UTRECHT UNIVERSITY

Coalition Formation between Self-Interested Heterogeneous Actors

by

Arlette van Wissen

August 2009

UTRECHT UNIVERSITY

Coalition Formation between Self-Interested Heterogeneous Actors

by

Arlette van Wissen

August 2009

A thesis submitted in partial fulfillment for the
degree of Master of Sciences

of

Artificial Intelligence

in the

Faculty of Humanities
Graduate School of Natural Sciences

60 ECTS

First Supervisor:	dr. Virginia Dignum	(Utrecht University)
Second Supervisor:	dr. Ya'akov Gal	(Harvard University)
Third Supervisor:	dr. Harald Kunst	(Utrecht University)

Declaration of Authorship

I, Arlette van Wissen, declare that this thesis titled, 'Coalition Formation between Self-Interested Heterogeneous Actors' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at Utrecht University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.

Signed: ________________________________

Date: ________________________________

Coming together is a beginning. Keeping together is progress. Working together is success.

Henry Ford (1863-1947)

UTRECHT UNIVERSITY

Abstract

Faculty of Humanities
Graduate School of Natural Sciences

Master of Sciences

by Arlette van Wissen

This thesis addresses successful coalition formation between heterogeneous self-interested actors. It focuses on interactions between self-interested humans and software agents that take place when they are required to cooperate. Cooperation can maximize the payoff of actors, but it also involves a higher risk of failure because they depend on the performance of other actors. In these interactions, the actors search for the best possible partner to carry out a task. It turns out that humans are not fully rational when it comes to choosing the best partner. Instead, social considerations strongly influence their decisions. We explore the effect of those considerations in a setting in which actors have to choose between different possible partners as their team members. More specifically, we will investigate how nature and trust influence people's decisions in mixed coalition formation. We will look at this in a fast-paced domain that is characterized by a high degree of uncertainty. Our findings show that in some situations people use different criteria for humans and agents. Also, a positive trust relation positively influences cooperation between actors.

Acknowledgements

I am grateful to many people for their support and collaboration during the course of my research. In my experience, coalition formation between scientists yields a much higher payoff than working individually.

First of all, I thank my supervisor Virginia Dignum. When I first started to work with Virginia, my ambitions were to learn about conducting research in general and to explore the world of academia. Virginia has given me the opportunity to do just that. She has helped me find my way in the Multi-Agent Systems community. Virginia has been very supportive and has always been willing to give advise, to discuss my work and to provide critical reviews. I was fortunate to have her as my supervisor.

Secondly, I owe many thanks to Ya'akov (Kobi) Gal. Kobi and I have worked together for little over a year and during this entire period he has helped me to improve my research through numerous discussions. My frequent talks with him were of great guidance. I also thank him for inviting me to set up and perform our experiment at the Artificial Intelligence Research Group (AIRG) at Harvard University.

I thoroughly enjoyed the experience of being part of the AIRG research group. I would like to thank all AIRG members for the pleasant and friendly atmosphere. In particular, I would like to express my gratitude to Philip Hendrix, Dimitrios Antos, Ece Kamar and Swapna Reddy for their input and feedback on the project. I also thank them for making me feel at home. Another special thanks goes out to Professor Stuart Shieber, who's enthusiasm and knowledge were of great value to our project.

Many thanks also to Maarten Engelen, who was always available to share with us his programming expertise. The brainstorm sessions with Maarten have helped shape both the conceptual setup and the implementation of our project.

I thank my third supervisor Harald Kunst for reviewing this thesis. His comments and suggestions are highly valued.

A special word of thanks to my family who have supported me in all my choices. I thank my parents for passing onto me the hunger to learn.

Finally, I thank Bart Kamphorst for his limitless support. He was part of the project from day one and his participation has made this project as comprehensive as it currently is. Bart has made countless contributions, varying from discussing the conceptual setup to implementing the domain, from to proof reading this thesis and putting up both with my excited babbling and my worries about the project. Bart has helped me to keep things in perspective and to stay motivated.

Contents

Declaration of Authorship iii

Abstract v

Acknowledgements vi

Abbreviations ix

1 Introduction 1
- 1.1 Hypotheses . . . 3
- 1.2 Research Method . . . 4
- 1.3 Contributions . . . 4
- 1.4 Thesis Outline . . . 5

2 Background 7
- 2.1 Multi-Agent Systems . . . 7
- 2.2 Mixed-Initiative Interactions . . . 8
- 2.3 Game Theory . . . 8
  - 2.3.1 Cooperativeness . . . 8
- 2.4 Coalition Formation . . . 9
  - 2.4.1 Building Blocks . . . 9
  - 2.4.2 Negotiation . . . 10
  - 2.4.3 Payoff . . . 11
- 2.5 Game Classification . . . 12
- 2.6 Social Factors . . . 12
  - 2.6.1 Trust . . . 12
  - 2.6.2 Fairness . . . 13
  - 2.6.3 Social Dilemma . . . 14

3 Related Work 15
- 3.1 Coalition Formation . . . 15
- 3.2 Social Dynamics . . . 17

4 Conceptual Approach 21
- 4.1 The Package Delivery Domain . . . 21
- 4.2 Colored Trails . . . 22
  - 4.2.1 A Testbed for Investigating Decision Making . . . 22
  - 4.2.2 Components . . . 23
    - Board . . . 23
    - Chips . . . 23
    - Phases . . . 23

Proposals 24
4.3 Conceptual Design 24
4.3.1 Adapting the PDD: MIPDD 24
4.3.2 Agent Implementation 25
4.3.3 Knowledge constraints 26
4.3.4 Choosing Team members 27
4.3.5 Defection 27

5 Experimental Design **29**
5.1 The Domain 29
5.2 The Game 32
5.2.1 Game configuration 32
5.2.2 Defecting 34
5.2.3 Game Flow 34
5.2.4 Considerations and Pilot Experiments 36
5.3 The Experiment 37
5.4 Evaluation 38

6 Results and Analysis **41**
6.1 Analysis 41
6.2 Coalition Formation 42
6.3 Social Factors 45
6.4 Concluding Remarks 47

7 Future Work **49**
7.1 MIPDD 49
7.2 BDI Agents 49
7.3 Pre-Established Payoff Distributions 50
7.4 Utilitarian Strategy Preferences 50

8 Conclusion **51**

A Handout **53**
A.1 Outline 53
A.2 Questions 53

B Questionnaire **59**
B.1 Outline 59
B.2 The Questionnaire 59

C Screenshots **67**

Bibliography **73**

Abbreviations

AI	**A**rtificial **I**ntelligence
BDI	**B**elief **D**esire **I**ntention
CF	**C**oalition **F**ormation
CT	**C**olored **T**rails
LP	**L**arge **P**ackage
MAS	**M**ulti **A**gent **S**ystem
MIPDD	**M**ixed **I**nitiative **P**ackage **D**elivery **D**omain
PDD	**P**ackage **D**elivery **D**omain
SP	**S**mall **P**ackage
UG	**U**ltimatum **G**ame

Chapter 1

Introduction

Social interactions between computer agents and humans are becoming increasingly important. Both in everyday life and in science humans use, control and cooperate with agents. When humans and agents form mixed networks, these networks can benefit from both the computational capabilities of the agents and the social capabilities of humans. Nonetheless, the inclusion of agents in human networks presents novel problems for the design of autonomous agent decision-making mechanisms. It becomes important for computer agents to operate in open and dynamic mixed environments. In these environments, humans encounter other humans and also computer agents, serving different people and organizations. Each of them could have different interests, plans and goals. In this thesis, we will refer to computer agents as 'agents' and we will use the term 'actors' to represent both humans and computer agents as entities that are capable of autonomously acting in an environment.

Common interactions with agents concern cooperative and helpful interactions, where agents are designed to help people make better decisions or carry out tasks. Examples of these kind of agent systems are pilot training, simulation-based learning environments, or feedback systems. Other examples of agents designed to assist humans are planning mechanisms for carpooling, agents that aid air traffic management by developing traffic schedules and agents that can help people to make optimal bids in negotiations. There are also many cases in which we interact with agents in a competitive setting. Examples of these interactions are e-commerce biddings or multi-player games with computerized opponents.

Humans are thus familiar with various kinds of human-agent interaction. These interactions are often called *mixed-initiative* interactions: both agent and humans can take initiative to perform an action. Although both cooperative and competitive human-agent interactions are well studied, a far less explored area concerns interactions that consist of a combination of cooperative and competitive motives. In realistic scenarios, humans and agents may be required to cooperate even though they are not a member of a specific team beforehand. They do not need to share the same goal and can have hidden motives or plans. Think for example about a poker game where two players may want to work together to remove a third player from the game. They will cooperate while ultimately, they are still opponents with the goal of eliminating all players and maximizing their own utility.

These actors are 'self-interested' in the sense that they aim to fulfill their own goal as effectively and efficiently as possible. Nonetheless, self-interested actors may have interest in working together. Cooperating may yield a larger profit than exploiting a purely competitive strategy. However, actors will only be keen to interact if they are sufficiently confident that a cooperation will be beneficial. In other words, cooperation will take place if there exists a positive *trust* relation between the actors. With 'trust' we refer to a degree of confidence about relying on someone else's actions. This degree can be increased or decreased by positive or negative experiences. Humans use trust as a basis for almost all their decisions and behavior[16, 21]. As Jonker et al.

put it: “Trust is omnipresent in all our interactions with people.”[43] In a world without trust, our society would collapse due to a lack of cooperation.

Ideally, people would interact with computer systems in a manner similar to the way that people interact with each other. In order to achieve this objective, we need to establish human conditions for cooperative behavior. Since trust is such an important notion in human interaction, it is also important to consider it when examining cooperations between humans and agents. Research in human-agent interaction can help us understand the conditions under which cooperation with agents is effective. Also, human responses to agents could reveal some basic conditions for cooperation. We will refer to the difference between human responses to agents and humans as a difference on basis of *nature*. The nature of an actor is, in our case, either 'human' or 'agent'.

In this thesis we will focus on these competitive interactions within heterogeneous groups of self-interested actors who may need to cooperate to perform a certain task. When three or more actors negotiate, an agreement does not necessarily include all participants. Instead, an agreement may be reached between a subset of all actors involved. Differences in capabilities, knowledge and resources imply that cooperation could lead to more beneficial results for the agents in question. In these scenarios, actors search for the most beneficial partners to carry out specific tasks. This cooperation method is called *coalition formation*. Coalitions are ‘temporary, means oriented, alliances among individuals or groups which differ in goals’ [29]. A more formal definition will be give in Chapter 2.

The central question we will address is: **How nature and trust influence people’s decisions in mixed coalition formations?** This question is motivated by the fact that humans are not fully rational when it comes to formation decisions [30]. There are several motivations that can lie behind cooperation, for example creating a band of trust, following social norms or displaying a sense of helpful behavior. Social considerations are shown to play a role in creating and maintaining coalitions: humans use principles of fairness, familiarity and trust whilst interacting with humans [14, 24, 53]. Additionally, social factors also play a role in human interactions with software agents. Recent work indicates that there is a human tendency to display the same social behavior towards agents and humans which dominates other behavioral conducts. It appears that humans apply the same social rules to computers as to humans [55, 74]. However, these studies examine only interactions in which the partner or opponent is predetermined. In this thesis we will extend these results to a scenario in which actors can choose their partners. When actors are allowed to choose certain aspects, such as the set of ‘acceptable behavior’ of an agent or human or the notion of in-group and out-group members, may become more dominant in their interactions.

Since social factors influence human decision making, decisions about whom to choose and how to split the payoff may be influenced by considerations about previous interactions. We will therefore also examine how previous interactions and trust relations affect coalition formation. We would also like to explore whether humans show different conducts of behavior towards agents and humans in a setting in which actors can choose between different partners as their team members. More specifically, we would like to investigate whether forming coalitions is influenced by the nature of the players (agent or human). To answer these questions, we will analyze who reaches an agreement with who in human-agent negotiations and how the agreement is reached.

So on the one hand coalition formation is motivated by social principles. Players use allocation norms in order to estimate the gain from joining a coalition. On the other hand, coalition formation is motivated by self-interest. This creates a social dilemma: if everyone cooperates, everyone is better off and obtains a higher payoff. However, defecting could lead to a short-term benefit for an individual. Once players start to defect, others will anticipate this by not cooperating. We will examine whether trust and the nature of others influence human behavior in these social dilemmas.

1.1 Hypotheses

Most research on coalition formation concentrates on either human coalition formation or agent coalition formation. The interaction between humans and agents in coalition formation has been mostly neglected. This thesis can be placed at the intersection of human coalition studies and coalition formation within multi-agent environments. We will use methods and results from both disciplines. We place self-interested agents and humans in the same domain to examine how cooperation evolves. Here we define 'cooperation' as the acceptance and proposing of splits of the coalition payoff.

This work focuses on one of the most important problems of coalition formation: how to reach an agreement on the allocation of the payoff [30]. This question poses several sub-questions, which we will address:

1. To what extent do trust and fairness influence team formation in mixed networks?
2. How does the nature of actors affect the way people relate to their actions?
3. Do actors develop stable relationships over time?

We will examine these questions in the light of the research question how nature and trust influence people's decisions in mixed interactions. We have developed the following hypotheses:

Hypothesis 1 A trust relationship between the actors affects the cooperation between them.

Hypothesis 2 The nature of other actors affects the cooperation between them.

As we will discuss in Chapter 3, prior work has shown that trust is an important factor for humans to base their decisions on. We suspect that our coalition formation scenario will stimulate the development of trust relations between both agents and humans. The setup uses iterated interactions to promote the development of trust relations and cooperation. The actors can perform several actions to individualize themselves. Their chosen actions make the difference between being more or less trustworthy. Actors can dynamically and simultaneously interact with each other and are allowed to defect from their commitments to a coalition. This creates a social tension between the actors. We suspect that actors with a stronger trust relation will remain committed to their joint task. We also suspect that people behave less cooperative towards actors who have proven to be untrustworthy.

We examine social dilemmas on several levels that relate to trust:

1. Will actors complete smaller tasks of lesser value with certainty or choose to form teams for completing larger tasks with associated uncertainty?
2. Do actors renege on a commitment to a team to obtain more lucrative deals?

