	2.14 (100%)	2.12 (100%)
All Defect	2.34 (0%)	2.27 (0%)	2.27 (0%)	2.25 (0%)
TFT	2.35 (70.1%)	**2.35 (68.4%)**	**2.31 (67.5%)**	**2.36 (67.2%)**
SARSA	**2.35 (39.4%)**	2.26 (39.1%)	2.27 (39.3%)	2.27 (37.7%)
Active	2.26 (44.1%)	2.23 (44.8%)	2.23 (43.5%)	2.20 (43.9%)
Trustful	2.27 (71.6%)	2.26 (71.2%)	2.24 (70.5%)	2.28 (71.7%)
Moody	2.29 (50.5%)	2.26 (50.4%)	2.30 (50.5%)	2.26 (50.2%)

as large as the increase in payoff that less cooperative agents receive in the other arenas, allowing the empty arena to yield the highest average payoff.

The above effect is not seen in the Stag Hunt as the temptation payoff is less than the mutual cooperation payoff. Since defecting still receives less retaliation in the empty arena this does not result in the increased payoff that would be expected in the Prisoner's Dilemma. In general the results show that the environment has less of an effect on the strategies in the Stag Hunt. This highlights how the payoff matrix of a social dilemma and the environment are interlinked. Note that TFT was the most successful as per [5], and that Trustful was also a successful strategy across both social dilemmas, supporting $H1$ further.

6.3 Conclusion

To summarise this chapter, it presented an experiment with a number of different strategies and evaluated them in the Prisoner's Dilemma and the Stag Hunt games. The agents played the games in a number of different network topologies and their equivalent arenas with mobile agents. Through these experiments the results have shown how arenas lower the average payoff, and the level of cooperation, and increase convergence times when compared to the equivalent static network. This difference was attributed to an inherent property of mobility in an arena, namely the range of opponents that will be faced and how a number

Table 6.2: Average score (percentage cooperation chosen) of each strategy in each of the static environments in the Prisoner's Dilemma. The most successful strategy is shown in bold.

	Environment			
Strategy	Complete	SW Static	Reg Static	Ran Static
WSLS	2.37 (74.8%)	2.37 (74.5%)	2.36 (75.3%)	2.38 (74.5%)
Random	2.23 (50%)	2.23 (50%)	2.24 (50%)	2.23 (50%)
All Coop	2.16 (100%)	2.16 (100%)	2.15 (100%)	2.16 (100%)
All Defect	2.18 (0%)	2.17 (0%)	2.19 (0%)	2.17 (0%)
TFT	**2.39 (66.5%)**	**2.39 (66.2%)**	**2.39 (66.6%)**	**2.39 (66.3%)**
SARSA	2.22 (37.3%)	2.21 (37.4%)	2.23 (37.5%)	2.20 (37%)
Active	2.18 (45%)	2.18 (44.8%)	2.19 (45%)	2.18 (45.3%)
Trustful	2.28 (72.6%)	2.27 (72.5%)	2.27 (72.6%)	2.27 (72.3%)
Moody	2.23 (50%)	2.22 (50.1%)	2.23 (50.3%)	2.22 (50.1%)

Table 6.3: Average score (percentage cooperation chosen) of each strategy in each of the mobile environments in the Stag Hunt. The most successful strategy is shown in bold.

	Environment			
Strategy	Empty	Small World	Regular	Random
WSLS	2.05 (80.7%)	2.11 (79%)	2.09 (79.3%)	2.06 (75.9%)
Random	1.54 (50%)	1.53 (50.1%)	1.55 (50%)	1.50 (50%)
All Coop	2.06 (100%)	2.14 (100%)	2.12 (100%)	2.13 (100%)
All Defect	1.34 (0%)	1.33 (0%)	1.33 (0%)	1.32 (0%)
TFT	**2.13 (68.6%)**	**2.15 (68.3%)**	**2.17 (69.6%)**	**2.17 (67.4%)**
SARSA	1.61 (37.7%)	1.63 (39.8%)	1.62 (38.6%)	1.64 (39.3%)
Active	1.81 (43.5%)	1.84 (45%)	1.83 (44.3%)	1.85 (45.5%)
Trustful	2.05 (70.4%)	2.07 (70.6%)	2.10 (72%)	2.15 (72.4%)
Moody	1.59 (51.3%)	1.55 (50.6%)	1.56 (51.1%)	1.50 (50.4%)

Table 6.4: Average score (percentage cooperation chosen) of each strategy in each of the static environments in the Stag Hunt. The most successful strategy is shown in bold.

	Environment			
Strategy	Complete	SW Static	Reg Static	Ran Static
WSLS	2.10 (74.9%)	2.10 (74.8%)	2.09 (75.3%)	2.10 (74.5%)
Random	1.50 (50%)	1.49 (50%)	1.50 (50%)	1.49 (50%)
All Coop	2.16 (100%)	2.15 (100%)	2.16 (100%)	2.16 (100%)
All Defect	1.29 (0%)	1.29 (0%)	1.30 (0%)	1.29 (0%)
TFT	**2.20 (66.6%)**	**2.20 (66.5%)**	**2.20 (66.8%)**	**2.20 (66.7%)**
SARSA	1.61 (38.2%)	1.61 (38.1%)	1.62 (38.1%)	1.60 (38%)
Active	1.85 (45.1%)	1.85 (45.1%)	1.85 (45.1%)	1.84 (45%)
Trustful	2.17 (72.7%)	2.18 (72.8%)	2.17 (72.8%)	2.17 (72.6%)
Moody	1.50 (50.6%)	1.50 (50.4%)	1.51 (50.9%)	1.49 (50.2%)

of them will only be faced once, effectively making these one-shot interactions and thus limiting an agent's ability to retaliate against defection.

The results show that the more open an arena, the larger the range of opponents and number of singular interactions. For an agent to take advantage of the lack of retaliation, the payoff matrix needs to also support defection over cooperation for an individual agent. The Prisoner's Dilemma has this support, so defecting agents achieve high levels of payoff in the empty arena. Conversely the Stag Hunt does not have this support as the temptation payoff is lower than the mutual outcome payoff for an individual agent.

The results find that mobile agents show a lower level of cooperation when compared to static agents. The lower cooperation is due to mobile agents facing a larger range of opponents and interacting with these opponents fewer times than static agents. Mobile agents in an open arena will have a higher level of defection than in arenas with more obstacles, however there is only an increase in the society's payoff in the Prisoner's Dilemma and not the Stag Hunt for the given scenarios. This difference is due to temptation payoff being higher than mutual cooperation in the Prisoner's Dilemma and lower in the Stag Hunt. Overall the experiment has provided strong results for answering **SRQ3**. For the larger research question of the these, the results generalise the notable effects on agents that mobility has on agent societies.

Chapter 7

Evolutionary Stability

The work completed in the previous chapters has implemented and experimentally explored cooperation between emotional and moody agents. The aim of this chapter is to examine the emotional and mood models implementations within social dilemmas and where these strategies fit into the wider literature of social dilemma strategies. Section 7.1 gives an analysis of the evolutionary stability of emotional and moody agents in the Prisoner's Dilemma. Section 7.2 gives a simulation of the agents to show how the stability analysis is applicable in practice. A conclusion of the chapter is then given in Section 7.3. The analysis in this chapter will provide answers to the main **RQ** and **SRQ2** by capturing how competitive emotional and moody agents are when compared to other social dilemma strategies.

To analyse the evolutionary stability of these agents, the Prisoner's Dilemma game will be used as this allows the analysis to effectively look at the cooperation these societies achieve and whether the cooperation is sustainable against invading strategies. The work in this chapter has been published in the following papers: [28, 27].

7.1 Evolutionary Stability Analysis

To analyse whether emotional and moody agents can be considered an evolutionarily stable strategy, there needs to be an opponent strategy that will take the largest advantage of these agents and minimise their payoff. By designing such a strategy the analysis can show that if emotional and moody agents are able to remain the dominant strategy, then no other strategy can invade the emotional or moody agents.

The strategy will be termed the *oracle*. The effectiveness of the strategy is achieved by breaking an assumption of the Prisoner's Dilemma, namely that players have no knowledge of the opponent's move, as reflected in the name. Intuitively the oracle strategy will always cooperate with itself, and when faced with another strategy will choose the worst outcome for the opponent, effectively making it the worst case scenario for the opponent. The oracle strategy targets the conditions needed to be an evolutionarily stable strategy, allowing effective analysis of evolutionary stability.

For example, if an opponent chooses to cooperate, the oracle strategy is guaranteed to defect, giving the oracle strategy the T payoff and the opponent the S payoff. For a society of agents to successfully survive an oracle invasion, that society must have perfect cooperation among themselves, and protect themselves from the oracle by always defecting against the opposing strategy.