Coalition partners may be interesting for different reasons: they can have beneficial capabilities or they can have a good history of cooperation and be very trustworthy. Another possibility is that *nature* (i.e., being a human or an agent) is the discriminating factor. We expect there is a difference between the payoff division offered to humans and to agents, given their nature and their interaction history. This hypothesis is bidirectional because no previous work has been done on choosing coalition members with a different nature. We suspect that nature makes a difference, such that humans can either prefer humans or agents for certain tasks. Perhaps humans prefer working with agents because they think they use the same strategy over a period of time, which makes them 'predictable'. Humans might also prefer agents because they expect agents to be cooperative and humans to be self-interested [34, 74]. We can also think of motivations for humans to prefer other humans. For one, they are part of the same 'species'. Second, it might be beneficial to team up with a human: they might be more inclined to offer fair proposals or to stay committed because of 'moral' obligations and expectations towards other humans. Whatever direction turns out to be the right one, we think nature is a discriminating factor in cooperation.

Although we stated our hypotheses as two separate ones, we are aware that trust and nature are sometimes intertwined. For example, a human might trust another human more than an agent, *because* he is human. In the analysis of our results we will take the interplay between trust and nature into account.

1.2 Research Method

The aim of this study is to conduct an empirical study to determine the factors that influence people's decisions in mixed interactions. Our research method is therefore a combination of constructive methods and empirical research. The research consists of three major parts: a literature study, the design and implementation of agents and the environment and the conducting of the experiment. First we will perform a literature study to get familiar with coalition formation theories and mainstream research within the field. As we gain more knowledge about methods and approaches of coalition formation analysis, we will shift our attention to more detailed work that directly relates to our work. This is reflected in the layout of the thesis. Once we are familiar with the relevant work, we will construct an environment that is suitable for our experiment. This needs to be a dynamic environment in which both agents and humans can interact and form groups. At this point we need to implement agents that are able to interact with humans in this domain. After the development of the agents and the environment, we will conduct an empirical study. Eventually we will evaluate this study by looking at the distribution of the payoff. The distribution of the payoff between the group members can provide us insights into the different social factors that play a role. For example, an equal distribution of the payoff between the coalition members may indicate that principles of fairness play a role in the allocation. The fact that actors can choose their partners and can exclude members from their coalitions, may influence the way they negotiate.

1.3 Contributions

A coalition is an alliance between several self-interested parties or persons, who cooperate in joint action, mostly for a short period of time. Basically it is a temporary team, with characteristics of responsibility, commitment, performance monitoring and within-team interdependency [23]. Although coalitions often allow their members to join and leave at any time, the members of a coalition are expected to act on behalf of everyone in the coalition. Cooperation is one of the most fundamental building blocks for successful interactions, both in multi-agent systems as in human society. For AI researchers, modeling cooperation in teams and coalitions is interesting for three main reasons:

1. **The development of agents that are able to interact and cooperate in an intelligent way on behalf of their users**
   Presently in Artificial Intelligence (AI) research, there is a great emphasis on designing agents that can support humans in everyday tasks or decisions. Examples of such agents are a pocket negotiator that negotiates on behalf of the human buyer [52] or a scheduling-agent for personal appointments [76]. These agents take over some of the tasks that humans otherwise would carry out. A model of human team work and cooperation can help to predict what the user would want.

2. **The development of agents that are able to interact and cooperate with humans in a manner similar to how humans would interact with other humans.**
   Agents engaged in group activity settings must choose among alternative actions and make decisions that appear realistic and are acceptable to the humans with whom they are interacting. Human cooperation models can provide the agents with information of how to react more human-like.

3. **The development of simulation environments for human trainees.**
   Models of human-agent coalition or team formation can be used to train humans to gain collaborative skills, to improve their decision making in complex tasks and for gathering information on how to build teams consisting of agents and humans who can work together to solve intelligence-intensive problems more efficiently. These models concepts and principles from various areas such as economics, philosophy, logic and the social sciences.

Our study makes several important contributions to the field of both behavioral and computational studies. First of all, we are one of the first to present a study of mixed coalition formation in which both agents and humans can take initiative. Examining these interactions is useful for creating models of human behavior. It enables agents to perform optimally, either in supporting humans or in out-performing humans.

Second, we present a coalition game where actors can choose their coalition members based not only on the proposed payoff distribution, but also on their history of interaction and the nature of the actor. We are not familiar with any study that lets humans choose with whom they would rather like to work: agents or other humans. The effects of this study has great implications for the design of software agents. Once we know the ways in which human behavior towards agent differs from their behavior towards humans, we can make the agent adaptive to these behaviors and so create agents that are more acceptable to humans [11].

Third, we provide an analysis of the interplay between trust and fairness and their effect on performance in mixed networks. Although trust and fairness are well studied in behavioral sciences, social dynamics between agents and humans are a relatively new field of interest. Our results will add some more understanding and knowledge of these dynamics to the field.

Lastly, we adapted a test-bed for decision-making to the study of fast-paced domains. The test-bed is now able to handle complex and dynamic domains where actors interact simultaneously. This adaptation can be used by other researchers to perform experiments on individual or team interactions.

1.4 Thesis Outline

The thesis follows the different stages of the research. In Chapter 2 and 3 we discuss literature that is relevant to our work. Chapter 2 presents a general overview of the field of coalition formation and of social dynamics in negotiation. In this chapter important definitions are presented that are used in the remainder of this thesis. Related work on both human and agent coalition formation, trust and human social considerations for agents can be found in Chapter 3. This chapter highlights work that is an important inspiration for our research. Chapter 4 and 5 address the design and implementation of our coalition analysis framework. In Chapter 4 we will present the conceptual design of the experiment. Here we present the framework Colored Trails and the Package Delivery domain. Chapter 5 takes a closer look at the experimental configuration and setup. In Chapter 6 we present the analysis and discussion of the results of the surveys and logs of the subjects. This chapter also offers our view on important challenges for our work and how the results contribute to research in the field. We will then in Chapter 7 discuss remaining open issues and how we would like to improve our research. Finally, we conclude the thesis in Chapter 8.

Chapter 2

Background

This chapter provides a brief introduction to the theories and formalizations we use throughout the thesis. We will introduce the language of games as provided by game theory. We will then in Chapter 4 define our domain using this language. The aim of this chapter is not to provide a full exposition of game theory or coalition theory, but to familiarize the reader with the terminology and concepts concerned with coalition formation.

2.1 Multi-Agent Systems

The term 'agent' it is used to indicate many different things. Within the research field of computer science and AI, 'agent' usually refers to a computer software entity. In other cases, it is used to describe all parties in a domain that are able to carry out actions, including humans. There is no universally adapted definition of the term. All that is agreed upon is that an agent has some degree of autonomy. To avoid confusion, we will use the term *actor* to refer to all entities in a domain that can make decisions and act upon them. So the actors in our domain are software agents and humans. We will reserve the term *agent* for intelligent software components that meet Definition 1.

Definition 1 (Agent)**.** An *agent* is a computer system that has the following properties (adopted from [75]):

1. situated in some environment
2. capable of autonomous action in this environment
3. reactiveness
   (the ability to perceive and respond to the environment)
4. proactiveness
   (the ability to exhibit goal-directed behavior by taking the initiative to satisfy objectives)
5. social abilities
   (capabilities to interact with other agents and humans)

A *Multi-Agent System* (MAS) is then a collection of agents that act and interact in an environment.

Definition 2 (Multi-Agent System)**.** A *Multi-Agent System* is a group of individual agents with the following properties [23]:

1. no explicit global control
2. distributed resource, expertise, intelligence, and processing capabilities
3. typically working in an open environment with many uncertainties
4. emphasis on social agency and social commitment

In typical MAS's, agents interact only with other agents. In this work we will look at systems where not only agents, but also humans will interact. More specifically, we will focus on *mixed-inititative* interactions.

2.2 Mixed-Initiative Interactions

There are many different dimensions of human-agent interactions. Three fields are developing in particular: human-centered teamwork, human-guided teamwork and mixed-initiative teamwork. The focus in the former two fields lies with the strengths and authority of the human actors involved. Humans are the crucial element and agents should be adapted to serve humans needs. This requires agents to have a detailed understanding of people's goals, desires and beliefs and how they use these to make decisions.

Mixed-initiative teamwork on the other hand addresses collaboration between humans and agents in order to benefit from the strengths of both parties. In these interactions, both agents and humans, as individuals or as groups, are autonomous and work alongside each other. They can take initiative and decide what to do next. In such environments, humans and agents may share goals or have conflicting goals, and they may collaborate or compete for resources. Actors often need to interact, cooperate and communicate in order to be successful. The big question is how they can do this optimally.

2.3 Game Theory

Game theory tries to answer this question. In order to study social phenomena such as coalition formation, social psychologists have over the years increasingly adopted an approach that has its roots in game theory. Game theory attempts to capture behavior in strategic situations in which an individual's success in making choices depends on the choices of others. If behaviors can be captured, social disciplines can use these strategic models to analyze and predict human behavior.

Game theory assumes that people behave and reason rationally. In this approach, parties are called *players* and the format in which these players interact are *games*. Moreover, the outcome that is obtained is usually a quantitative measure such as money or points. This outcome is referred to as the *payoff*.

2.3.1 Cooperativeness

Our work focuses on the interactions between *self-interested* actors and how they can form effective teams. Self-interested actors act primarily in their own interest and motivate their choices by trying to maximize this interest. The notion of a *utility function* is used in game theory to express differences in the actor's satisfaction. The utility is a real number that represents the level of contentment an actor has in that state of the world. The higher the utility, the better. The utility function defines this number over all possible states. The utility function is usually determined by payoff (e.g., money, points) but can also carry weight from for example social utilities (e.g. having friends, being a leader). We say that a self-interested actor wants to maximize its own utility. Note that this does not say anything about other actors' utility. It is not necessarily the case that a self-interested actor also wants to minimize the utility of other actors. Games where the focus lies on interactions between self-interested actors are called *non-cooperative* games.

Definition 3 (Non-cooperative games)**.** A game that is a competition between self-interested players and has the following properties:

- interactions take place between self-interested players
- the players can make non-binding agreements
- the basic modeling unit is the individual

Non-cooperative game theory is the study of games in which any cooperation that emerges is fully explained by the strategy the player deploys. However, when dealing with humans in complex domains, it might be difficult to trace their exact strategy. Cooperative game theory tries to find characterizations of what agreement the players will reach, instead of focusing on how they reach it. The actors in a coalition formation game typically interact in a scenario where they are able to communicate and cooperate. These are aspects of *cooperative* interactions. Therefore, coalition formation is traditionally categorized under cooperative game theory. It is important to note that the term 'cooperative' here does not mean that each actor is agreeable and follows all rules. Rather, it means that the basic modeling unit is the group rather than the individual actor.

Definition 4 (Cooperative games)**.** A game that is a competition between groups of players and has the following properties:

- the players make binding agreements
- communication and negotiation between the players is allowed
- groups of players may enforce cooperative behavior
- the basic modeling unit is the group

Although the cooperative games of coalition formation are our domain of interest, we will focus on the individual actor and their preferences and motivations. This indicates that we use aspects of both cooperative and non-cooperative game theory. Therefore, the interactions we are interested in are *hybrid* games: the coalitions are formed in a cooperative game, but the players play in a non-cooperative fashion. On the one hand parties need to cooperate in order to reach their goal. But on the other hand, they are in conflict because they need to allocate an outcome.

Different from non-cooperative motives, the motivations behind cooperative behavior are usually more difficult to capture. Exactly these motivations might cause people to care about fairness, trustworthiness and reputation. Game theory shows that people do not play completely rationally; social factors influence there decision making (see Section 3). This means that when analyzing a hybrid game, one has to look both at the strategic motivations players make as well as other motivations that might influence their game play. We will focus on two of these motivations: trust and nature.

2.4 Coalition Formation

2.4.1 Building Blocks

Coalition formation enables actors to jointly perform tasks that they would otherwise not be able to do, or would perform poorly. Coalition formation theory [29, 48] specifies the following conditions of a coalition situation:

Definition 5 (Coalitional situation)**.** A coalitional situation is one in which the following conditions are present [29]:

1. There is a decision to be made and there are more than two social units attempting to maximize their share of the payoffs.
2. No single alternative will maximize the payoff for all participants.
3. No participant has dictatorial powers, i.e., no one has initial resources sufficient to control the decision by himself.

4. No participant has veto power, i.e., no member *must* be included in every winning (successful) coalition.

In these situations, a subset of the players come together to form a *coalition*. A formal description of 'coalition' is given in Definition 6.

Definition 6 (Coalition)**.** A *coalition* is a subset S of the set of players N where [48]:

1. agreements take place involving approval by every player in S
2. no agreement between any other members of S and any member of $N - S$ is permitted

Coalitional theory identifies three different processes involved in group formation. In order to say anything meaningful about who will join with whom in any specific instance, these processes have to be modeled. To become part of successful coalitions, one has to deal with the following problems [30]:

1. how to identify the members with whom it is most likely to form a successful coalition

2. how to reach an agreement on the allocation of the payoffs

3. how to make a tradeoff between personal benefit and social strategy preferences (e.g., fairness, social welfare, care for out-group members)

The first problem addresses the issue of selecting the most appropriate agents to invite to the group. By avoiding players that are likely to defect, an actor increases the expected utility of a group. By avoiding players that are likely to reject offers, an actor maximizes the efficiency of the formation process. The second problem concerns the question of how much to offer a player to join a coalition and how large a payoff to accept to join a coalition. This process is put nicely into words by Gambarelli:

> "The process is developed as follows. Each player, in a preliminary examination of the game, asks the question: how much must I ask to take part in a coalition? If I ask too little, I won't optimize my win, while if I ask too much, the coalition will refuse me. More precisely, the more I ask, the less probability I have of belonging to a coalition which will actually be formed, up to the extreme case in which I don't belong to any such coalition and therefore I can obtain nothing." [28]

The third problem focuses on non-utilitarian strategy preferences. It recognizes that a coalition might yield payoff to more than just the members of a coalition. For instance, it can contribute to the overall welfare of a system. Furthermore, coalition members or the coalition as a whole might grant goods or privileges to out-group members.

2.4.2 Negotiation

Players who want to form coalitions have to make agreements on the distribution of the payoff. This can be considered a bargaining problem, which has been a challenging subject in game theory for a long time. The problem is stated in the following way:

> "Two individuals have before them several possible contractual agreements. Both have interests in reaching agreement but their interests are not entirely identical. What will be the agreed contract, assuming that both parties behave rationally?" [2]

The problem is how to reach an optimal solution when self-interested actors would to best to cooperate. The Ultimatum Game (UG) [35] is a commonly used game setting developed by experimental economists to simulate and analyze negotiation behavior. In the classic ultimatum game, two people are given the task to divide a sum of money. One player proposes how to divide the money between them and the other player can accept or reject this proposal. If the second player rejects, neither player receives anything. In the original game, the players interact only once. The rational solution is for player A to propose the total sum - k, where k is the smallest non-zero unit possible, and Party B should accept k. In this thesis we will treat the negotiation between coalition members as an UG.

2.4.3 Payoff

All coalitions have a guaranteed payoff that is specified by the game. This payoff is commonly defined as the *coalition value*, but in this thesis we will use 'payoff' because this is a general term that can also be applied to other areas in game theory. If the payoff may be freely distributed among the members of a coalition, the game is called a *coalitional game with transferable utility*. The definition of the coalition utility over all coalitions is known as the *characteristic function* of the game. The empty coalition has no utility since this coalition cannot be formed.

Definition 7 (Coalitional game with transferable utility)**.** A *coalitional game with transferable utility* is a pair (N, v) where [70]:

- N is the set of all players
- v is the characteristic function that associates with each coalition $S \subseteq N$ a real-valued value $v(S)$ that the coalition's members can distribute among themselves. We assume that $v(\emptyset) = 0$.

As an example of a characteristic function, consider the 'researcher's problem':

> "Imagine a situation where there are three researchers (players), A, B and C, each of whom has some resource needed to run and study. No researcher can run the study alone (so that $v(A) = v(B) = v(C)$), but any two of them can collaborate to run the study. Assume that the combined resources of A and B permit them to run the "best" study (say $v(AB) = 95$); that A and C run the next "best" one (let $v(AC) = 90$) and B and C the worst one (let $v(BC) = 65$). Assume further that the coalition ABC does not yield any value, so that $v(ABC) = 0$ (there is a limit on the number of researchers who can be involved). Thus, the characteristic function v has been defined completely, so the situation constitutes a coalition game. Now, researcher A must ask the question, "With whom should I propose a coalition, and how should I propose to allocate the resources assigned by v?"" [17]

In general it is advantageous to join some coalition, but it is up to each individual player to obtain some part of $v(S)$ as his own payoff. The transfer of payoff from one member of a coalition to another is called a *side payment*. Some coalitions may have a greater guaranteed payoff than others. If we assume rationality, players should be drawn to those coalitions that provide them the largest individual payoff. The property of *superadditivity* of a game ensures that the largest coalition that can be made has the highest payoff among all coalitional structures. This property specifies that the value of two disjoint coalitions is as least as high as the sum of their separate values.

Definition 8 (Superadditivity)**.** A game G = (N,v) is *superadditive* if $v(S \cup T) \geq v(S) + v(T)$ for all S, $T \in N$ such that $S \cap T = \emptyset$

2.5 Game Classification

Game theory distinguishes between different types of games. To be able to classify our game, we will discuss some of these distinctions.

A major dichotomy of games is whether or not they are *constant-sum*. In a constant-sum game, the sum of all players' payoffs is the same for any outcome. Hence, a gain for one participant is always at the expense of another. These are games of pure competition.

Definition 9 (Constant-sum games)**.** A game is *constant-sum* if when one player gains in proposing one outcome over another, the remaining players collectively lose. [48]

Definition 10 (Nonconstant-sum games)**.** A game is *nonconstant-sum* if the gains realized by one player when moving from one outcome to another need not be compensated for by losses of the other players. [48]

Cooperative games with transferable utility are said to be in *characteristic function form*. Most cooperative games are presented in the characteristic function form, while the *normal form* is used to define noncooperative games. The normal form is usually a matrix of players, strategies and payoffs. The characteristic function form uses the characteristic function to represent a game.

Definition 11 (Characteristic function form)**.** The *characteristic function form* is a game representation of a cooperative game where the characteristic function determines the payoff of each coalition and the empty coalition obtains a payoff of zero.

Definition 12 (Normal form)**.** The *normal form* represents a non-cooperative game by any function that associates a payoff for each player with every possible combination of action.

Another distinction can be made between *simple* and *multi-valued* games. Simple games attribute to the coalition a value of 1 in case of success and 0 in case of failure. This means that all successful coalitions obtain the same reward, as well as the unsuccessful ones. Multi-valued games on the other hand specify a different value for each possible coalition. We are interested in a restricted from of multi-valued games. Large coalitions obtain a higher payoff than smaller ones but the payoff of a coalition depends only on its size, not on who fulfills the positions within a coalition.

Experimental research based on game theoretic approaches often use one-shot negotiation games, since they are easy to explain and analyze. Many of the rational solutions found in game theory can be turned upside down by changing a *one-shot* game into an *iterated* game. In a one-shot game, players interact only once. Cooperation, fairness and equity are less prominent when the opponent has no opportunity to reciprocate or punish. If on the other hand players repeatedly play a game against each other where they, critically, can see each others' actions in all rounds, cooperating becomes the rational thing to do [7]. As Hinckley describes it, "from this shift in time perspective comes the importance of trust, future success, or past hostility, the importance of norms developed over time or else invoked to keep the group together". [36]

2.6 Social Factors

Individuals mindlessly apply social rules and expectations to computers [56]. In this section we take a closer look at some of these social behaviors.

2.6.1 Trust

In human-agent interactions, trust is of particular importance. When actors decide to cooperate, the responsibility for achieving a goal is distributed over the parties involved. Cooperation always involves a degree of risk arising from the uncertainties of interacting with autonomous and self-interested actors. In environments where coalitions formation is important and defections are possible, the use of trust allows social norms to be dened and compared. An actor can use this information when invited to join a group or coalition, to decide whether or not its utility will be increased by joining. By avoiding those who are likely to defect, actors can maximize the efficiency of the formation process. If achieving the goal has consequences of importance, humans are often not very willing to transfer responsibility of executing the task properly. They want to minimize the risk. Trust represents the actor's assessment of the risk that the involved party does not fulfill its commitments [16, 22]. As long as humans remain responsible for the outcomes, they want effective authority and the means to influence the performance [18]. Accordingly, they will only be inclined to cooperate if they sufficiently trust the actors involved. This is the case in human-human interaction, but even more so in human-agent interaction since agents never have final responsibility. Humans remain responsible for the actions of their agents. As agents are becoming increasingly more autonomous, trust is becoming a very relevant issue. For example, a trust relation between the user and his personal assistant agent (and, in general, his computer) is important because the user wants his assistant to make decisions in his best interest. As agents become increasingly autonomous in crisis-management, trusting the agent to make correct decisions on behalf of humans is crucial.

Definition 13 (Trust)**.** For an actor a to be said to trust another actor b with respect to a particular goal g, a must have the following beliefs [16, 22]:

- **Competence Belief**:
  b is useful for achieving g and is able to provide the expected result
- **Disposition Belief**:
  b is not only capable, but also willing to do what is necessary to achieve g
- **Dependence Belief**:
  the results and rewards of achieving g depend on the involvement of b
- **Fulfillment Belief**:
  g will come about due to b's involvement

Trust can be divided into two main categories: *experience-based* trust and *recommendation-based* trust [31]. If the actor uses experience-based trust, it solely relies on its own experiences. According to whether this experience was a *trust-negative* or a *trust-positive* experience [44], the agent can update its risk assessment. Actors can use recommendation-based trust to find out from others how trustworthy the actor in question is perceived to be.

2.6.2 Fairness

The most straightforward answer to the question of how payoffs should be divided amongst the coalition members is that the split should be *fair*. In this work, we will use the following definition for 'fair':

Definition 14 (Fair)**.** A *fair* split is a distribution of the payoff that is considered socially just.

Traditionally, economic models assume that people are exclusively pursuing their material self-interest and do not care about social considerations. Recent research puts more emphasis on the fact that people feel very strongly about principles of fairness, especially in repeated interactions. Similar to insights based on equity theory and distributive justice, players in a coalition game seem to demand and expect a share of the reward that corresponds to their source of entitlement [1, 73].

In game theory, the *Shapley value* [69] defines a fair distribution of payoff that is obtained by cooperative action. The Shapley value assigns a value to all members that can be added to a coalition. This value is the average of all contributions over all possible ways that a coalition can be formed with that player in it.

2.6.3 Social Dilemma

When groups, teams or organizations interact, there is often a tradeoff that needs to be made between individual utility and a global/collective interest. Definition 15 gives two general properties of social dilemmas.

Definition 15 (Social dilemma)**.** A social dilemma is a situation in which [20]:

1. each actor receives a higher payoff for a socially defecting choice than for a socially cooperative choice, no matter what other individuals in society do, but
2. all individuals are better off if all cooperate than if all defect

These social dilemmas are abundant in everyday life. A students who needs to hand in a homework assignment would prefer to copy everything from his friend in stead of spending hours trying to figure it out by himself. But if everyone would do so, no-one would come up with the answer. A soldier who fights in a large battle can reasonably conclude that no matter what his comrades do he is personally better off to hide and safe himself; yet if no one takes chances, the result will be a certain death for everyone.

Literature divides social dilemmas into two groups:

1. social dilemmas in which a common resource needs to be distributed over a group of actors (example: Ultimatum Game)

2. social dilemmas in which every agent has to invest a cost in order to obtain a higher payoff, given that everyone invests (example: Public Goods Game, Prisoners Dilemma)

In this thesis, we will study a domain that entails both dilemmas.

Now that we have presented the basic concepts we will use in this work, we move on to describe the most relevant literature in more detail in the next chapter.

Chapter 3

Related Work

This thesis combines several issues that have only been studied in isolation within the behavioral and computational sciences. We look at social dynamics that emerge in the formation and negotiation process of coalitions between humans *and* agents. Importantly, we look at this in the light of how coalition members are chosen and how trust and nature influence this decision. As far as we know, no previous work focuses on the formation of coalitions with human and agent members who can freely choose their partners. In this chapter, we will highlight research that has been most relevant for our work and point out how this work differs from our own. We will discuss previous work on coalition formation and social dynamics in group interactions along the lines of computational and behavioral sciences.

3.1 Coalition Formation

Social psychologists have been interested in the subject of coalition formation for many decades (see e.g., [3, 48, 57]). Coalition formation entails topics like negotiation, cooperation, selfish behavior and interaction within a team. Von Neumann and Morgenstern [57] state in their influential book on game theory dating from 1944 that "[o]ur [. . .] discussion of "games of strategy" will show that the role and size of "coalitions" is decisive throughout the entire subject." In later years, Kahan and Rapoport extensively studied and formalized human coalition formation [46, 48]. They used human subjects in a computerized experiment to test human behavior in coalition formation games and compared actual human performance with the theoretical predictions from solution concepts found in game theory[1]. Traditionally, cooperative game theory uses these solution concepts and values such as the Shapley value (see Chapter 2) to specify normative predictions for coalition formation given that players follow a set of rationality conditions [17, 29]. However, these concepts do not do well when studying whether the human mechanism really does work in the way a cooperative solution concepts work [8]. Although Kahan and Rapoport found that human behavior did not significantly deviate from game theory predictions, we cannot to use game theory to analyze our coalition formation interactions. In the present study, all coalition members have the opportunity to defect from their coalition which creates a social dilemma: they can choose between staying committed to the group or pursuing and maximizing their own profit. What is more, in our study subjects can choose to not cooperate at all while still obtaining payoff. For our experiment we will design a coalition game were we allow (i) single players to have non-zero values and (ii) imperfect information,

[1] An example of such solutions are *quota solutions*, which distinguishes coalitions by their size and the reward available to them. For a more in-depth discussion of quota games and quota solutions, refer to [48].

conditions that are not captured in traditional game theory.[2] We will thus need to rely on more social-psychological models to determine how these factors influence human and agent behavior in coalition formation interactions.

Over the years, researchers have discovered many interesting phenomena that occur in human coalition formation. These phenomena can be divided into two main categories: (i) fairness considerations greatly influence the human coalition formation and negotiation process [45, 61]; and (ii) previous experiences strengthen or weaken a trust relation between coalition members [37]. These findings stress the fact that humans are not the rational players game theory sometimes assume they are. In 'Rules of Encounter', Rosenschein and Zlotkin [64] emphasize that transferring utility through side payments induces promises and commitments between actors to perform future actions. These social commitments blur rationality. For example, literature shows that in 2- and 3-player coalition games, people respond by acting negatively reciprocally. They punish proposers by rejecting unfair offers [4, 58]. Furthermore, when selecting future group members people are biased towards actors with whom they have developed strong working relationships in the past [38]. We will use these results to analyze whether humans display the same behavior towards agents as they do to humans. We hypothesize that if they can choose between humans and agents, trust and fairness will be more important in their relation towards humans than it is towards agents.

Another well-known phenomenon within behavioral studies is that people favor other people who they perceive to be in the same group [6]. Self-oriented actors, that is, actors focused on the improvement of their own payoff, are more likely to exclude out-group members than in-group members. Such processes are especially pronounced in inter-group settings [5]. Results from [56] suggest that people tend to rely on social groups when interacting with computers. These social groups can be defined by any common property, such as the color of a team or a number. In these experiments, humans pair up with computers as easily as they do with other humans. However, these studies use predefined groups, with assigned properties of categorization. The current work explores whether these results hold in relation to groups of agents and humans when humans are allowed to choose their group members. It might be the case that people have more affiliation with members of the group with the property of 'being human' than with members of the group with the property 'in the same coalition'. If this is the case, this creates a social dilemma when humans have to choose between commitments between out-group humans and in-group agents.

Inter-agent coalition formation in MAS's traditionally focuses on the same game theoretic approaches, looking for optimal and efficient solutions. However, optimal rationalizations may not always be possible due to noise, uncertainty or time constraints. Especially in a domain where the agents are required to interact with humans or other agents. In a real-world domain, the agent should be highly autonomous and adaptive in order to overcome these challenges.