The expected value V after one round, for the oracle strategy o against strategy b, where $Ac(b,o)$ returns the action b would use against o, can be calculated as:

$$V(o,b) = \begin{cases} R & \text{IF } b \equiv o \\ T & \text{IF } b \not\equiv o \text{ AND } Ac(b,o) = C \\ P & \text{IF } b \not\equiv o \text{ AND } Ac(b,o) = D \end{cases} \tag{7.1}$$

The oracle is the most effective strategy at minimising the payoff of the emotional agents, which is shown in Theorem 1. To prove this theorem, we start with the following Lemma.

Lemma 1. An emotional agent will not change its subsequent action against an opponent if its opponent's action mirrors the emotional agent's action.

Intuitively this means that if an emotional agent is cooperating and its opponent is also cooperating, then the emotional agent will not switch to defection and vice-versa. Table 4.7 shows why emotional agents will not change their action, when both agents have the same action.

Theorem 1. The expected payoff of emotional agents using the defined characteristics, in the Prisoner's Dilemma with the payoffs defined in Table 3.1, is minimised by the oracle strategy, with no other strategy being able to lower the expected payoff further.

Proof. If the emotional agent is initially defecting then the payoff achieved by the emotional agent is $V(e, o) = Pn$ where n is the number of rounds. Neither the oracle nor the emotional agent will ever change their action, as per Lemma 1.

When the emotional agent is initially cooperating then the payoff the emotional agent receives is S as the oracle defects. By the definition of the emotional agent, the emotional agent will change to defection when the Anger level of that agent reaches the Anger threshold. The emotional agent will change its action to defection, against the oracle agent. Once the emotional agent has changed its action the oracle will continue to defect, now both agents are defecting. As both agents are defecting they will continue in mutual defection indefinitely as per Lemma 1, and the emotional agent will receive the P payoff. The expected value of an initially cooperative emotional agent against the oracle is $V(e, o) = S \cdot An + P(n - An)$ where An is the Anger threshold and n is the number of rounds.

Assume there is a strategy x where the payoff achieved by the emotional agent is $V(e, x) < S \cdot An + P(n - An)$ when initially cooperating and $V(e, x) < Pn$ when the emotional agent is initially defecting.

If the strategy x only defects then the payoff of an initially defecting emotional agent is $V(e, x) = Pn$, and for an initially cooperative agent $V(e, x) = S \cdot An + P(n - An)$. This contradicts the assumption as $V(e, x) < S \cdot An + P(n - An)$, therefore only defecting is not the strategy x.

If the strategy x only cooperates then the expected payoff of an initially defecting agent is $V(e, x) = T \cdot G + R(n - G)$ where G is the Gratitude threshold. The payoff for the initially cooperative agent is $V(e, x) = Rn$. This leads to a contradiction as T and R are both larger than S and P in the Prisoner's Dilemma. Strategy x therefore cannot only cooperate.

Strategy x must therefore be a mixed strategy. By Lemma 1, repeating the emotional agent action leads to an indefinite repetition. Therefore the strategy of doing the opposite of the emotional agent needs to be considered. When the emotional agent is cooperating the strategy x will defect and when the emotional agent is defecting the strategy x will cooperate. Therefore the expected value of the emotional agent is $V(e, x) = (T \cdot G + S \cdot An)\frac{n}{G+An}$.

This is the minimal strategy since if the strategy x switches a cooperative action for a defection action, then the emotional agent will receive a P payoff, but will not switch to cooperation, effectively removing the S payoffs it would have received. By definition of the Prisoner's Dilemma $P > S$ so the emotional agent's expected value will increase. If the strategy x switches a defection for a cooperative move, then the emotional agent will receive a R rather than a S payoff, and will not switch to defection leading to further R

payoffs. By definition $R > S$ so the emotional agent's expected value will increase.

As strategy x must be the mixed strategy of doing the opposite action of the emotional agent, therefore the following must hold:

$$V(e,x) < V(e,o) \tag{7.2}$$

The equation where $V(e,x) = (T \cdot G + S \cdot An)\frac{n}{G+An}$ and $V(e,o) = S \cdot An + P(n - An)$ with an initially cooperating emotional agent will be evaluated. The initially cooperating agent is chosen as $S \cdot An + P(n - An) < Pn$.

Given the list of possible characteristics in Table 4.1 for the emotional agents and using the values for the Prisoner's Dilemma in Table 3.1, the characteristics that will be chosen are the characteristics that will play the maximum number of cooperative moves and the lowest amount of defection moves to minimize the number of T payoffs and maximize the number of S payoffs. This is the characteristic Trustful. Plugging in the values in the above equation results in the following:

$$\begin{aligned}
(5 \cdot 1 + 0 \cdot 3)\frac{n}{4} &< 0 \cdot 3 + 1(n-3) \\
\frac{5}{4}n &< n - 3 \\
\frac{1}{4}n &< -3 \\
n &< -12
\end{aligned} \tag{7.3}$$

For the equation to hold, strategy x must yield a lower expected payoff to the emotional agent than the oracle, for any number of rounds n. This is therefore a contradiction as n must be positive by definition. Figure 7.1 shows this by plotting the expected payoff of the emotional agent after n rounds for both opponent strategies.

Therefore strategy x is not the mixed strategy. It was also shown that only cooperating and only defecting are not strategy x.

$\therefore$ The expected payoff of emotional agents using the defined characteristics, in the Prisoner's Dilemma with the payoffs defined in Table 3.1 is minimised by the oracle strategy, with no other strategy being able to lower the expected payoff further. $\square$

The oracle agent is the most effective agent at minimising the expected payoff of the emotional agents, for the given characteristics and the given values for the Prisoner's

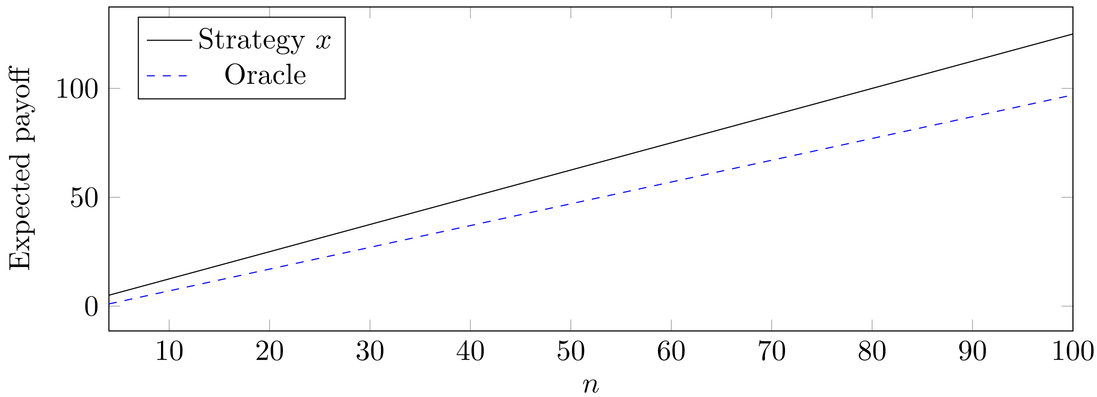


Figure 7.1: Expected payoff of an initially cooperative emotional agent, playing against the mixed strategy x and the oracle agent after n interactions.

Dilemma. By restricting the analysis to the given agent characteristics and values, the ability to analyse the emotional agents is more effective by looking at one opposing strategy rather than all possible strategies.

7.1.1 Emotional Agents

This section will start by showing that the emotional agents are not an evolutionarily stable strategy when there are no restrictions on reproduction and interaction speed. This will be shown for both for the initially cooperative emotional agent and the initially defecting emotional agent.

Theorem 2. Emotional agents are not an evolutionarily stable strategy in the initial phase.

Proof. Assume that emotional agents are an evolutionarily stable strategy. Given a majority of emotional agents, with an invasion force of oracle agents, by definition of an evolutionarily stable strategy, Equation 3.3 must hold for the emotional agent strategy M and the invading oracle strategy I. The initially cooperative emotional agent gives the following,

$$R > T \text{ OR } (R = T \text{ AND } S > R) \tag{7.4}$$

and for the initially defecting emotional agent

$$P > P \text{ OR } (P = P \text{ AND } P > R). \tag{7.5}$$

A contradiction has been reached for each line in both the initially cooperative emotional agent and the initially defecting emotional agent. $P > P$ is a contradiction, $T > R$ and $R > P$ (from Table 3.1) contradict the equations.

$\therefore$ Emotional agents are not an evolutionarily stable strategy in the initial phase. $\square$

Emotional agents are not an evolutionarily stable strategy, due to how the agents respond initially to the oracle strategy. Emotional agents are able to respond to the opponent on an individual agent level, that is, the action the emotional agent gives depends on who the opponent is. Section 4.4.2 in Chapter 4 has shown that all initially cooperative Emotional agents will cooperate with each other indefinitely. Emotional agents will eventually choose to defect indefinitely against the oracle strategy given enough time to adjust. The number of interactions needed per pairing of emotional and oracle agents is the same as the anger threshold of the Emotional agent.