> "Coalition formation mechanisms proposed to date [...] commonly provide [strategies and protocols for the agents], however they include several restrictive assumptions, which do not hold in real-world domains where coalitions are necessary. [...] Under the assumptions of incomplete information, heterogeneous task valuations, and short time for completion of the coalition process, traditional coalition mechanisms are inadequate." [51]

In summary, game theory research on coalitions has developed methods for formal analysis concerning issues of solution stability, fairness, and payoff allocations. However, the methods are abstracted and some of the underlying assumptions of the developed algorithms do not necessarily

[2] In their '77 paper, Kahan and Rapoport try but fail to extend their findings to the case where 1-person coalitions are allowed to have non-zero values [47]. Coalition structures appear to be dependent on the presence or absence of nonzero payoffs for 1-person, 2-person, and 3-person coalitions in 3-person games. However, it is not clear how they are dependent.

TABLE 3.1: Common Assumptions in Coalition Formation

Interchangeability

Previous Work	All potential members bring the same utility to a coalition (e.g., [12, 13]).
Our Work	**Some potential members are preferred over others.**
Motivation	In complex domains, or domains involving humans, this is generally not the case. Some members are more valuable to the coalition than others. When throwing a dinner party, someone who has good cooking skills or handy cookware is more valuable than someone who has never cooked before. Coalition members can have preference relations about who to work with. Doing homework exercises is probably most efficient with students that are very bright or with the ones you have worked with before. Some individuals are preferred over others because they have for example a better skill set or they have shown to be trustworthy.

Membership

Previous Work	Membership to a coalition is equally available to everyone (e.g., [12, 67]).
Our Work	**In order for membership to become available, there are some requirements members have to meet.**
Motivation	Coalitions often require members with certain capabilities or resources. To become a member of the Christian Coalition of America, one has to be a firm believer of the fundamentals of Christianity. Not everyone is automatically eligible to be part of the coalition.

Conflict

Previous Work	Conflict is eliminated by making agreements binding (e.g., [58, 72]).
Our Work	**No conflict-free environment since agreements are not binding.**
Motivation	In a dynamic world, coalitions change all the time. New members are invited and others are cast out. Often there exist competing coalitions who try to attract the same agents to their group. People for example change jobs in case they get a better job offer. Actors may have agreements and commitments to one coalition, but may break these to join another.

hold in real-world systems. Current work on coalition formation makes some strong but common assumptions. These assumptions often come in handy to analyze coalition structures, but are not very realistic. In our work, we try to model coalition formation in a more real-world environment, disregarding these assumptions. Our domain will use for example incomplete and uncertain information. In Table 3.1 we shortly describe the assumptions that are usually adopted and our motivation to discard them.

3.2 Social Dynamics

In this study we use an Ultimatum Game with repeated interaction to let the members of a coalition agree on a split of the coalition value. The setup of the game is very suitable to study social relations such as trust and fairness between different actors. In the traditional UG, people usually offer 'fair' (i.e., 50%:50%) splits, and most offers of less than 20% are rejected. This latter phenomenon is referred to as *inequity aversion*. Social factors, such as altruism, fairness and reciprocity are shown to have an influence on negotiation behavior (see e.g. [14, 53, 65]). Recall that it is a property of coalitions with transferable utility that the coalition members have to agree on a split of the utility (see Chapter 2). The members thus have to negotiate. These negotiations can be considered as an UG, where one of the members proposes the split and other can accept or reject. We use the UG to study trust and fairness in this negotiation behavior. We will examine how these behaviors change over time and how they differ between agents and humans. As far as we know, we are the first to use the Ultimatum Game in a coalition formation setting.

We allow for coalitions consisting of up to three actors, which means the bargaining over the side payoff will also involve 3 players. Only a few experimental studies extend the UG to more than two players and even fewer use it in a coalition formation setting. Okana and Riedl do use a 3-player UG in a Coalition Formation problem [58]. They found that in 2-person coalitions as wel as in 3-person coalitions responders behave negatively reciprocally. They punish proposers by rejecting positive but unfair offers. We will extend this line of research by looking at repeated interactions and different natures of opponents. The studies that have been performed show that in three player UG's new social considerations come into play. The presence of a third player has an effect on the bargaining process: humans make social comparisons between responders [63]. The relative payoffs between members clearly matter: there exists a strong correlation between relative payoffs and rejection behavior [10].

Within the field of human-agent interaction there exist contradictory findings when it comes to human conduct towards opponents of different natures. On one hand, people appear to display the same social behavior towards agents as towards humans. For example, human responders care about equality of outcomes while negotiating with agents [27]. On the other hand, experiments indicate that a significant difference exists between the behavior of humans towards agent and human opponents. People appear to perceive other humans differently from agents [9, 66]. Results from these works show that people have a different set of 'acceptable conducts' when working with people, versus working with agents. This difference in human behavior poses the problem of how humans actively treat agents while interacting with them.

Reeves and Nass were one of the first to address the issue of how humans treat computers and other media [62]. Through a series of experiments they showed that human behavior towards media is in many cases the same as towards other human beings. This phenomenon is called 'the Media Equation'. It is based on the thought that if interacting with a computer can be similar to interacting with another human, we can also expect certain social psychological dynamics from human-human interaction to apply to human-computer interaction. In media equation studies, the social dynamics surrounding human-human interactions are shown to exist in human-computer interactions (see e.g., [74]). Nass et al.[55] performed several experiments to demonstrate that humans affiliate with computers as a team in the same way they affiliate with human team members. As it turned out, making a human's performance dependent on a computer's performance (or even just saying this is the case), triggers perceptions of team affiliation.[3]

We use Nass' findings as a foundation for our research. We expect to find similar social psychological dynamics to occur between humans and agent as were found in the studies mentioned above. However, our focus will be on the changes in these dynamics when players are allowed to choose between humans and agents. This has not been investigated before. Although it can be the case that humans have similar attitudes towards agents when they are dependent on them or if they have no other choice than to work with them, we suspect the differences between human conducts towards humans and agents are strengthened when they can choose between them.

As pointed out in Section 3.1, multi-agent interactions are commonly concerned with optimizing strategies. These strategies may perform well in certain scenarios, but do not necessarily cause human-like behavior. Literature has shown that it can be advantageous for agents to implement a model with social principles. Models that take principles such as fairness and helpfulness into account were shown to explore new negotiation opportunities [40] and to find solutions that correspond to solutions found by humans [42]. Several approaches use inter-trust relations

[3] An elaborate introduction of the media equation and an overview of literature in the field can be found in [62]

between agents to try to optimize coalition formation (e.g., [12, 22]. Griffiths and Luck [32] use trust to decide which agents are most likely to be good group members, and use what is known about the agent's motivations to decide whether or not to join the group. Jones et al. [41] have explored a coalition formation setting where agents can pursue partners of varying trustworthiness and lure them away from their current coalition. This means they do implement a social dilemma. Their results show that in some circumstances an agent may profit from selecting less trustworthy partners, since they are more likely to defect from an existing team and subsequently form a successful coalition with the agent.

Although these studies show that agents that implement social dynamics often perform better than those who use a purely computational strategy, they have not used these strategies for interactions with humans. The works mentioned above use trust in a stable and limited environment. In this work, we want to create a dynamic environment where both agents and humans have to adapt to changes in the environment and changes in the strategy of other actors.

Chapter 4

Conceptual Approach

In this chapter we will discuss the domain we have chosen to implement. It is a mixed-initiative coalition game called the Package Delivery Domain. We will also introduce Colored Trails: a framework for examining decision making processes in multi actor interactions. Colored Trails is very suitable to implement our coalition game. We will present the rules of the game, as well as our motivations for the implementation of the rules.

4.1 The Package Delivery Domain

The Package Delivery Domain (PDD) is very suitable as a coalition formation game (see [67, 72]). It requires actors to have individual goals and to cooperate in order to satisfy their goals. In the case that cooperation is required, agents must decide when and who to ask for help and when and who to grant a favor. We will extend the domain in such a way that it allows us to examine human-agent coalition formation.

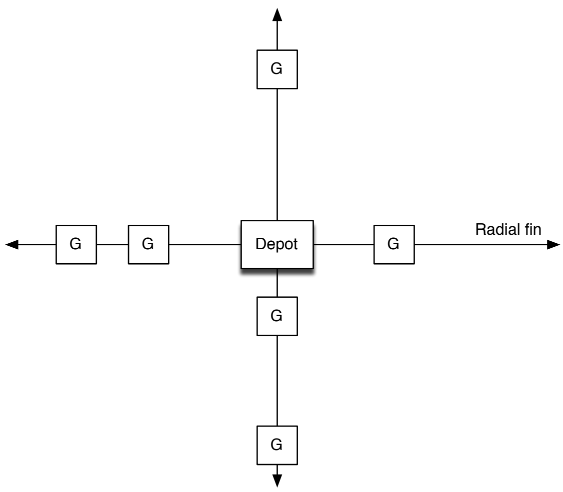


FIGURE 4.1: The original Package Delivery Domain by Sen.

In the original PDD [67] N agents are designed to deliver T packages each. All packages must be picked up from a central depot. On arrival at the depot, an agent is assigned a package that it has to deliver. The package destinations are located on one of R different radial fins (axes) at a distance between 1 and D from the depot. In Figure 4.1, these destinations are marked as a 'G'. Movement to and from the goal is only allowed via the different fins. It is not possible to move directly between the fins. The agent at the depot can decide to either deliver the package itself or to ask for help. An agent can ask for help from another agent who is currently also at the depot and who is assigned a package that has to be delivered on the same fin. However, there is a cost involved for helping other agents, which is one unit of extra cost per unit distance traveled when carrying the extra package.

This domain was designed to examine interactions between self-motivated agents that must adapt their behavior depending on the behavior of other agents in the environment. Sen and Dutta [68] use the domain in order to identify dominant strategies under different environmental conditions. They use mixed groups of agents with exploitative, reciprocative and philanthropic strategies to understand the dynamics in this more realistic scenario.

4.2 Colored Trails

We designed a coalition formation game that was inspired by Sen's Package Delivery Domain. For the implementation of our game we needed a flexible framework that provides an open architecture for actors. Colored Trails proved to be a very suitable framework for this purpose.

4.2.1 A Testbed for Investigating Decision Making

Colored Trails (CT) [34] is a testbed developed by Grosz and Kraus for the purpose of investigating decision making that arises when agents interact in task specific settings. The testbed consists of a game than can be played by multiple players whose task it is to reach a goal by exchanging resources. These interactions provide the basis for modeling, comparing and testing the performance of humans and computational strategies deployed by software agents. Colored Trails is especially suitable for our research, since the architecture allows the game to be played by groups of people, computer agents, or heterogeneous mixes of the two. CT enables us to examine the behavior of self-interested agents and humans in group settings.

> "A key determinant of CT design was the goal of providing a vehicle for comparing the decision-making strategies people deploy when they interact with other people with those they deploy when computer systems are members of their groups." [34]

Another important feature of CT is that it enables the experimenter to hide the nature of players. It is possible to deny the participant information about whether he is playing against an agent or a human.

A variety of domains can be implemented in CT, ranging from abstract to complex. It is very well-suited to implement popular games from behavioral economics (e.g., Prisoners' Dilemma, Public Goods Game, Ultimatum Game, Dictator Game) [25, 34, 74], but it also serves as a basis for more real-world scenarios because of the clear analogues to task settings and the possibility to provide situational contexts. The game parameters can be set to represent environmental features, such as reward structures, dependencies between agents, resources and skills. We will elaborate on how we used these parameters in Chapter 5.

We will briefly describe some of the prior experiment conducted with CT, to give an impression of the range of possibilities. CT was first used in an experiment to compare agent's decision making strategies with those of humans [34]. The game was set up as a simple negotiation game, in which players could exchange resources by making non-binding agreements. In recent work, CT was used to implement a very different game to test agent strategies for deciding whether to help other members of a group whose members are engaged in a collaborative activity [49]. In this game players could negotiate to help a member of a group reach his goal.

We will use CT to design a game that differs significantly from games that have been implemented in CT thus far. Our game is more dynamic, interactive and complex than previous games. We will give the details of our setup later in this chapter.

4.2.2 Components

CT can be divided into two parts: (i) fixed components that provide the player with the basic functionality and (ii) the game rules and configuration as implemented by the programmer. Figure 4.2 shows the fixed components as they appear in a basic CT configuration. In this section we will elaborate on each of the components.

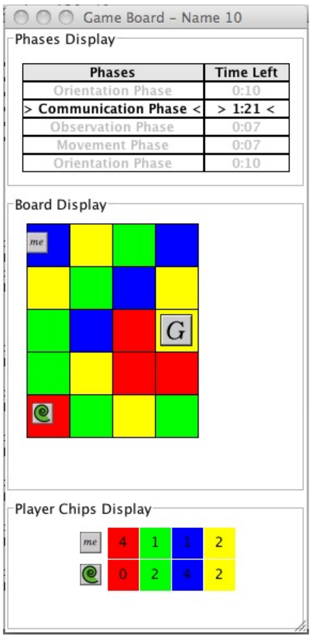


FIGURE 4.2: Basic components of Colored Trails: phases, board and chips

Board CT is played on an N×M board of colored squares in which one or more squares can be designated as a goal. In order to move to an adjacent position, the player has to hand in a chip that has the same color as the square he[1] wants to move to. It is not possible to make diagonal moves. Players are allowed to occupy the same square. The path a player chooses to reach the goal represents the plan or recipe an actor has in the world to fulfill a goal, where each square represents an individual task.

Chips Players can reach the goal by exchanging chips. Each player is given a set of chips with colors taken from the same palette as the squares of the board. The different colors correspond to different capabilities or resources the actors have. The number of chips represent the number of resources every player has at his disposal.

Phases The game can be set up to contain several phases that specify which actions may be undertaken by the players. This ensures both a clear structure and good playability of the game. For example, during a 'Communication Phase', players are allowed to make proposals and respond to them but they are not allowed to move. The players are permitted to move during the 'Movement Phase'.

[1]The players in this game can be men, women or computer agents. For purposes of consistency and clarity, we will refer to all players as male. This is however completely arbitrary.