Lemma 2. All emotional agents will converge to defection with all oracle agents given a converting number of interactions and sufficiently randomness in pairing.

As other agents cannot affect the action choice of either the emotional agent or the oracle, the oracle will only defect which only increases the agent's Anger value. The emotional agent will not switch back to cooperation.

Theorem 3. Initially cooperative emotional agents that have fast interactions and slow reproduction are an evolutionarily stable strategy

Proof. Assume a fraction ϵ of the population is replaced by the invading oracle strategy. There is also the assumption that interactions between all agents are fast and reproduction of the population is slow. Given the fast interactions, and slow reproduction with respect to time, all emotional agents will be defecting against any other oracle agents that may be residual in the populous, as per Lemma 2. This gives both the oracle strategy and the emotional agents the P payoff. No emotional agent has adjusted to the newly invading ϵ-oracle agents, and as such are able to receive the S payoff.

Thus the expected payoff of the emotional agents against the oracle agents is $V(e, o) = S\epsilon + P(1 - \epsilon)$, and the expected payoff for the oracle agents against the emotional agents will therefore be $V(o, e) = T\epsilon + P(1 - \epsilon)$.

Using these values in Equation 3.3 gives the following:

$$\begin{aligned} &R > T\epsilon + P(1-\epsilon) \text{ OR} \\ &(R = T\epsilon + P(1-\epsilon) \text{ AND } S\epsilon + P(1-\epsilon) > R) \end{aligned} \tag{7.6}$$

The equation will therefore hold, given that ϵ is sufficiently small as per the definition of an evolutionarily stable strategy [41]. The expected value that the oracle agent gets from the emotional agents will be sufficiently close to P such that the first line will always hold. The emotional agents are protecting themselves from defection of the oracle agents. The ϵ number of new oracle agents are unable to take a large enough advantage of the emotional agents that they can break the stability.

$\therefore$ Initially cooperative emotional agents are an evolutionarily stable strategy, when interactions are fast and reproduction is slow. □

In summary, initially cooperative emotional agents are an evolutionarily stable strategy, as no strategy is able to minimise the payoff of the emotional agents more than the oracle agent. The assumptions of fast interactions and slow reproduction, are to allow the emotional agents to adapt to all the oracle agents before the next reproduction. A sufficiently small epsilon in this case is less than half, if using Table 3.1 as the payoff matrix, given the assumption that the emotional agents are the majority as per the definition of an evolutionarily stable strategy [41]. The assumptions of fast interaction and slow reproduction are part of an efficient evolution and learning process [59], with fast interactions allowing the agent to learn which, in turn guides the reproduction process.

Considering the initially defecting emotional agents, they have already adapted to the invading oracle agents. However the initially defecting emotional agent being able to protect its payoff is not enough for it to be considered an evolutionarily stable strategy. The initial defection will prevent the emotional agents from cooperating as a group, and with no potential for breaking the defection, this allows the oracle agents, which do work together, to be a fitter strategy.

7.1.2 Moody Agents

This subsection will now be comparing the emotional agents to moody agents, again using the oracle strategy. The moody agents analysed are similar to emotional agents as they both use OCC-inspired emotions as part of their decision-making process. However the

Mood Level	Moody Agent Cooperating	Moody Agent Defecting
Very High ($An > 90$)	No change	Cooperate
High ($70 > An \geq 90$)	No change	Cooperate against new opponent
Neutral ($30 \geq An \leq 70$)	No change	No change
Low ($10 \leq An < 30$)	Defect against new opponent	No change
Very Low ($An < 10$)	Defect	No change

Table 7.1: How different simulated mood levels change the action selection in moody agents

addition of the Mood model on top of the emotions changes how the moody agents react in certain circumstances. The moody agents have been shown in Chapter 5 to perform better in self-play than the emotional agents. Analysing the evolutionary stability of these agents shows the effects the Mood model has, when compared to emotional agents.

To analyse the moody agents, different mood levels need to be taken into account as they affect how the moody agents respond to the oracle. Therefore each mood level needs to be analysed individually to be able to gain insights into the Mood model as a whole. Table 7.1 shows each Mood level and when the Mood value will override the action selection of the agent. Knowing that emotional agents need to be initially cooperative to be considered an evolutionarily stable strategy, for this analysis there is the assumption that the moody agents are also initially cooperative. The proofs commence by considering very high levels of mood down to very low moods. For simplicity, the analysis is presented in respect of each different mood level

Theorem 4. Moody agents that are in an initially very high mood are not an evolutionarily stable strategy

Proof. Assume moody agents in a very high mood are an evolutionarily stable strategy. Given that a fraction ϵ of the population is replaced by the invading oracle strategy, the expected payoff of two moody agents in a very high mood is R since by definition all moody agents are cooperating. Thus the expected payoff of an oracle agent against the moody agent will be T and the moody agent will receive S. Equation 3.3 holds by definition of an evolutionarily stable strategy. Therefore the following equation should be true:

$$R > T \text{ OR } (R = T \text{ AND } S > R) \tag{7.7}$$

This is a contradiction, both sides of the disjunction are false. By definition of the Prisoner's Dilemma, $T > R$ which contradicts both sides of the equation.

$\therefore$ Moody agents in an initially very high mood are not an evolutionarily stable strategy. $\square$

Moody agents in very high moods are not an evolutionarily stable strategy. This is due to these particular agents being functionally equivalent to a fully cooperative strategy, which is known to not be an evolutionarily stable strategy [17, 73]. Moving onto high moods, it will be shown that initially cooperative agents are equivalent to moody agents in a neutral mood. This will help to avoid repeating proofs.

Lemma 3. Moody agents that are in an initially high mood are functionally equivalent to moody agents in a neutral mood.

In high moods the only effect on decision making is that the moody agent will always cooperate with an unknown opponent. By the assumption above, moody agents are initially cooperative, and therefore moody agents in a high mood are functionally identical to moody agents in a neutral mood.

Neutral moods are functionally equivalent to emotional agents and this will be shown next.

Lemma 4. Moody agents that are in an initially neutral mood are functionally equivalent to emotional agents.

The mood value has no effect on action selection; by definition of moody agents, they will respond using the emotion model as defined for the emotional agent. The result for moody agents in a neutral or high mood are functionally equivalent to emotional agents means that Theorem 3 holds. Therefore moody agents in a neutral mood or high mood can be considered evolutionarily stable strategies given the same conditions.

The next part of this section will go on to show that initially very low, and low, mood levels are not an evolutionarily stable strategy, to demonstrate that both types of moody agents are functionally equivalent against an oracle agent.

Lemma 5. Moody agents that are in an initially low mood are functionally equivalent to moody agents against an oracle agent.

The moody agents will defect with all other moody agents as the initial cooperation is broken. The low moods change the first action to defection, so the moody agents will continue to defect indefinitely as per Lemma 1. The moody agents in a very low mood will defect by definition. When playing against an oracle both the oracle and the moody agent in a low mood will play defection. The oracle will also defect against a moody agent in a very low mood. Regardless of whether the moody agent is in a low mood or a very low mood, they will defect indefinitely with both other moody agents and oracle agents.

Theorem 5. Moody agents that are in an initially low mood or are in an initially very low mood are not an evolutionarily stable strategy

Proof. Assume moody agents in low mood or a very low mood are an evolutionarily stable strategy. Given that a fraction ϵ of the population is replaced by the invading oracle strategy, the expected payoff of two moody agents is P. This is valid for both very low and low moods as per Lemma 5.

The expected payoff of an oracle agent against a defecting moody agent is P, and they will be in mutual defection indefinitely. Equation 3.3 holds by definition of an evolutionarily stable strategy. Therefore the following equation should be true:

$$P > P \text{ OR } (P = P \text{ AND } P > R) \tag{7.8}$$

Two contradictions have been reached, $P > P$ and $P > R$ since by definition of the Prisoner's Dilemma $R > P$.

$\therefore$ Moody agents that are in an initially low mood or are in an initially very low mood are not an evolutionarily stable strategy. □

To conclude that moody agents overall are not an evolutionarily stable strategy, there is a need to now show that the moody agents' mood level will always lead to the evolutionarily unstable mood levels. As only neutral and high moods are possibly an evolutionarily stable strategy, only these two mood levels need to be considered.

Theorem 6. Moody agents in an initially neutral or initially high mood will move to the very high mood level, when there is a sufficiently small ϵ invasion of oracle agents.

Proof. The expected payoff of a moody agent(k) in either a neutral or high mood will be $V(k, o) = S\epsilon + P(1 - \epsilon)$ against an oracle agent and $V(k, k) = R$, as Theorem 3 applies as per Lemma 4, since moody agents are functionally equivalent to emotional agents.