Proposals Players can negotiate to obtain the chips they need for reaching the goal. The exchange of chips corresponds to actors providing resources to each other, performing tasks for them, or enabling them to do tasks they otherwise could not do. The proposals and responses are the only way the players can communicate. This is a fixed but expressive messaging protocol.

Anyone who wants to create a CT game, can manipulate these components in several ways. These manipulations are basically the rules and configuration of the game. Players can for example have different or multiple goals. The scoring function determines the payoff for the individual players and can be completely adjusted by the programmer. It can for example be composed of several different factors such as the Manhattan distance of the player to the goal. It can also depend on whether the player reached the goal or the number of chips left in possession of the agent. The scoring function can also reflect more complex utilities such as the social welfare of the group or the completion of certain subtasks.

There are several ways in which the game gives the programmer the ability to promote interdependency of players. If players are *task-dependent*, they are depending on other players supplying the chips they need for the performance of their own task (reaching the goal). This generates a social dependency because the score of one player depends on the helpful behavior of another. A second dependency can be created by making the players *reward-dependent*. This indicates that the score of one player depends in some way on the score of the other players. The scoring function thereby strengthens or weakens the competitive relation between the proposer and the responder. For example, by varying the relative weights of individual and group goods in the scoring function we can make collaborative behavior become more or less beneficial.

The visibility the players have of the board, of the other players and of the chipsets corresponds with different knowledge conditions. If the opponent's goal is hidden on the board, then this corresponds to being unable to know the intentions and goals of the other player. Concealing the chips of other players corresponds to not having any knowledge about their capabilities.

4.3 Conceptual Design

4.3.1 Adapting the PDD: MIPDD

We adapt the PDD from Sen to create a coalitional game with transferable utility containing resources, players and capabilities. This is done in order to create realistic interaction situations. We call this domain the *MIPDD*: Mixed Initiative Package Delivery Domain. In this name the original domain of inspiration as well as the influence of both agents and humans are captured.

In the MIPDD game, N players (humans and agents) aim to maximize their utility, given a time limit that is unknown to them. A player can increase its utility by delivering packages to the central depot. The players are randomly placed on the board, as are the packages. The packages are not assigned to particular players, instead, each player can walk around freely and collect either a small package (SP) or form a coalition to collect a large package (LP). Figure 4.3 shows a simple schematic representation of the domain.

The fact that our game has both cooperative and non-cooperative aspects is made explicit in the payoff distribution. In cooperative games payoffs usually are not given for individual players, but only for coalitions. In our scenario, coalitions are indeed rewarded with a payoff. However, individual players are also able to obtain payoff, be it a significantly smaller one. Any player can deliver a small package on its own in exchange for a small payoff. Though the payoff of large packages is higher, the cost is also higher because large packages can only be delivered with the help of at least one other player. Each player can only (help to) deliver one package at the time. Although it is not obligatory for players to form coalitions, it does pay off to be in a (large) coalition because they yield a higher payoff.

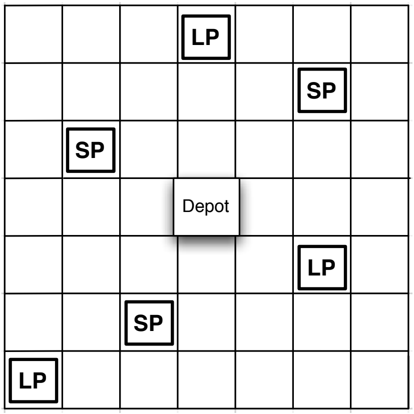


FIGURE 4.3: A schematic representation of MIPDD.

Within the game, we distinguish between *initiators* and *members*. The initiator is allowed to shape the process of negotiation by choosing the players he would like to join his coalition. The initiator also decides how much of the total payoff he offers the members. He can propose the split to the members. The proposal takes the form of one round of the Ultimatum Game: the initiator proposes a split of the playoff and the member can accept or reject. If the member rejects, both players get nothing. In this sense, it is a one-shot constant-sum negotiation (as defined in Section 2.5). Everything initiator A proposes to keep to himself, will be subtracted from the payoff member B will receive. Although the negotiations about the payoff are one-shot, during the game the players interact and negotiate repeatedly. This causes trust and social commitment to become more important. The game as a whole is a nonconstant-sum game, since players can deliver packages and increase their payoff without decreasing other players' payoff.

4.3.2 Agent Implementation

As stated in Chapter 1, one of our main goals in this work is to investigate whether humans have a preference for working together with other humans or with agents. Importantly, we want the agents to be as human-like as possible, since we can only say something meaningful about preferences on the basis of nature when the actors display very similar behavior. In the case humans and agents have very different strategies and conduct, any preference might be related to this difference. Ideally we would create a model of human behavior in our domain and use this to develop our agents. This would ensure that our agents are equipped with advanced strategies that capture social considerations in stead of simple algorithmic strategies which might be easy to identify and predict.

However, no matter how the agent strategy is designed, any differences in preference could still be traced back to the implementation of the agents. We therefore chose an alternative approach that ensures that humans and agents have similar strategies. To each player we assigned a distribution of how they perceived each other. Given the set N of all players in the game, we created a mapping M for each player $p \in N$ that p's perception of every other player in N to either a human or a computer agent: $M\colon N \to \{\text{h,c}\}$. Even though the subjects played the game solely against each other, they randomly perceived half of the players as human and the other half as agent. Naturally, they always perceived themselves as human. Each game, these pairings are randomly assigned. So when a subject perceives another subject as 'human' in one round, he can perceive him as 'agent' the next round. Also, the pairings were non-symmetrical: subject 1 could perceive subject 2 as an agent while subject 2 perceived subject 1 as a human. Importantly, we did not tell them that they played solely with humans. We deceived them into believing they were playing against both humans and agents.

4.3.3 Knowledge constraints

The knowledge available to the actors greatly influences their performance. For instance, if someone knows he is the only candidate for a job position, he will be self-confident at the interview and maybe not try his very best to impress. However, if he knows he has to compete with 20 other applicants, he will try very hard to distinguish himself and impress the interviewer. Game theory uses the terms *imperfect* and *incomplete information* to distinguish between games with different knowledge constraints for the players [60]. Imperfect information relates to a state of knowledge about the behavior of the players, whereas incomplete information refers to knowledge about the structure and rules of the game (such as the payoff function and the number of players).

Definition 16 (Incomplete information)**.** A state of knowledge where the player is uncertain about the behavior of the other players.

Definition 17 (Imperfect information)**.** A state of knowledge where the player is uncertain about some of the elements which define the rules of the game.

In our experiment, players have complete information in the context of the game. The game is however a game of imperfect information because the players do not know all actions of the other players. Consequently, actors can only use experience based trust to create an expectation of a player's trustworthiness. This is not a reputation game where actors can use recommendation based trust to update their trust value of another actor.

We created a setup with imperfect information to (i) prevent players from being overwhelmed with information and (ii) to limit the variables that can influence decision making. The game is highly complex because it is played with multiple players simultaneously and players' actions are not restricted by different phases of the game. Because our objective is to create a less abstract and more dynamic environment than is used in other CT games, phases do not play a role in our setup. The players continuously interact with the environment and the other players. To help players focus on the task at hand, namely choosing coalition partners, we limited their knowledge:

- Players have no information about other players' locations on the board, except for the ones that are in their coalition.
- Players view no coalitions and proposals other then the ones they are involved in.
- In three-player coalitions, invited members do not get to see how much payoff the third party receives.

During the pilots we noticed that it was distracting and stressful to see all players moving around, making coalitions and delivering packages. It was a logical choice to remove other player's positions, since in realistic situations actors often do not have full knowledge about the actions and whereabouts of other actors. This implementation choice makes the game clearer and more realistic. Another reason for introducing uncertainty is that we focus on nature and trust as factors in the decision process. Removing the position of other players excludes proximity as a possible reason for inviting players to a coalition.

The fact that other coalitions and proposals do not appear ensures that the player is focused on maximizing its own payoff instead of being affected by other players' progress.

We also did not display third party relative payoffs. That is, in the UG in which players negotiate over the payoff distribution, only the payoff of the initiator and one member can be compared. The reason we chose to limit knowledge about this is that, as we pointed out in Chapter 3, the presence of a third player has a clear effect on the bargaining process. Players already have to take several variables into account, such as offers from different players, their preference for the player the proposal originates from and coalitions of different sizes. We ruled out the effect of third player presence in order to analyze their negotiation behavior solely in these terms of trust and nature without having to account for differences due to third party involvement.

4.3.4 Choosing Team members

In Chapter 3, we addressed several assumptions often made in coalition formation theory. These assumptions restrict the domain, making it more abstract and easier to formalize. However, we created a more dynamic domain by loosening these constraints. We have realized this in the following way:

- **Members of coalitions are not defined:** Coalitions are not imposed, each actor has the freedom of choosing which coalitions to join and who he wants to invite to his own coalition.
- **Some potential members are preferred over others:** The actors have different natures. they are either human or agent. Nature is a discriminating factor on basis of which an actor can choose its partner. On top of that, the actors interact repeatedly. Accordingly, actors may favor other actors they successfully interacted with in the past.
- **Actors have to meet requirements in order to become members:** In order to create a coalition, the members of the coalition must have colors that correspond to the colors of the path to the goal. Therefore, only actors with a required color are eligible as a member of that particular coalition.
- **Agreements are non-binding:** An actor who is already involved in a coalition can join or create another coalition and thereby defect from his current coalition.

4.3.5 Defection

The criteria humans use to discriminate between their possible partners are of great interest to us. We stressed the importance of freedom of choosing team members in realistic scenarios. In our setup we use nature as one of the distinguishing factors between actors. We have also argued that repeated interactions can help to develop trust relations, which in turn can be a distinguishing factor as well. To stimulate actors to develop these trust relations even further, the game allows for non-binding agreements. In other words, players are not forced to commit to their obligations to the group. As these agents are self interested, it may be reasonable for an agent to quit one coalition and to join another that provides a higher payoff. Importantly, *players cannot be part of more than one coalition at the time.*

It is one thing for the game to allow defections, it is another for the team members to allow this. The members might not accept defection and this may show in their future interactions with the player that defected. We suspect defections to have an influence on the dynamics of a trust relation. When defection is allowed, being part of a coalition is not always beneficial for an actor. The cost to join a coalitions has increased because there is a risk that the coalition will break up despite the actor's effort. The actors must manage the risk associated with interacting with others who may have different objectives, or who may fail to fulfill their commitment. The success of a coalition is not entirely dependent on the actor anymore, but instead is a shared responsibility of all members in that coalition. Therefore, players can use information about another player's defections to strengthen or weaken their trust- and preference-relation for a that player. Risk-averse actors may find the cost for joining a coalition exceeding the expected payoff they receive for successful delivery [32]. In short, our domain fulfills the two requirements of a social dilemma: defecting is individually beneficial, but all are better off if all cooperate.

Chapter 5

Experimental Design

We have created a Mixed Initiative Package Delivery Domain (MIPDD), inspired by Sen's Package Delivery Problem [67]. This domain has different objectives than the original domain, but they still share the main characteristics that stimulate cooperation between competitive agents. Self-interested actors can choose to form coalitions that could increase their utility. The objective of our game differs from the original one since our purpose is to examine human-agent interaction, a subject not addressed by Sen. His implementation of the domain is not designed to deal with human decision making. Our implementation is more dynamic and stimulates frequent interactions between different actors. We performed several pilot experiments to shape the experiment after our requirements and to find a baseline for the values of our variables. In this chapter we will describe the implementation details of MIPDD and illustrate the setup of the experiment.

5.1 The Domain

In this experiment each game had 6 subjects interact simultaneously. As explained in Section 4.3.5, each subject randomly perceives half of the participants as agents and the other half, including himself, as human. From now on we will say that the game was played with 3 agents and 3 humans. In a game with n players, each tile on the board can have one of $n/2$ colors. With 6 players, there are 3 colors in every game: red, green and blue. The board is constructed with these colors in a random distribution of colored squares. Each color is represented by both a human and an agent. So, in each game there are 2 red players (1 agent, 1 human), 2 green players (1 agent, 1 human) and 2 blue players (1 agent, 1 human). This allocation is known to all players.

We use the colors to specify the requirements a coalition must meet in order to deliver a large package. The cost of an LP is expressed in terms of the colors of the chosen path from the package to the goal. More specifically, the LP can only be delivered if a path exists from the package to the goal, such that the colors of that path are represented by the members of the coalition. Look for example at the simple configuration in Figure 5.1. The SP on the right can be delivered by any player on its own. To deliver the LP on the left however, players have to form a coalition. Imagine a green player who intends to form a coalition to deliver the LP. There are three shortest paths to the goal: [blue, red], [blue, green] and [red, green]. Notice that the the color of the square the package is on and the neutral white color of the goal square do not have to be considered. In order to successfully deliver the package, the colors of the taken path must be represented by the colors of the players in the coalition. Given that the player has the color green, he can now form a coalition with either a red or a blue player to deliver the package.

Each color in the path has to be represented by only one player. If, for instance, a path contains three red squares, the coalition needs only one red player. *A coalition can never contain multiple*

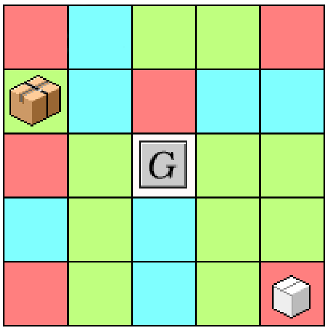


Figure 5.1: Package delivery in the MIPDD game

players of the same color. This forces players who intend to form a coalition to always choose between a human and an agent. By imposing the choice between a human and an agent partner, players are stimulated to reveal their preferences concerning the nature of other players.

Importantly, the players are able to invite more players than necessary to get to the goal. So a package than needs to be delivered by 2 can also be delivered by 3. If certain actors prefer to work together a lot, they can decide to invite each other even though their membership is not a necessity for the coalition to be successful. Note however that a coalition can never contain more than 3 members since a coalition can never contain multiple players of the same color and there are 3 colors in the game.

The payoff of the different packages is as follows: a SP yields a payoff of 3 and an LP yields a payoff of 60 for a 2-player coalition and a payoff of 180 for a 3-player coalition. The packages are randomly distributed over the board, where the amount of large packages is 12 ($2 \times n$) and the number of small packages is 6 (equals n).[1] This number stays the same during the whole game, so when players deliver a package to the goal, a new package of the same type appears. Table 5.1 shows the differences between the types of packages.