When a moody agent receives a payoff its mood level updates. Therefore $\Omega_i(j) \approx \mu_i$ when two moody agents interact. The majority of interactions are between two moody agents, therefore making their averages (μ_i and μ_j) approximately equal, which means Ω is not changing the perception of the reward in Equation 5.2.

The mood value is updated by the equation $m_i \leftarrow m_i + (r_i - \mu_i) + \Omega_i(j)$. The majority of the interactions being between two moody agents is therefore $r_i - \mu_i \approx 0$. As $\Omega > 0$, the m_i will increase in the majority of cases indefinitely as the invasion of oracle agents is ϵ small.

$\therefore$ Moody agents in an initially neutral or initially high mood will move to the very high mood level, given a sufficiently small ϵ invasion of oracle agents. $\square$

In conclusion, moody agents are not an evolutionarily stable strategy. While moody agents may be an evolutionarily stable strategy in neutral and high moods, with the same conditions as the emotional agents, the moody agents will move into the other mood levels that are not an evolutionarily stable strategy. If the mood level of moody agents was to stay stable over time, this would go against the design principles of the model [23]. The psychological grounding of the moody model requires that mood levels change over time as per the psychology literature [79, 34].

7.2 Simulation

This section will provide a simulation of an agent society, interacting in the Prisoner's Dilemma against an invasion force of oracle agents. The agent society will consist of either emotional agents or moody agents. The aim of this simulation is to show how the previous theoretical analysis is accurate and applicable to these agents in practice. The aim for the simulation using the emotional agents is to show that the requirements for evolutionary stability are achievable in a reasonable time frame. The moody agent simulation aim is to give further justification to the claim that the moody agents are not an evolutionarily stable strategy.

The simulations produce a graph which shows the evolution of the average payoff for each different strategy, and how they compare to each other. The Prisoner's Dilemma values that are used are given in Table 3.1. The simulation uses a total of 1000 agents, where 50 agents are the invading oracle strategy and 950 will be either the emotional agents or the moody agents, depending on the scenario. The agents are paired randomly and play

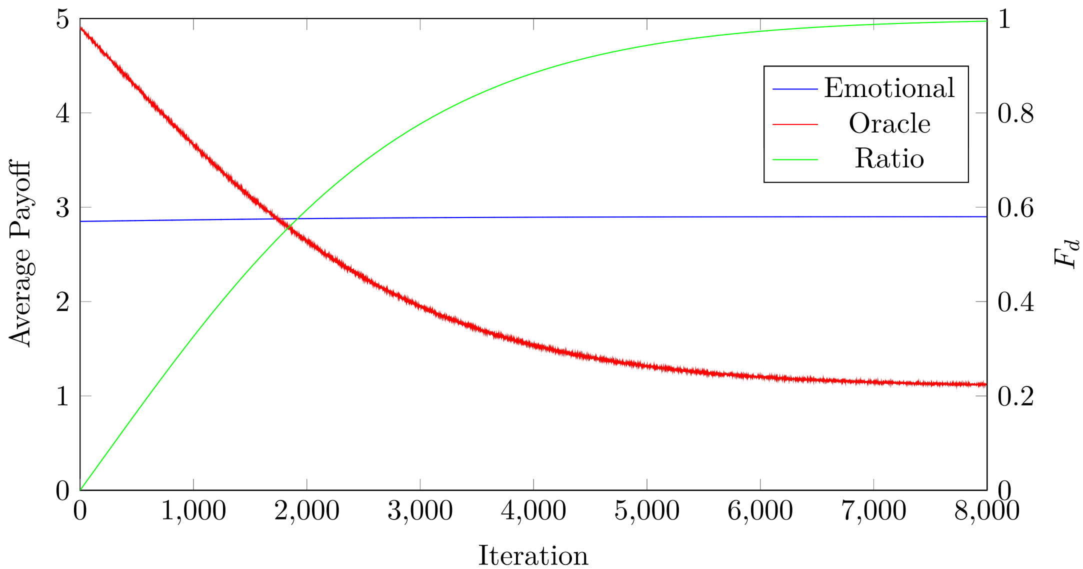


Figure 7.2: Average payoff of the emotional strategy and the oracle strategy over 8000 interactions, averaged over 1000 simulations. The ratio of emotional agents who choose to defect against all oracle agents is plotted as Ratio.

one round of the Prisoner's Dilemma. The process of randomising the agents and playing a single round is repeated for 8000 iterations. To ensure that the values that are produced are accurate, the simulation will repeat the above process 1000 times. The emotional agents and moody agents are both initially cooperative, with the moody agents being initialised in a neutral mood, and both use an equal mix of the characteristics in Table 4.1.

The results of the emotional agents interacting against the oracle agents are given in Figure 7.2. Note that after around 2000 iterations the oracle agent does worse in terms of payoff than the emotional agent, and continues to do so indefinitely. This reflects the theoretical analysis and shows that emotional agents are able to adapt to the oracle agent in a reasonable time. Next are the moody agents, where the simulation was conducted identically but with moody agents rather than emotional agents. The average payoff over time is shown in Figure 7.3.

The figure shows that moody agents cannot protect themselves from the oracle agents. The number of moody agents who would normally choose to defect against the oracle increases similarly to the emotional agents. The mood level of the moody agents is kept very high by having a majority of moody agents. This causes the moody agents to override

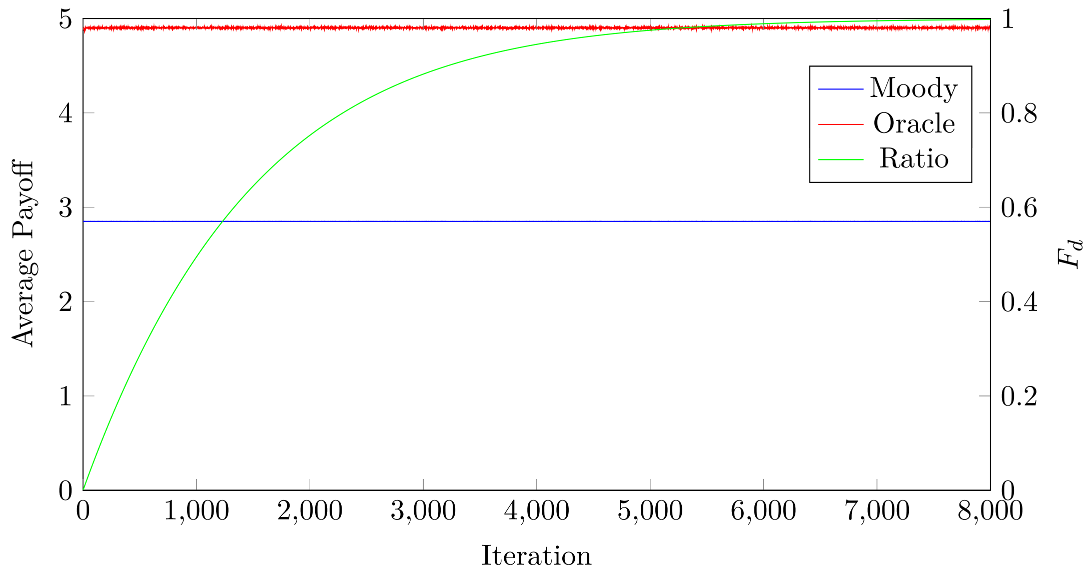


Figure 7.3: Average payoff of the moody strategy and the oracle strategy over 8000 interactions, averaged over 1000 simulations. The ratio of moody agents who choose to defect, before any mood level changes, against all oracle agents is plotted as Ratio.

the defection choice, preventing them from protecting themselves when playing against the oracle agents, which are able to take full advantage.

7.3 Conclusion

To summarise the chapter, analysing the emotional and moody agents has allowed us to show the inherent risk that cooperation brings in the Prisoner's Dilemma. The work has also showed how different strategies are better deployed in different scenarios. emotional agents are better suited to a mixed group of agents with differing strategies than the moody agents, while moody agents are better suited than emotional agents when only one strategy exists.

The analysis has shown that emotional agents that use a model of emotions as part of their decision-making can be considered an evolutionarily stable strategy when they initially cooperate with new partners and are able to adapt to an invading strategy before reproducing. The emotional agents take a more defensive strategy that allows cooperation to remain stable over time. The analysis also showed that moody agents using a simulated model of mood alongside the model of emotions as their decision-making process, are not part of an evolutionarily stable strategy. This is because some mood levels break the assumption that moody agents cooperate together and will always protect themselves from invading strategies. Opponents are able to take advantage of the moody agents as they try to create cooperation. The emotional and moody agents were tested against an oracle strategy, which was shown to be the most effective at minimising the expected payoff of the emotional agents and can successfully invade the moody agents.

To support the findings of the evolutionarily stable strategy it was further shown that emotional agents meet the required conditions for stability, and in a reasonable time, through simulations with 950 emotional and 50 oracle agents. The emotional agents were able to reach a point where they had adapted in around 2000 iterations on average, when playing against the invading oracle agents.

The chapter has contributed to answering **SRQ2** as showing that emotional agents are an evolutionarily stable strategy demonstrates that they are competitive against a wide range of opponents. The moody agents not being an evolutionarily stable strategy shows how they are not as competitive as the emotional agents; this lack of competitiveness is the cost of being able to create cooperation in more friendly environments.

Chapter 8

Conclusion

The aim of this thesis was to explore how simulated emotional and moody agents could cooperate as a society and be considered competitive, when placed within a mobile society. This aim was expressed in the main research question **RQ** which is described in Chapter 1 and was:

> *Can agents with simulated emotions and simulated mood be cooperative and competitive in a self-interested mobile society?*

In order to effectively address the question, a number of different aspects needed to be analysed which to a number of sub-research questions. These sub-research questions were identified as:

SRQ1 *How can we computationally model decision making using emotions and mood?*

SRQ2 *How can we capture cooperation and competitiveness in emotional and moody agent societies?*

SRQ3 *How does mobility affect agents in their interactions within a society?*

In order to start exploring the questions, there was a need to understand what the current psychological literature states when considering emotions, mood, and decision making. Chapter 2 analysed the psychology research and showed that emotions are difficult to define, however there are a number of models of emotions, which present various levels of difficulty when considering whether they can be adapted to a computational model.

Mood was shown to be even more difficult to define, with the psychology research often having conflicting definitions of mood. Given these conflicting definitions of mood, there was a need to decide on a characterisation that would be used throughout. For this thesis the characterisation of mood was that mood is synonymous with what is defined in the psychology research as positive and negative affect.

Analysing whether previous implementations of emotional and moody agents have strong psychology underpinnings was explored in Chapter 3. This chapter found that some implementations of simulated emotions were weakly underpinned by psychology research, while other implementations had a strong underpinning in psychology research. Implementations that use simulated mood were often weakly validated by psychology research and often used simulated emotions alongside the simulated mood. The chapter also explored how the agents using simulated emotions and mood were to be evaluated. The agents in this thesis are tested experimentally within social dilemmas, notably the Prisoner's Dilemma. Considering whether the simulated emotional and moody agents are evolutionarily stable strategies was also shown to be an effective way of evaluating the agents in a wider context.

Following on from this, Chapter 4 explored whether or not mobility had an effect on a previous implementation [69] of a simulated emotional agent. The effect of mobility was observed in a first experiment, before a second experiment was conducted to further understand the effect of mobility. The chapter contributed significantly to answering the sub research questions. For **SRQ1** by simulating emotional agents in a mobile setting. The second sub research question **SRQ2**, was addressed through the experiments taking place in a Prisoner's Dilemma setting, which captures cooperation among agents, and in analysing the average payoff of the self-interested agents, the competitiveness was captured. By simulating the agents in different arenas, **SRQ3** was addressed.

Chapter 5 provided a generic computational model of mood grounded in psychology. The model had two implementations, one that included the previous emotional agents and another that integrated reinforcement learning. Both the implementations were tested in an arena where the agents had mobility. The mobility given to the agents allows the results to answer **SRQ3**, and also showed the effects mobility had on the SARSA reinforcement learning algorithm. The main contribution in the Chapter was to **SRQ1** where the generic mood model along with the two implementation was given. **SRQ2** was answered in a similar manner to Chapter 4 by using the Prisoner's Dilemma as the game to play in the agent interactions.

The effects of mobility studied in Chapters 4 and 5 were further examined in a more general sense in Chapter 6. This chapter used a number of different strategies in the four mobile arenas and compared them to the same strategies in equivalent network constructions. The experiment also expanded the range of agents that were tested. The main aim was to expand the investigation on mobility and environment structure and so contributed significantly to answering **SRQ3**.

Chapter 7 placed the emotional agents and the moody agents in the context of the wider literature on Prisoner's Dilemma strategies. This context was achieved by analysing whether the emotional agents and moody agents can be considered as evolutionarily stable strategies. The results of this analysis have contributed to answering **SRQ2**, in the context of competitiveness among all strategies in the Prisoner's Dilemma.

The remainder of the final chapter will set out the contributions this thesis has made to the wider literature. After the contributions are summarised in Section 8.1 there will be a discussion of the limitations of the work conducted, and what further questions this work has raised; that discussion is given in Section 8.2. The thesis concludes with some final remarks in Section 8.3.

8.1 Contributions

This section will list the major contributions that presented in the thesis. The contributions in this thesis have been achieved through both experimental and theoretical work. Chapters 2 and 3 provided the background and highlighted the gaps in the literature on which the main research question **RQ** of the thesis is based. Chapters 4, 5 and 6 make their contributions through experimental work that is based on simulating agents in a number of different scenarios. The theoretical analysis in Chapter 7 provides an evolutionarily stable strategy analysis. The major contributions from each individual chapter are:

Chapters 2 & 3 give the background literature review of the thesis, the notable part of these chapters is showing the gaps in the literature. The gaps that were noted were that there was *a need for a model of mood that is computational and is psychologically grounded* when considering the outward effects of mood. Mood was shown to have multiple definitions in the psychology literature, with this thesis defining mood as affect. The chapters go on to show how *the environment is a factor that is often a missing aspect in experimentation* when considering agent interactions. When simulating emotional and moody agents the analysis is often experimental, there is therefore a need to analyse these kinds of agents in a theoretical way to provide further conclusive results about the agents' behaviour.

Chapter 4 started by describing the framework for the experiments that were conducted for the thesis. The first experiment showed that *mobility had an effect on the simulated emotional agents*. The effects that were noted included the changes to average payoff and which emotional characteristics were successful. The emotional characteristic that was the most successful in the past literature was shown to be the least successful in the scenarios where mobility is introduced. Mobility therefore has an impact on the agents. The second experiment expanded to four arenas which were created so that the density of the agents was the same throughout. This second experiment found that the number of unique interactions affected the simulated emotional agents' performance, and that the number of unique interactions was linked to the structure of the environment. Overall, *successful emotional characteristics were able to respond to defection quickly while taking a small advantage*.

Chapter 5 provided the general computational model of mood, along with the psychological grounding of this model. An implementation of the mood model was then given. The first experiment conducted showed that *Moody agents were able to create and sustain cooperation in the Prisoner's Dilemma*. The average level of mood is related to the average level of cooperation. Overall the average level of cooperation increased over time for the Moody agents whereas the emotional agents were stable over time. When the Moody agents were played against an invasion force of pure defectors, the level of cooperation collapsed among the Moody agents, when the mood level was high. A second implementation of mood replaced the underlying simulated emotions with a reinforcement learning process. *There was an increase in the average level of cooperation and average payoff for the Moody agents when compared to SARSA*. The structural effects of the environment that were noted previously were shown to have the same effect for the Moody agents, namely that *different characteristics are successful depending on the environment*. More open arenas supported agents that took advantage, while more closed off arenas supported cooperative agents.

Chapter 6 gave an in-depth look at the effects environment structure and mobility has on agents in general. *Mobility was shown to lower the average payoff of the agents,* *the average level of cooperation, and increase convergence times.* The reason for this is that mobility removes guarantees on the number of opponents an agent will face, with significant numbers of opponents only being faced once. The structure of the arena showed that the number of singular interactions increased, and the number of unique opponents increased, thus increasing the mobility effect. In addition it was shown that agents are able to take advantage of the effect when it is supported by the payoff matrix: *the payoff matrix needs to provide higher payoffs for defection for an individual agent than cooperation.* The Prisoner's Dilemma supports the effect of mobility, which is shown throughout the previous chapters. The Stag Hunt game does not support this effect, and as such the cooperative strategies are able to be more successful.

Chapter 7 provided a theoretical understanding of the simulated emotional and moody agents. The understanding was achieved through an evolutionarily stable strategy analysis. The results of the analysis showed that *moody agents are not an evolutionarily stable strategy.* The reason is due to how the moody agents will try to create cooperation, which opens them up to being taken advantage of. *Emotional agents were shown to be an evolutionarily stable strategy*, given some reasonable requirements. The requirements are that the emotional agents have time to adjust to the invading strategy through quick interactions and slow replication, and that the emotional agents are initially cooperative. An example was presented that showed that these requirements can be achieved in a reasonable time, with 950 emotional agents achieving a higher average payoff than the 50 invading oracle agents in around 2000 interactions.

The code for all the experiments can be found on GitHub. `https://github.com/JoeCol/ThesisExperiment`

8.2 Limitations and Future Work

The aim of this section is to review the limits of the work presented, along with avenues for further exploration that may unearth interesting findings. These avenues for further exploration may be of interest to researchers interested in studying simulated emotional and moody agents, and mobile agents that interact with an environment such that this affects the agents' decision making.