Table 5.1: The different types of packages

	Small	**Large**
Color	white	brown
Points	3	60 or 180
Delivery	1 player	2 or 3 players (coalitions)

A player who wants to deliver an LP should invite players to his coalition. He can do so by standing on a large package and making a proposal to another actor. Remember that the location of other players is unknown until they join the coalition. The initiator can therefore base his decision on who to invite on color, nature and previous interactions. To make a proposal, the initiator needs to specifiy:

1. how many members the total coalition will consist of
2. which player he wants to make the proposal to
3. the payoff he is willing to transfer to this player in exchange for joining the coalition and delivering the LP

Figure 5.2 shows how players can propose. In this figure, an initiator is ready to make an offer to a green agent (displayed as a computer icon) to join a 2-player team. Under 'Your Chips' the

[1] In Subsection 5.2.4 we will elaborate on how we determined these values.

FIGURE 5.2: Making a proposal

initiator can specify how much payoff he wants to offer the member. Notice that a reminder of the maximum payoff for this coalition (60) is also displayed.

Responders will receive a proposal that shows the size of the coalition that they are invited to (2 or 3). This provides the necessary information to reason about the equity of the split. In case the proposed coalition has size 3, the identity of the third player remains unknown until he has accepted. Note that the initiator can invite only one player at the time. Importantly, the UG takes place only between the proposer and the member he proposes to. As explained in Section 4.3.3, this member will be aware of the total number of people in the coalition, but not the share of the payoff transferred to another member. He can base his decision to accept or reject only on the payoff the initiator offers.

Once the responder has accepted, he becomes part of the same coalition as the initiator. The position of all coalition members subsequently become visible on the board. Meanwhile, the initiator of the coalition can add other members to the coalition by proposing to them. After a player accepts a proposal it is important that he moves toward the package the initiator is positioned on. The coalition will only be complete when all members have arrived at the location of the package. If it turns out that the members of the coalition do not have the required colors to travel the distance to the goal, a message is displayed in the Coalition Window of all members. The coalition will only be dissolved when one of the members or the initiator defects.

If on the other hand all members are at the package and they have the required colors, the game controller will automatically perform a sequence of actions that leads to the successful delivery of the package. First, the payoff will be transferred as agreed upon in the proposals. Transferring the payoff at this point in the game ensures that players cannot transfer payoff without forming a coalition. Second, all the members will jointly deliver the package to the goal. All coalition members are transformed into one single icon: a van. The van then automatically finds the shortest path containing the colors of the coalition members and moves to the goal. Finally, after they successfully delivered the package, the coalition dissolves and the players are placed at a random location on the board.

Table 5.2 gives an overview of the setup of our domain. It compares Sen's original PDD with our adaptation. The setup of MIPDD is such that it can easily be implemented in Colored Trails, which allows us to create human-agent interactions in a game setting. The representation of a grid in which players can move in all directions to pick up packages is also straight forward since CT provides a board with exactly these properties. CT includes the concept of different types of goals to represent both the packages and the depot. Delivering packages to the depot is be the main goal. Picking up a package will formally be the completion of a subgoal as part of the plan to reach the main goal. The problem of allocation payoff can be represented very nicely with the negotiation protocol CT uses to exchange resources.

TABLE 5.2: A comparison of Sen's Package Delivery Problem with our MIPDD

	PDD (Sen '96)	**MIPDD**
actors	agents	humans & agents
package pickup	at depot, no difference in package size	scattered large (LP) and small (SP) packages
package delivery	depots along the radial fins	one central depot
assignment	at central depot	no assignment
movement	radial grid	grid-based
delivery costs	travelled distance	SP: none LP: colors of path
helping costs	additional travelled distance	additional travelled distance
delivery payoffs	none	SP: 3 LP: 60/180
helping payoffs	none	distributed payoff
coalition formation	ask for help at the depot	ask anyone for help to deliver LP

We can now specify our game as a coalitional game with transferable utility (N,v), where:

set of players N	$N = 6$
characteristic function v	$v(n_i) = 3$
	$v(n_i, n_j) = 60$
	$v(n_i, n_j, n_k) = 180$

Players who act on their own have a non-zero value of 3. Note that our game meets the requirements of superadditivity, since all coalitions S that can be formed have a value that is higher than the added values of the smaller coalitions S can consist of ($180 > 60 + 3$, $180 > 3 + 3 + 3$ and $60 > 3 + 3$).

The difficulty of our approach lies in the more fundamental part of the domain: its highly dynamic and fast paced properties. CT normally uses discrete states, the phases, in which the actions of the player are specified and restricted. In MIPDD this does not suffice if we want (i) packages to dynamically appear, (ii) players to be able to pick up and drop packages and (iii) players to interact with each other all the time. At each point in time, the players should simultaneously be able to perform these actions. This requires the game controller to continuously check for updates of the board and respond to them. We have implemented such a controller.

5.2 The Game

5.2.1 Game configuration

Figure 5.3 shows the components of the board as they are viewed by the player:

- **Board Window** (*center*): This is the board where upon packages and players are positioned. Players can see all packages but only their own position. The position of other coalition members becomes known after a player accepts a proposal. In the configuration displayed in the figure, the 'me' player is standing on the red square, just left to the goal. If this player wants to pick up the LP on the bottom right of the window, he would at least need a green player because there is no path from that package to the goal without moving

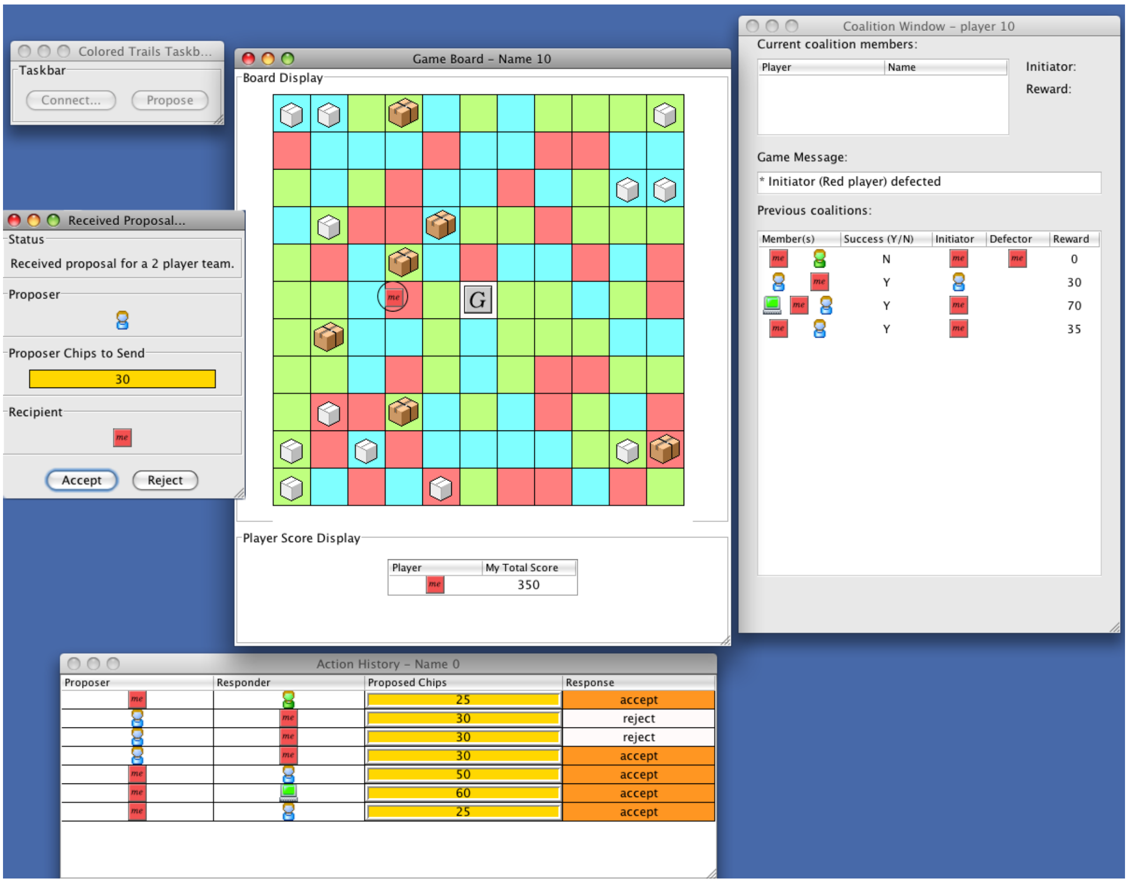


FIGURE 5.3: The game components

across a green square (remember that players are not allowed to move diagonally). Players can move freely around the board in two ways: they can either drag-and-drop their icon using their mouse, or they can use the keyboard cursor keys. When other players pick up packages, these packages disappear from the board. The score and color of the player are displayed at the bottom of the Board Window.

- **Task Bar** (*upper left*): Contains the 'propose' button, which becomes active after a player moves on top of an LP. By pressing this button, a Proposal Window will appear. This window requires the player to select the player he wants to propose to and to specify how many chips and how many players the total coalition will involve.

- **Coalition Window** (*right*): Keeps track of all past and present coalitions of the player. The white panel at the top shows the coalition the player is currently part of. In this case the panel is empty, meaning that the player is currently not part of a coalition. Short game messages (such as defection notifications and the joining of new coalition members) appear in the small white panel directly beneath the first panel. The large white panel contains all previous coalitions the player was enrolled in, showing whether the coalitions were successful or not and how many points the player obtained by being in that coalition. It also shows who was the initiator and in case of defection who was the defector. The Coalition Window is very important for players who want to keep track of the players with whom they had successful interactions. In this example it shows that the latest coalition this player was in, was one involving the green human. The player initiated the coalition himself, but also defected from it. The coalition was therefore not successful and did not yield any payoff.

- **Action History Window** (*bottom*): Keeps track of all proposals and the responses that were sent from and to the player. The most recent proposal is displayed at the top of the window. Here, the most recent proposal for this player was one of 25 points he offered to the green player. The proposal was accepted. This window serves is a tool for players to remember the history of proposals.
- **Proposals** (*left*): Proposals will appear to the left of the board. If the player accepts a proposal, he will be part of a coalition. The proposal consists of the size of the proposed coalition, the icon of the sender, the proposed split of payoff and the icon of the recipient.

5.2.2 Defecting

We argued in Section 4.3.5 for the use of a defection mechanism to study trust relations between the players. There can be several reasons why a player would choose to defect. We distinguish two main categories:

1. **Untrustworthy team member(s)**
   A player might be dissatisfied with the other member(s) of his coalition. Once a member joins a coalition, his whereabouts become known to all coalition members. This information can be used to see if the member is trustworthy. For example, players can now see if the member is collecting packages on his own or that it takes him very long to join the other members at the package.
2. **Better outside offer**
   When the player is presented with a better offer than the one he has currently accepted, he may be inclined to break his team commitment and accept this offer.

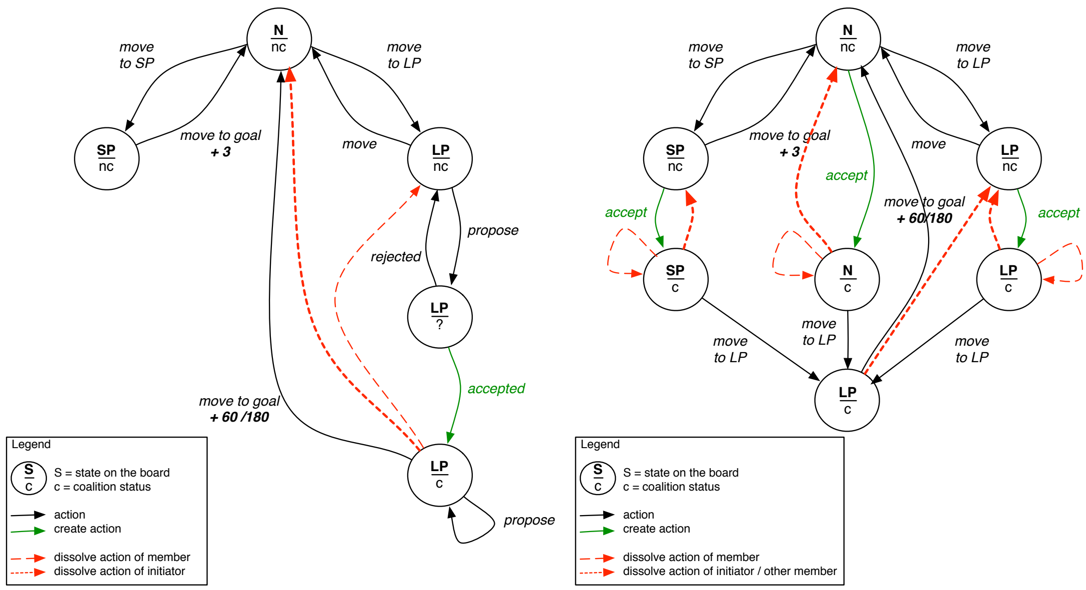
Of course there can be other reasons for the player to defect that do not fall under a clear and distinct category. For example, the player can make a mistake and accidentally defect. Or he might change his mind and want to deliver a different package. Although we ask for a motivation from the subjects on why they defected, we assume for now that these cases are a minority.
In the MIPDD, a player can defect in the following ways:

1. As an initiator (creator) of a coalition:
   - By accepting a proposal from someone outside his current coalition.
   - By stepping away from the package he intended to deliver.
2. As a member of a coalition:
   - By accepting a proposal from someone outside his current coalition.
   - By proposing to form a new coalition.

If a player defects, this information will appear in the Coalition Windows for the players involved in the coalition. Furthermore, a defection always causes the coalition to be dissolved by the game controller.

5.2.3 Game Flow

Now that we have explained the game's components, we will describe the sequence of the game in more detail. We created two state diagrams that show the game flow for an initiator and for a member of a coalition. Respectively: keep in mind that at any point of the game a player can change from being an initiator to being a member and vice versa. This can be achieved by accepting another proposal or by taking initiative and creating a coalition.
Let's first look at the state diagram for initiators as represented in Figure 5.4(a). The general sequence of the game is as follows:



(a) State diagram for initiators (b) State diagram for members

FIGURE 5.4: State diagram for the MIPDD

- The player starts at a random position on the board. The game controller will position the player such that he on a *neutral* (**N**) spot on the board (e.g., not on any package) and not part of any coalition (nc). This is denoted as $\frac{\mathbf{N}}{nc}$ in the diagram.

- The player now chooses to either move towards a small package or a large package. If he moves to a small package, the player automatically picks up this package, resulting in state $\frac{\mathbf{SP}}{nc}$. Note that because the player automatically picks up the package, he can not move away from the square to return to $\frac{\mathbf{N}}{nc}$. Instead, he can return to a neutral state by delivering the package to the goal, by which he gains 3 points.

- If he moves to a large package ($\frac{\mathbf{LP}}{nc}$), the player has two options: (1) return to a neutral state by simple stepping away from the package or (2) creating a proposal to invite another player to his coalition (resulting in $\frac{\mathbf{LP}}{?}$ since the initiator does not know yet whether his proposal will be accepted). He is then required to select the total number of players of this coalition, the color and nature of the player he wants to invite and the payoff he is willing to transfer.

- The initiator now waits for the member to respond to his proposal. If it is accepted, the players form a coalition ($\frac{\mathbf{LP}}{c}$). If it is rejected, the player goes back to the previous state where he can choose again to propose or to move away.

- Depending on how large he intended his coalition to be, the initiator can now invite other members to his coalition. Note that the player stays in the same coalition state, regardless whether his proposal is accepted or rejected. However, both the initiator and the members can at this point dissolve the coalition by defecting according to the game rules explained in Subsection **??**. If a member defects, the initiator is still on the LP and so in coalition state $\frac{\mathbf{LP}}{nc}$. If an initiator defects, he ends up again in coalition state $\frac{\mathbf{N}}{nc}$. He will also reach this state if the coalition successfully delivers the package. Delivering the package also yields the payoff corresponding to the coalition size.

The coalition state diagram for a player who becomes a coalition member is represented in Figure 5.4(b). Although the first two actions that the player can perform are the same as the ones of the initiator, the game sequence as a whole is slightly different:

- The player starts at a random position on the board, not on any package and not as part of any coalition ($\frac{\mathbf{N}}{nc}$).
- The player now chooses to either move towards a small package or a large package. Again, if the player moves to a small package, he will automatically pick up this package ($\frac{\mathbf{SP}}{nc}$). He can return to a neutral state by delivering the package to the goal. If the player moves to a large package, his state on the board changes. If the player now chooses to make a proposal, he becomes an initiator and his game flow changes to the one in Figure 5.4(a). By moving away from the LP, the player ends up again in $\frac{\mathbf{N}}{nc}$. There is however also a third situation that can occur: the player can receive a proposal and if he accepts it, he will become part of a coalition.
- When accepting a proposal directly from the $\frac{\mathbf{N}}{nc}$ state, the player is still at the same position on the board, but is now part of a coalition (resulting in $\frac{\mathbf{N}}{c}$). The same happens if the player accepts a proposal in a $\frac{\mathbf{SP}}{nc}$ or a $\frac{\mathbf{LP}}{nc}$ state (resulting in $\frac{\mathbf{SP}}{c}$ and $\frac{\mathbf{LP}}{c}$, respectively). Note that a player who is carrying a SP can still join a coalition.
- From these coalition states, where the player is in a coalition but not yet on the position of the LP that needs to be delivered, three actions are possible. The simplest action is if the member moves toward the LP in question. The other two actions represent defection actions that are performed by either the member or the initiator. If the member defects, then this results in a similar coalition state where the member is still in a coalition and at the same position. However, if the initiator or another member of the coalition defects, the coalition will be dissolved and the player will return to a state with a coalition status 'nc'.
- Ideally, if the player has arrived at the LP, the coalition is complete and they deliver the package. Or, in case the player joined a 3-player coalition, he might have to wait for the other member to arrive at the package. The other scenario is that the initiator or the member defects, leaving the player at the package without a coalition ($\frac{\mathbf{LP}}{nc}$).

5.2.4 Considerations and Pilot Experiments

In this section we give a motivation for the values of the variables in our domain, such as the number of SP's and LP's, the payoff of the different packages and the size of the board. Furthermore, we will elaborate on some of the implementation details.

We let the game be played by 6 players, which was a logical choice for us since we wanted 2- and 3-player coalitions to be formed. 3-player coalitions must consist of 3 players with different colors. Because each color is represented by 1 agent and 1 human, the minimal number of players is 6.

Given that the game is played by 6 players, we ran several pilot experiments to determine the board size, the number of packages and the different payoffs. Eventually we found a relation between the number of players k and the size of the board that ensured a good game play. Given a number of players k, the size of the board is $N \times M$, where both N and M are defined as: $2 \times k$ - 1. The - *1* is added in order to create borders of an uneven number of squares which allows us to place the main goal exactly in the middle of the board. Accordingly, our game has a 11×11 board.

To establish the number of packages we were concerned with the fact that players might try to deliver the same package. Since the position of other players is not visible, players expressed frustration to see a lot of packages disappear in front of them. Because it takes more effort to collect large packages, we placed more SP's than LP's on the board. Another reason for this is that we did not want to force players to cooperate, each player should have enough opportunity to act individually. Eventually, we placed as many LP's on the board as there were players and twice as many SP's as LP's .

In determining the value of the payoffs, we took into account the following constraints:

1. In order to enable players to make fair splits, the 2-player coalition payoff should be divisible by 2 and the 3-player coalition payoff should be divisible by 3.
2. The payoff for a SP should be much smaller than for an LP because it takes less effort to deliver a SP. However, the payoff should not be as small that players neglect SP's.
3. The payoff ratio should stimulate both cooperation and competition.

We tried different values for the payoff of the SP's and LP's. The pilots showed that a payoff ratio of 3 for SP's and $(s \times 30) \times (s - 1)$ for LP's with a coalition size of s meets the above mentioned requirements.

In the implementation of our game — under the hood, so to speak — we complied with the default way CT handles movement. The resources of all agents consist of chips of one color. A red player, for instance, has a set of red chips. Technically, for a coalition to move to the goal each player in that coalition has to hand in at least the number of his chips that correspond to the number of squares of the same color in the path to the goal. Because we are interested only in the colors of a path and not in the number of squares of one color, we made sure that the players did not have to worry about the number of chips to provide in order for the coalition to move. The chip transfers of the agents were done automatically when a coalition was complete and ready to move. Players were not aware of the number of their resources, nor could they accidentally run out of chips because we ensured that all players had an abundantly large number of chips to accept all proposals.

5.3 The Experiment

We ran two experiments with a total of 18 subjects from different levels of society. 44% of the subjects was male, 56% female. 50% of the subjects was younger than 25, 44% was of an age between 25-29 and 6% was between 30-34. The majority (72%) of the subjects were students, all enrolled in different majors.

The game consisted of several rounds in which subjects played with one configuration of the board. Each round, the game controller randomly divided the subjects into groups of 6. The duration d in minutes of each round is randomly assigned, where $5 \leq d \geq 14$. Earlier work has shown that if subjects know when a game ends, the last interactions suffer the same consequences as one-shot games: non-cooperation is the rational action. Using a fixed value for the length of the round would increase predictability of the game, with similar consequences for the subject's decision process. Our participants are not told the duration of the rounds to ensure that their behavior is not influenced by time. Each player participated in 5 rounds.

We introduced subjects to the game by giving a tutorial. The tutorial consists of four parts: a handout, a video, a small questionnaire and a trial round. First, players are given a handout to get familiar with the game and the rules (see Appendix A). This handout is discussed publicly with one of the experimenters. Secondly, we show a video which helps to give subjects an idea of the dynamics of the game. We use the video as a basis to present a walk-through of the game in which we explain all actions that can be performed and the situations that can occur. Afterwards subjects are allowed to ask questions. Then, the subjects have to fill out some questions to demonstrate that they understand the purpose and rules of the game. We

Figure 5.5: The Decision Science Lab. Left: controller screen. Right: subject's cubicle.

deliberately overbooked the experiment with respect to the number of participants, so we could dismiss some of them who did not understood the game properly or showed a disinterest. Finally, we let the subjects play a trial round, so that they could get a feel for how to move, how to propose and how to pick up packages. During all parts of the tutorial we were careful to use 'neutral' terminology. That is, in order to minimize our influence on the behavior of players in the game we did not use terminology that suggested a competitive or cooperative domain. For example, we used 'participant' instead of 'opponent' and 'interaction' instead of 'game'.

In order to give players an incentive to maximize their own payoff, we paid in a positive linear correspondence to their performance in the game: they earned more money if they performed better. This ensures that the players aim to maximize their score. How well they did is expressed by the sum of the payoff they obtained over all rounds. We made the subjects aware of this payment function. Players received on average $23 for participating.

The experiments were held at the Harvard Decision Science Laboratory of the Harvard Kennedy School, Cambridge, Massachusetts. This laboratory is especially designed to look in a systematic way at the factors that lead people to make different kinds of decisions, and ultimately to find ways in which people will make better decisions. Also, the lab has an advanced technological system that is very suitable for the display of computer-driven presentation technologies to conduct empirical research. The laboratory features 36 cubicles for subjects and three large controller screens to control all individual computers (see Figure 5.5). Removable partitions offer researchers the opportunity to run 12, 24, or 36 subjects at the time.

5.4 Evaluation

When the focus of a research project lies on human behavior, questionnaires are of essence. A questionnaire can provide insights into how humans perceive themselves and others and what motivations they had for the decisions they made. The questionnaire shows how cooperative aspects come about and how these were influenced by the different factors at play.

We use questionnaires because they are a good tool to gather more detailed information about how decisions come about. However, we are aware that there are several disadvantages of questionnaire techniques: they depend on the subject's motivation, honesty, memory and ability to formulate thoughts and feelings. On top of that, questionnaires are not appropriate to study complex social phenomena, since the individual is not the best unit of analysis in all cases. Because of this, we keep track of all actions that occur during the game: which coalitions were formed, which were successful, who defected, what was proposed, etc. We use the information from these logs to study the cooperation and competition between actors and to look at emergent behavior at the higher social level. On the one hand logs can validate comments of the players from the questionnaire and on the other hand, the questionnaire can motivate players' behavior as observed in the logs.

Our questionnaire has a major focus on the strategic choices players made during the game. This is motivated by our interest to capture the effects of nature and trust in repeated interactions between the players. In our pilot tests we identified (i) the questions that proved difficult to answer or did not address the correct issues and (ii) the questions that provided the answers we were interested in. We used this initial analysis to improve our questionnaire. The final questionnaire and its motivation is discussed in Appendix B.

Chapter 6

Results and Analysis

In this chapter we will use our data analysis to answer our main research question: how do nature and trust influence people's decisions in mixed coalition formations? The results we discuss relate to two experiments. In Experiment 1, 6 subjects played the game while Experiment 2 involved 12 subjects. We collected data over a total of 12 rounds of the game. In this chapter we will combine the data of both experiments since the conditions of both experiments were the same.

6.1 Analysis

In this chapter we address our sub-questions mentioned in Chapter 1:

1. To what extent do trust and fairness influence team formation in mixed networks?
2. How does the nature of actors affects the way people relate to their actions?
3. Do actors form stable relationships over time?

In order to answer these questions we look at the data provided by our logs and questionnaire. Results 1 - 10 as presented in this chapter are statistical significant results from the logs. We will refer to data from the questionnaire in order to explain or illustrate results from the logs. In some cases we will explicitly refer to a subset of all subjects. For example, the subjects of Experiment 2 filled out a more detailed questionnaire than the subjects of Experiment 1.

To make our analysis more clear, we will first introduce the terms we use to analyze our data. Successful coalition formation in our domain can be divided into two 'steps':

Step 1 The initiator invites the member(s) who thereafter join(s) the coalition.

Step 2 The coalition delivers a package.

In our analysis we will make a distinction between the successful completion of stage 1 and 2. A coalition that has not yet successfully completed stage 1 is called an *attempted* coalition.

Definition 18 (Attempted coalition)**.** A coalition is *attempted* if:

- The initiator has at least made one offer to a member.
- Not all members agreed (yet) to be in the coalition.

A coalition is *formed* when step 1 is successfully completed.

Definition 19 (Formed coalition)**.** A coalition is *formed* when the initiator has invited *n-1* members for a n-player coalition and the members have agreed to be part of that coalition.

Consequently, a 2-player coalition is formed when 1 player accepts the offer of the initiator and a 3-player coalition is formed when 2 players accept the offer of the initiator. Note that every coalition is in attempted state before it is formed. We will now define a *successful* coalition using the notion of a 'formed' coalition:

Definition 20 (Successful coalition)**.** A *successful coalition* is a formed coalition that succeeded to deliver a large package to the goal.

Definition 21 (Unsuccessful coalition)**.** An *unsuccessful coalition* (or *failed* coalition) is a coalition that was dissolved before it was able to deliver the package. The coalition is dissolved when a player who is part of the coalition defects.

We will analyze our results in terms of *throughput* and *stability* [12, 72]. Throughput is an important measure of performance as it illustrates the amount of work an actor is able to complete in a given time. In our case, we will define throughput as the number of small and large packages players were able to deliver.

Definition 22 (Throughput)**.** The number of tasks completed or goals achieved in a given time frame.

We will furthermore use stability to see how coalitions developed over time. Stability enables us to look at the preference of players to work together since it analyzes changes in coalition structures.

Definition 23 (Stability)**.** The rate and change in coalition size and composition over time.

Stability can thus help us to identify trust relations between players. In our case, stable coalitions are coalitions that often occur the same composition. If stable coalitions are formed, this implies that it's members prefer working each other over working with others.

6.2 Coalition Formation

There is a wide variety of player strategies in the rounds we collected. The questionnaire and logs show that some subjects were often initiator, while others preferred to deliver small packages while waiting to be invited to a coalition. Some subjects defected easily while others tried not to defect from their coalition. Our analysis of the logs shows that group membership is a significant predictor of performance (score). We found that the number of times participants joined coalitions is significantly positively correlated to performance (pearson correlation $r = 0.56$). Initiating coalitions is also positively correlated with performance (pearson correlation $r = 0.42$).