The focus of this thesis has been on simulating emotions and mood, and emotions and mood have a relation to ethical considerations. Ethics is becoming more prominent within AI as AI technologies become more embedded in real world applications. Simulating emotions and mood may be a topic of interest for implementations of agents that need to make ethical decisions. The agent may analyse how a certain decision can affect both its own emotional state and the other agents' emotional state. The ability to incorporate ethical decision making into agents is becoming more prominent as agents are starting to be deployed in scenarios where they have to make decisions that do not necessarily have a correct answer and may impact the safety of human beings. A popular example is the self-driving car, which needs to exist among other human driven cars. The self-driving car may find itself in an ethically challenging situation of the type that has been widely studied in the philosophy literature on trolley problems [46].

The thesis has gone into depth analysing both simulated emotions and simulated mood, with frequent reference to the previous literature. While the background of the thesis reviewed many different implementations of simulated emotions and simulated mood, the analysis presented in this thesis is limited to one specific implementation. The level of analysis provided would be useful to apply to different kinds of implementations with an analysis of what differences the alternative implementations give. Once this has been completed it would be useful to see whether any patterns emerge in the differences and whether such patterns apply at a broader level.

Chapter 6 started with the exploration of other social dilemmas and in this chapter the second social dilemma used was the Stag Hunt. The evolutionarily stability analysis in Chapter 7 was independent of specific payoff matrix values for the Prisoner's Dilemma. This highlights how the thesis has mainly focused on the Prisoner's Dilemma. The Prisoner's Dilemma is an effective way of analysing cooperation for self-interested agents but there may be some interesting results when considering different social dilemmas, such as whether particular kinds of social dilemma are better at putting pressure on cooperation than others. There may be more interesting results when extending to analysing social dilemmas independently of specific payoff matrix values, which would require further analysis; such as consideration of evolutionarily stable strategies.

Moving on to the structure of the environments, the thesis has looked at four different environmental structures. For agent designers to gain further knowledge on how environmental structures will affect their agents, a wider view of graph structures is needed. Chapter 3 introduced the scale-free network which would be the next logical investigation to conduct. Given the way this thesis constructed the environmental structures, such as the small-world, regular, and random environments, there maybe some interesting results when looking at the finer points of randomisation, and what effects increasing the randomness in the construction of the network and arenas has on the agents.

Staying with the arenas, the agents that are presented throughout the work all use a random walk. How an agent interacts with the environment will have an effect on the results. The manoeuvrability of the agents will change the effects that the environment has, so agent designers who are interested in how mobility impacts their agents will need to take into account how their agents move. Chapter 3 has a small mention of this by comparing to Ranjbar et al.'s work [95], where the authors' agents were allowed to physically interfere with the movement of the other agents, causing differences in how the environmental structure affected the agents.

8.3 Closing Summary

Simulating emotions and mood is an approach to allowing dynamic systems to respond to actions that are not well defined by the system. Using simulated emotions and mood allows for responses to be developed on both a global and individual levels, allowing the system to be more adaptable over time. These dynamic systems are applicable to a wide range of societies of agents, from real world implementations such as self-driving cars which can adapt to other road users independently to other more theoretical societies of self-interested societies.

Simulating mood is a natural extension to the work on simulated emotions, when considering how decision-making in agents can be informed by these concepts. There has been little work in applying mood as a generic computational model. The work in this thesis has contributed to establishing models of computational mood within agents and how such models can be applied with simulated emotions or any other underlying decision making process. When considering computational agents, especially those which use simulated emotions and mood, there is a focus on using experimental work that uses networked interactions. This thesis has presented a number of experiments with emotional and moody agents that are not only experiments with mobile agents but also include theoretical results on an implementation of emotional and moody agents.

This thesis has shown that mobility is an important aspect to consider when analysing agent interactions, especially given the increased interest in deploying agents in real world scenarios where the agents will not be in neat constrained networks. Not only is mobility important, but the structure of the environment will also have an effect on the agents, which changes outcomes regarding which simulated emotional agents are successful. The simulated emotional agents were shown to converge to an outcome when playing against another agent using the same strategy. The algorithm to show how to calculate which set of actions the two emotional agents will converge to was given, along with why the algorithm holds for all possible simulated emotional agents. A generic computational model of mood was given to expand on simulated emotions, and two implementations were given. A reflection was then given on the simulated emotional and moody agents in the context of the wider literature on Prisoner's Dilemma strategies by analysing whether emotional or moody agents can be considered evolutionarily stable strategies. The above as presented in this thesis is a significant contribution to the state-of-the-art regarding simulated emotional and moody agents in a mobile agent society, which gives a solid foundation for further investigation to drive forward research in AI on cooperative, automated decision making.

Bibliography

[1] André, E., Klesen, M., Gebhard, P., Allen, S., and Rist, T. *Integrating Models of Personality and Emotions into Lifelike Characters.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2000, pp. 150–165.

[2] Arellano, D., Perales, F. J., and Varona, J. Mood and its mapping onto facial expressions. In *Articulated Motion and Deformable Objects* (Cham, 2014), F. J. Perales and J. Santos-Victor, Eds., Springer International Publishing, pp. 31–40.

[3] Ashlock, D., and Rogers, N. A model of emotion in the prisoner's dilemma. In *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology* (2008), IEEE, pp. 272–279.

[4] Aspinwall, L. G. Rethinking the role of positive affect in self-regulation. *Motivation and emotion 22*, 1 (1998), 1–32.

[5] Axelrod, R., and Hamilton, W. D. The evolution of cooperation. *Science 211*, 4489 (1981), 1390–1396.

[6] Barabási, A.-L. Scale-free networks: A decade and beyond. *Science 325*, 5939 (2009), 412–413.

[7] Barrat, A., and Weigt, M. On the properties of small-world network models. *The European Physical Journal B-Condensed Matter and Complex Systems 13*, 3 (2000), 547–560.

[8] Baumeister, R. F., Bratslavsky, E., Finkenauer, C., and Vohs, K. D. Bad is stronger than good. *Review of general psychology 5*, 4 (2001), 323.

[9] Baxter, P. E., Wood, R., Morse, A., and Belpaeme, T. Memory-centred architectures: Perspectives on human-level cognitive competencies. In *AAAI Fall Symposium Series* (2011).

[10] Bench-Capon, T., Atkinson, K., and McBurney, P. Using argumentation to model agent decision making in economic experiments. *Autonomous Agents and Multi-Agent Systems 25*, 1 (2012), 183–208.

[11] Bilderbeck, A. C., Reed, Z. E., McMahon, H. C., Atkinson, L. Z., Price, J., Geddes, J. R., Goodwin, G. M., and Harmer, C. J. Associations between mood instability and emotional processing in a large cohort of bipolar patients. *Psychological Medicine* (2016), 1–10.

[12] Bloembergen, D., Caliskanelli, I., and Tuyls, K. Learning in networked interactions: A replicator dynamics approach. In *Artificial Life and Intelligent Agents Symposium* (2014), Springer, pp. 44–58.

[13] Bloembergen, D., Ranjbar-Sahraei, B., Bou Ammar, H., Tuyls, K., and Weiss, G. Influencing social networks: An optimal control study. In *Proceedings of ECAI'14* (2014), pp. 105–110.

[14] Bloembergen, D., Tuyls, K., Hennes, D., and Kaisers, M. Evolutionary dynamics of multi-agent learning: a survey. *Journal of Artificial Intelligence Research 53* (2015), 659–697.

[15] Bonsall, M. B., Wallace-Hadrill, S. M., Geddes, J. R., Goodwin, G. M., and Holmes, E. A. Nonlinear time-series approaches in characterizing mood stability and mood instability in bipolar disorder. *Proceedings of the Royal Society of London B: Biological Sciences* (2011), rspb20111246.

[16] Bower, G. H. Mood and memory. *American psychologist 36*, 2 (1981), 129.

[17] Boyd, R., and Lorberbaum, J. P. No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game. *Nature 327* (05 1987), 58 EP –.

[18] Broekens, J., and Haazebroek, P. Emotion and reinforcement: affective facial expressions facilitate robot learning. In *Artifical Intelligence for Human Computing*. Springer, 2007, pp. 113–132.

[19] Broome, M. R., Saunders, K. E., Harrison, P. J., and Marwaha, S. Mood instability: significance, definition and measurement. *The British Journal of Psychiatry 207*, 4 (2015), 283–285.

[20] Carver, C. S., and Scheier, M. F. Origins and functions of positive and negative affect: A control-process view. *Psychological review 97*, 1 (1990), 19.

[21] Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. The effect of mobility and emotion on interactions in multi-agent systems. In *Proceedings of STAIRS'16* (2016), D. Pearce and H. S. Pinto, Eds., IOS Press, pp. 39–50.

[22] Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. Mobility effects on the evolution of co-operation in emotional robotic agents. In *Proceedings of ALA'16* (2016), pp. 114–121.

[23] Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. Modelling mood in co-operative emotional agents. In *Proceedings of Distributed Autonomous Robotic Systems* (2016), pp. 573–586.

[24] Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. Environmental effects on simulated emotional and moody agents. *The Knowledge Engineering Review 32, e19* (2017), 1–24.

[25] Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. Mood modelling within reinforcement learning. In *Proceedings of ECAL'17* (2017), MIT Press, pp. 106–113.

[26] Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. On the role of mobility and interaction topologies in social dilemmas. In *Proceedings of ALIFE'18* (2018), pp. 447–484.

[27] Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. Stability of cooperation in societies of emotional and moody agents. In *Proceedings of ALIFE'19* (2019), pp. 467–474.

[28] Collenette, J., Atkinson, K., Bloembergen, D., and Tuyls, K. Stability of human-inspired agent societies. In *Proceedings of AAMAS'19* (2019), pp. 1889–1891.

[29] Crandall, J. W., and Goodrich, M. A. Learning to compete, compromise, and cooperate in repeated general-sum games. In *Proceedings of ICML'05* (2005), ACM, pp. 161–168.

[30] Crawford, V. P. Learning and mixed-strategy equilibria in evolutionary games. *Journal of Theoretical Biology 140*, 4 (1989), 537–550.

[31] DeSteno, D., Bartlett, M. Y., Baumann, J., Williams, L. A., and Dickens, L. Gratitude as moral sentiment: emotion-guided cooperation in economic exchange. *Emotion 10*, 2 (2010), 289.

[32] Devlin, S., Grześ, M., and Kudenko, D. Multi-agent, reward shaping for robocup keepaway. In *Proceedings of AAMAS'10* (2011), IFAAMAS, pp. 1227–1228.

[33] Dewey, J. The theory of emotion. *Psychological Review 2* (1895), 13–32.

[34] Diener, E., Larsen, R. J., Levine, S., and Emmons, R. A. Intensity and frequency: dimensions underlying positive and negative affect. *Journal of personality and social psychology 48*, 5 (1985), 1253.

[35] Digman, J. M. Personality structure: Emergence of the five-factor model. *Annual review of psychology 41*, 1 (1990), 417–440.

[36] Durán, O., and Mulet, R. Evolutionary prisoner's dilemma in random graphs. *Physica D: Nonlinear Phenomena 208*, 3 (2005), 257–265.

[37] Eldar, E., and Niv, Y. Interaction between emotional state and learning underlies mood instability. *Nature communications 6* (2015).

[38] Enke, B. Kinship, Cooperation, and the Evolution of Moral Systems*. *The Quarterly Journal of Economics 134*, 2 (01 2019), 953–1019.

[39] Erdős, P., and Rényi, A. On random graphs i. *Publ. Math. Debrecen 6* (1959), 290–297.

[40] Erev, I., and Roth, A. E. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *The American Economic Review 88*, 4 (1998), 848–881.

[41] Eshel, I. Evolutionary and continuous stability. *Journal of theoretical Biology 103*, 1 (1983), 99–111.

[42] Farrell, J., and Ware, R. Evolutionary stability in the repeated prisoner's dilemma. *Theoretical Population Biology 36*, 2 (1989), 161 – 166.

[43] Fehr, E., Fischbacher, U., and Gächter, S. Strong reciprocity, human cooperation, and the enforcement of social norms. *Human nature 13*, 1 (2002), 1–25.

[44] Fehr, E., Naef, M., and Schmidt, K. M. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. *The American economic review 96*, 5 (2006), 1912–1917.

[45] Fehr, E., and Schmidt, K. M. A theory of fairness, competition, and cooperation. *Quarterly journal of Economics 114* (1999), 817–868.

[46] Foot, P. The problem of abortion and the doctrine of double effect. *Oxford Review 5* (1967), 5–15.

[47] Freud, S. *Civilization and its discontents*. Broadview Press, 2015.

[48] Gadanho, S. C., and Hallam, J. Robot learning driven by emotions. *Adaptive Behavior 9*, 1 (2001), 42–64.

[49] Gerkey, B., Vaughan, R. T., and Howard, A. The player/stage project: Tools for multi-robot and distributed sensor systems. In *Proceedings of ICAR'03* (2003), pp. 317–323.

[50] Gibson, E. L. Emotional influences on food choice: sensory, physiological and psychological pathways. *Physiology & behavior 89*, 1 (2006), 53–61.

[51] Gintis, H. *Game theory evolving: A problem-centered introduction to modeling strategic behavior*. Princeton university press, 2000.

[52] Gradin, V. B., Pérez, A., Macfarlane, J. A., Cavin, I., Waiter, G., Tone, E. B., Dritschel, B., Maiche, A., and Steele, J. D. Neural correlates of social exchanges during the prisoner's dilemma game in depression. *Psychological medicine 46*, 06 (2016), 1289–1300.

[53] Haley, W. E., and Strickland, B. R. Interpersonal betrayal and cooperation: Effects on self-evaluation in depression. *Journal of Personality and Social Psychology 50*, 2 (1986), 386.

[54] Hebb, D. O. *The organization of behavior: A neuropsychological theory.* Psychology Press, 2005.

[55] Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., and McElreath, R. In search of homo economicus: behavioral experiments in 15 small-scale societies. *American Economic Review 91*, 2 (2001), 73–78.

[56] Hepburn, L., and Eysenck, M. W. Personality, average mood and mood variability. *Personality and Individual Differences 10*, 9 (1989), 975–983.

[57] Hertel, G., Neuhof, J., Theuer, T., and Kerr, N. L. Mood effects on cooperation in small groups: Does positive mood simply lead to more cooperation? *Cognition & emotion 14*, 4 (2000), 441–472.

[58] Hilbe, C., Traulsen, A., and Sigmund, K. Partners or rivals? strategies for the iterated prisoner's dilemma. *Games and Economic Behavior 92* (2015), 41–52.

[59] Hinton, G., and Nowlan, S. How learning can guide evolution. *Complex Systems 1* (1987), 495–502.

[60] Hofmann, L.-M., Chakraborty, N., and Sycara, K. The evolution of cooperation in self-interested agent societies: a critical study. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 2* (2011), International Foundation for Autonomous Agents and Multiagent Systems, pp. 685–692.

[61] Ichinose, G., Satotani, Y., and Nagatani, T. Network flow of mobile agents enhances the evolution of cooperation. *EPL (Europhysics Letters) 121*, 2 (2018), 28001.

[62] Jacobs, E., Broekens, J., and Jonker, C. Emergent dynamics of joy, distress, hope and fear in reinforcement learning agents. In *Proceedings of ALA'14* (2014).

[63] Keltner, D., and Gross, J. J. Functional accounts of emotions. *Cognition & Emotion 13*, 5 (1999), 467–480.

[64] Kim, J. H., and Vu, V. H. Generating random regular graphs. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing* (2003), ACM, pp. 213–222.

[65] Leahy, R. L. Clinical implications in the treatment of mania: Reducing risk behavior in manic patients. *Cognitive and Behavioral Practice 12*, 1 (2005), 89 – 98.

[66] Leeper, R. W. A motivational theory of emotion to replace'emotion as disorganized response.'. *Psychological Review 55*, 1 (1948), 5.

[67] Levenson, R. W. Human emotion: A functional view. *The nature of emotion: Fundamental questions 1* (1994), 123–126.

[68] Lieberman, E., Hauert, C., and Nowak, M. A. Evolutionary dynamics on graphs. *Nature 433*, 7023 (2005), 312.

[69] Lloyd-Kelly, M. *Modelling Emotions and Simulating their Effects on Social Interactions in Agent Systems.* PhD thesis, University of Liverpool, 2014.

[70] Lloyd-Kelly, M., Atkinson, K., and Bench-Capon, T. Developing co-operation through simulated emotional behaviour. In *13th International Workshop on Multi-Agent Based Simulation* (2012).

[71] Lloyd-Kelly, M., Atkinson, K., and Bench-Capon, T. Emotion as an enabler of co-operation. In *ICAART (2)* (2012), pp. 164–169.

[72] Lloyd-Kelly, M., Atkinson, K., and Bench-Capon, T. Fostering co-operative behaviour through social intervention. In *Proceedings of SIMULTECH'14* (2014), IEEE, pp. 578–585.

[73] Lorberbaum, J. No strategy is evolutionarily stable in the repeated prisoner's dilemma. *Journal of Theoretical Biology 168*, 2 (1994), 117 – 130.

[74] Lount Jr., R. B. The impact of positive mood on trust in interpersonal and intergroup interactions. *Journal of Personality and Social Psychology 98*, 3 (2010), 420 – 433.

[75] Maass, W. Networks of spiking neurons: The third generation of neural network models. *Neural Networks 10*, 9 (1997), 1659 – 1671.

[76] Malhi, G. S., Byrow, Y., Fritz, K., Das, P., Baune, B. T., Porter, R. J., and Outhred, T. Mood disorders: neurocognitive models. *Bipolar disorders 17*, S2 (2015), 3–20.

[77] Marsella, S., Gratch, J., and Petta, P. Computational models of emotion. *A Blueprint for Affective Computing-A sourcebook and manual* (2010), 21–46.

[78] Masuda, N., and Ohtsuki, H. A theoretical analysis of temporal difference learning in the iterated prisoner's dilemma game. *Bulletin of mathematical biology 71*, 8 (2009), 1818–1850.

[79] Mayer, J. D., and Hanson, E. Mood-congruent judgment over time. *Personality and Social Psychology Bulletin 21* (1995), 237–237.

[80] Mehrabian, A. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology 14*, 4 (1996), 261–292.

[81] Mehrabian, A. Analysis of affiliation-related traits in terms of the pad temperament model. *The Journal of Psychology 131*, 1 (1997), 101–117. PMID: 9018859.

[82] Moerland, T. M., Broekens, J., and Jonker, C. M. Emotion in reinforcement learning agents and robots: a survey. *Machine Learning 107*, 2 (Feb 2018), 443–480.

[83] Mondada, F., Bonani, M., Raemy, X., et al. The e-puck, a robot designed for education in engineering. In *Proceedings of ICARSC'09* (2009), pp. 59–65.

[84] Nash, J. Non-cooperative games. *Annals of Mathematics 54*, 2 (1951), 286–295.

[85] Ng, A. Y., Harada, D., and Russell, S. Policy invariance under reward transformations: Theory and application to reward shaping. In *Proceedings of ICML'99* (1999), Morgan Kaufmann, pp. 278–287.

[86] Nowak, M., Sigmund, K., et al. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature 364*, 6432 (1993), 56–58.

[87] Ojha, S., and Williams, M.-A. Ethically-guided emotional responses for social robots: Should i be angry? In *Proceedings of ICSR'16* (2016), Springer, pp. 233–242.

[88] Oosterbeek, H., Sloof, R., and van de Kuilen, G. Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics 7*, 2 (Jun 2004), 171–188.

[89] Ortony, A., Clore, G. L., and Collins, A. *The cognitive structure of emotions*. Cambridge university press, 1990.

[90] Phelps, E. A. Emotion and cognition: insights from studies of the human amygdala. *Annu. Rev. Psychol. 57* (2006), 27–53.

[91] Posner, J., Peterson, B. S., and Russell, J. A. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology 17*, 3 (2005), 715–734.

[92] Randløv, J., and Alstrøm, P. Learning to drive a bicycle using reinforcement learning and shaping. In *ICML* (1998), vol. 98, Citeseer, pp. 463–471.

[93] Ranjbar-Sahraei, B., Ammar, H. B., Bloembergen, D., Tuyls, K., and Weiss, G. Theory of cooperation in complex social networks. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence (AAAI-14)* (2014).

[94] Ranjbar-Sahraei, B., Bou Ammar, H., Bloembergen, D., Tuyls, K., and Weiss, G. Evolution of cooperation in arbitrary complex networks. In *Proceedings of AAMAS'14* (2014), pp. 677–684.

[95] Ranjbar-Sahraei, B., Groothuis, I. M., Tuyls, K., and Weiss, G. Valuation of cooperation and defection in small-world networks: A behavioral robotic approach. In *Proceedings of BNAIC 2014* (2014), pp. 103–110.

[96] Ranjbar-Sharaei, B., Bloembergen, D., Ammar, H. B., Tuyls, K., and Weiss, G. Effects of evolution on the emergence of scale free networks. In *Proceedings of the 14th International Conference on the Synthesis and Simulation of Living Systems* (2014), pp. 376–383.

[97] Rapoport, A., Chammah, A. M., and Orwant, C. J. *Prisoner's dilemma: A study in conflict and cooperation*, vol. 165. University of Michigan press, 1965.

[98] Rinck, M., Glowalla, U., and Schneider, K. Mood-congruent and mood-incongruent learning. *Memory & cognition 20*, 1 (1992), 29–39.

[99] Rolls, E. T. On the brain and emotion. *Behavioral and brain sciences 23*, 2 (2000), 219–228.

[100] Russell, J. A. Evidence of convergent validity on the dimensions of affect. *Journal of personality and social psychology 36*, 10 (1978), 1152.

[101] Russell, J. A. A circumplex model of affect. *Journal of personality and social psychology 39*, 6 (1980), 1161.

[102] Rusting, C. L. Personality, mood, and cognitive processing of emotional information: three conceptual frameworks. *Psychological bulletin 124*, 2 (1998), 165.

[103] Sakellariou, I., Kefalas, P., Savvidou, S., Stamatopoulou, I., and Ntika, M. The role of emotions, mood, personality and contagion in multi-agent system decision making. In *Artificial Intelligence Applications and Innovations* (Cham, 2016), L. Iliadis and I. Maglogiannis, Eds., Springer International Publishing, pp. 359–370.

[104] Santos, F., and Pacheco, J. Scale-free networks provide a unifying framework for the emergence of cooperation. *Physical Review Letters 95* (2005), 1–4.

[105] Santos, F. C., Santos, M. D., and Pacheco, J. M. Social diversity promotes the emergence of cooperation in public goods games. *Nature 454*, 7201 (2008), 213–216.

[106] Santos, R., Marreiros, G., Ramos, C., Neves, J., and Bulas-Cruz, J. Personality, emotion and mood simulation in decision making. In *Proceedings of EPIA'09* (2009).

[107] Scherer, K. R., Schorr, A., and Johnstone, T. *Appraisal processes in emotion: Theory, methods, research.* Oxford University Press, 2001.

[108] Schwarz, N. Emotion, cognition, and decision making. *Cognition and Emotion 14*, 4 (2000), 433–440.

[109] Shteingart, H., Neiman, T., and Loewenstein, Y. The role of first impression in operant learning. *Journal of Experimental Psychology: General 142*, 2 (2013), 476.

[110] Skyrms, B. *The stag hunt and the evolution of social structure.* Cambridge University Press, 2004.

[111] Smith, J. M., and Price, G. R. The logic of animal conflict. *Nature 246*, 5427 (1973), 15.

[112] Starnini, M., Sánchez, A., Poncela, J., and Moreno, Y. Coordination and growth: the stag hunt game on evolutionary networks. *Journal of Statistical Mechanics: Theory and Experiment 2011*, 05 (2011), P05008.

[113] Steger, A., and Wormald, N. C. Generating random regular graphs quickly. *Combinatorics, Probability and Computing 8*, 4 (July 1999), 377–396.

[114] Steunebrink, B. R., Dastani, M., and Meyer, J.-J. C. A logic of emotions for intelligent agents. In *Proceedings of AAAI'07* (2007), vol. 22, p. 142.

[115] Sutton, R. S., and Barto, A. G. *Reinforcement learning: An introduction.* MIT press Cambridge, 1998.

[116] Szolnoki, A., and Perc, M. Resolving social dilemmas on evolving random networks. *EPL (Europhysics Letters) 86*, 3 (2009), 30007.

[117] Thomas, B. On evolutionarily stable sets. *Journal of Mathematical Biology 22*, 1 (Jun 1985), 105–115.

[118] Vassiliades, V., Cleanthous, A., and Christodoulou, C. Multiagent reinforcement learning: Spiking and nonspiking agents in the iterated prisoner's dilemma. *IEEE transactions on neural networks 22*, 4 (2011), 639–653.

[119] Vaughan, R. Massively multiple robot simulations in stage. *Swarm Intelligence 2*, 2-4 (2008), 189–208.

[120] Vukov, J., Szabó, G., and Szolnoki, A. Cooperation in the noisy case: Prisoner's dilemma game on two types of regular random graphs. *Physical Review E 73*, 6 (2006), 067103.

[121] Watkins, C. J., and Dayan, P. Q-learning. *Machine learning 8*, 3-4 (1992), 279–292.

[122] Watson, D. *Mood and temperament.* Guilford Press, 2000.

[123] WATSON, D., CLARK, L. A., AND TELLEGEN, A. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology 54*, 6 (1988), 1063.

[124] WATSON, D., AND TELLEGEN, A. Toward a consensual structure of mood. *Psychological bulletin 98*, 2 (1985), 219.

[125] WATTS, D. J., AND STROGATZ, S. H. Collective dynamics of 'small-world' networks. *Nature 393*, 6684 (06 1998), 440–442.

[126] YU, C., ZHANG, M., REN, F., AND TAN, G. Emotional multiagent reinforcement learning in spatial social dilemmas. *IEEE transactions on neural networks and learning systems 26*, 12 (2015), 3083–3096.

[127] ZEVON, M. A., AND TELLEGEN, A. The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology 43*, 1 (1982), 111.