Result 1. *Both joining and initiating coalitions is positively correlated with performance. Joining is significantly more correlated to performance than initiating.*

The number of individual small packages that were delivered was a significant predictor of performance as well (linear regression $r^2 = 0.765, p < 0.0001$). However, successful coalitions were a significantly better predictor of performance than individual package delivery ($t = 3.57, p < 0 = 0.001$). This implies that in our game coalition formation is preferable over the individual delivery of packages if a player wants to maximize his score.

Result 2. *Being part of a successful coalitions is a better predictor of performance than delivering small packages.*

In other words, delivering large packages results in an overall improvement in performance. This result shows that people understood that the purpose of the game was to cooperate and that the game has the desired property of making all players better of if all cooperate.

Table 6.1: Ratio of the number of attempted coalitions and their eventual successes during all rounds

	attempted	**formed**		**successful**		
		total	% of attempted	total	% of attempted	% of formed
2-p coalitions	202	122	60%	83	41%	68%
3-p coalitions	295	205	69%	114	39%	56%

To gain insight in the course of the game, we analyzed the formation of coalitions and the different structures of the coalitions. First recall that the payoff for 3-player coalitions (180) is substantially higher than 1.5 times the payoff of 2-player coalitions (60). This adds an incentive for players to form larger coalitions, despite the additional risk involved.

Over all rounds, players attempted to create 497 coalitions, of which 202 (41%) 2-player coalitions and 295 (59%) 3-player coalitions. Of the total attempted coalitions, 197 (40%) successfully delivered their package. Players had an average throughput of 7 packages with 2-player coalitions and 10 packages with 3-player coalitions per round. Overall, subjects delivered 2418 small packages, which results in an average throughput of 34 small packages per subject per round. As can be derived from Table 6.1, there were significantly more attempts to create 3-player coalitions than 2-player coalitions (goodness of fit $p < 0.0001$). There were 295 attempted 3-player coalitions vs. 202 2-player coalitions. The result that players attempted to create more 3-player coalitions follows directly from the fact that 3-player coalitions were potentially more lucrative than 2-player coalitions. The questionnaire shows that the difference in payoff was a strong motivation for players to create 3-player coalitions.

Table 6.1 shows the ratio of attempted 2- and 3-player coalitions and their formation and success rate. As shown by the table, 60% of all attempted 2-player coalitions and 69% of all attempted 3-player coalitions result in formed coalitions. The percentage of coalitions that was able to successfully deliver their package relative to the total number of attempted coalitions are very similar for 2-player and 3-player coalitions (41% and 39%, respectively).

As indicated in the previous chapters, being part of a 3-player coalition increases waiting time and the risk of defection of other players. We would thus expect 3-player coalitions to be less successful. However, the results from Table 6.1 seem to suggest that attempted 3-player coalitions turned out just as successful as attempted 2-player coalitions seems to contradict this expectation. However, if we look at how many of the *formed* coalitions resulted in successful coalitions, this shows us something else. The fact that attempted 2- and 3- player coalitions are as successful can be explained by the fact that (i) attempted 2-player coalitions result less often in formed coalitions than attempted 3-player coalitions and (ii) formed 3-player coalitions are subject to more defection than formed 2-player coalitions. While 69% of attempted 3-player coalitions is formed, only 60% of attempted 2-player coalitions is formed. 3-player coalitions form more often than 2-player coalitions (goodness of fit $p < 0.05$). This implies that offers for 2-player coalitions were more often rejected than for 3-player coalitions.

Result 3. *Attempted 3-player coalitions form more often than attempted 2-player coalitions.*

Furthermore, Table 6.1 shows that while 122 of the 2-player coalitions in attempted state are formed, only 83 make it to a successful coalition. This is 68% of the formed 2-player coalitions. Only 56% of the formed 3-player coalitions are successful. Recall that the only reason why a formed coalition would not be successful is defection from one of the coalition members or the initiator. These results imply that the defection rate in 3-player coalitions is much higher than in 2-player coalitions ($\chi^2(N(1, 304)) = 11.8, p = 0.001$). Players in 2-player coalitions are more likely to be loyal. Figure 6.1 shows the defections in 2- and 3-player coalitions in a histogram.

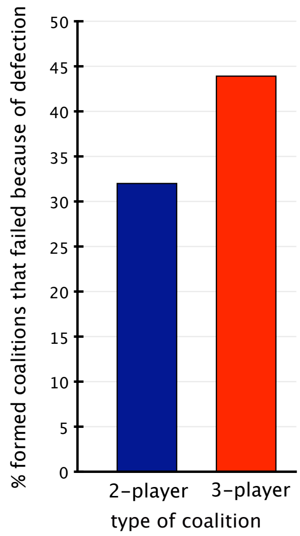


Figure 6.1: Defections from formed 2- and 3-player coalitions

These results are consistent with our expectation that 3-player coalitions entail a higher defection risk.

Result 4. *Players defect more from 3-player coalitions than from 2-player coalitions.*

The logs show that defections were not a significant predictor of performance, showing that defecting in itself did not necessarily result in a lower score. This suggests that although these players were less trustworthy, this had no overall negative influence on their performance.

Result 5. *Players with a high defection rate were as successful as those with a low defection rate.*

Players showed very different defection behavior. Table 6.2 demonstrates the defections of a player and the average number of successful coalitions he was part of. For example, a player who defected 2 times was on average part of 8 successful coalitions. From this table we can conclude that there was a lot of variance in defection behavior. Although players rarely defected more than 4 times per round, there is no clear evidence that supports the hypothesis that defections decrease the chance of being member of a successful coalition. There can be different explanations for this result:

1. Defectors optimized their score by choosing coalitions with the best offers and high potential to succeed.

2. The game was not designed as a reputation game.

First, players defect because they have more 'interesting' cooperation opportunities. In the questionnaire we asked subjects what their reason was for defecting as an initiator and as a member of a coalition. In both cases the majority stated that better offers were their most important motivation (this was the case for 75% of the initiators and 67% of the members). By selecting the best offers and defecting if necessary, subjects are thus able to seek out the most profitable coalitions. This would explain why their score does not deviate significantly from those who do not defect often.

The second explanation is motivated by the fact that players were not able to keep track of other players' defections because they only received information about players they worked with. Because the game was not a reputation game, players could only use experience-based trust. For

TABLE 6.2: Player defections in relation to their membership of successful coalitions per round.

number of defections	likeliness to occur	avg. successful coalitions
0	19%	7
1	21%	7
2	21%	8
3	15%	9
4	17%	5
5	3%	6
6	3%	13
7	1%	6

instance, initiator of coalition C could not always know that the player he invited defected on coalitions A and B. To confirm either of the above hypotheses, we have to do a more in depth analysis of the results. We intend to look at this issue more closely in future work.

6.3 Social Factors

In order to answer our research question, we analyzed the payoff distributions between subjects. Keep in mind that every player had a different perception of the nature of other players. We first calculated the optimal fair split and computed how the proposed split deviated from that optimal split. For example, in the case of a 2-player coalition the joint payoff is 60. A fair offer would then be an offer where the initiator proposes 30 to the member. This would be a 100% fair proposal. A proposed split of 20 for the member would be a 66.7% fair split. For 3-player coalitions it would be 100% fair if the initiator offered $\frac{1}{3}$ of 180 to a member, resulting in a payoff of 60 for the member.

On average, initiators of coalitions proposed splits that are 83% fair to recipients. Splits offered to people averaged 94% fair, significantly higher than the splits offered to computer agents, which averaged 82% (combined t-test $t(692), p < 0.0001$).

Result 6. *Players offer humans significantly more fair splits than they offer agents.*

The fairness results of offers to people correspond to results found in several Ultimatum Game studies between people: (i) there are virtually no offers that are more than 100% fair, (ii) the vast majority of offers in almost any study is in the fairness interval [80%-100%], (iii) there are almost no offers with a fairness lower than 20%, (iv) low offers are frequently rejected [15, 39, 71].

Although rare in occurrence, some initiators were willing to allocate slightly more to recipients than to themselves. We call this an *altruistic split* if it results in a split yielding more payoff for the responder than he would have received with a fair split. For example, in a 2-player coalition, the initiator would make an altruistic split if he offered the member 35, leaving himself with 25. We found no difference in the number of times initiators made altruistic splits to humans (20 occurrences) and to agents (19 occurrences).

The average split for offers that were accepted was 90% fair. The average split of offers that were rejected was 83% fair. In the questionnaire subjects describe their strategy for proposing a split. 56% mentions that they often or always propose an even split. 22% talks about creating proposals that are 'fair', 'decent' or 'reasonable'. 22% also mentions that they give less chips to agents than to humans, which corresponds to what we found in the logs. Noteworthy is that when we asked subjects to describe their strategy for choosing their team members, 22% says that they pick members who are fast responders in the game. So when subjects consider previous encounters with others, they do not only focus on successes and defections, but apparently also on reaction time.

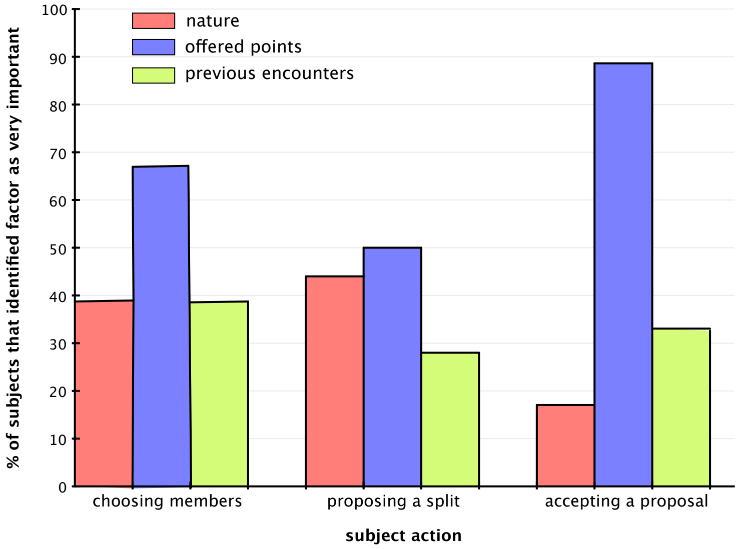


FIGURE 6.2: The factors with the highest importance for subjects' decisions according to subject questionnaire.

Although the logs show that initiators prefer inviting agents over humans (369 proposals to 328 proposals, respectively), this was not a significant result. There was also no significant difference between the amount of coalitions that were joined with a human and with a perceived agent as the initiator. More specifically, there is no significant difference between the percentage of proposals that was accepted from humans and from agents. Taking the results for accepting proposals and joining coalitions together we conclude:

Result 7. *The nature of participants does not affect the choice of coalition partners or the acceptance of offers.*

Figure 6.2 shows the percentage of subjects that stated in the questionnaire to find nature, offered points and previous encounters a very important factor for making particular decisions in the game. Even though 'offered points' is considered an important factor by most subjects, 'nature' as well is often identified as important. 39% finds the nature of the member important when they have to decide to be part of a team with this member. 44% finds nature important when it comes to how they propose a split. However, when it comes to accepting a split, nature is clearly less important than the offered points and even the history of interaction. The log data shows no significant difference in the average fairness of accepted offers given the nature: offers that were rejected from agents averages 80% fairness while offers that were rejected from humans averages 84%. It seems that what mattered most to people when accepting offers was fairness and not the nature of the proposer.

So far, our log results have shown that nature does not make a difference for joining teams and for the acceptance of proposals. Yet of all participants, 44% declared to sometimes or frequently use a different strategy towards agent and humans. To examine this more closely, we let the subjects in Experiment 2 answer more specific questions about their preference for cooperating with agents or humans. The results can be found in Table 6.3. Note that the results display that players had no clear preference for either humans or agents for accepting proposals, exactly as was displayed in Figure 6.2. Players also don't really care whether an agent or a human is the initiator of a team. Furthermore, Table 6.3 shows that some players clearly prefer humans while other prefer agents to be in their coalition. The distinction is most clear for 'proposing spits', where a remarkable stated to 50 % prefer agents. This can be explained by the fact that players proposed far less fair proposals to agents than to humans (see Result 6). Players made

Table 6.3: Players' preferences to cooperate with others of a certain nature.

	Game components			
Preferred nature	choosing members	accepting proposals	proposing splits	as initiator
human	33 %	17 %	8 %	17 %
agent	25 %	17 %	50 %	17 %
no preference	42 %	67 %	42 %	67 %

statements like "I felt I could 'pay' computers less without guilt." and " I tried to lure computer with a very low chip value." Apparently players prefer agents because they offer them less and so could keep more payoff to themselves.

We found that nature does not significantly influence the success of a coalition. Out of 55 defections by members of coalitions, 33 (60%) defected from what they perceived as human initiated coalitions, and 22 (40%) from agent initiated coalitions. This is not a significant difference. These results are supported by the data from the questionnaire as displayed in Figure 6.3, which shows that humans have no clear preference for either humans or agents as an initiator.

Result 8. *The nature of the initiator did not affect coalition success.*

A discriminating factor for fairness that we did not foresee is team size. Offers that were made for members of a 2-player coalition were significantly more fair (99%) than for members of a 3-player coalition (83%) (t-test $t(690) = 8.90, p < 0.001$).

Result 9. *Offers in 2-player coalitions were more fair than offers in 3-player coalitions.*

A possible explanation for this is the significant difference in payoff player can receive for 2- or 3-player coalitions. A completely fair split might be the baseline, but initiators of 3-player coalitions can offer less than a fair split and still offer a significantly higher payoff to the member than this member would obtain in a fair split from a 2-player coalition. If the initiator of a 2-player team offers 30 points to a member, this is a 100% fair split. This same offer would only be a 50% fair split in a 3-player team. Initiators of 3-player teams can make members very appealing offers without being fair.

Our analysis shows that some participants cooperated multiple times with those whom they successfully interacted before. This suggests that these actors were able to build lasting relationships on the basis of trust. To find stable relationships between players, we analyzed how often the same members were in a successful coalition together. This analysis looks purely at the members of a coalition, not at who initiated it. For example, if in one round the coalition of players with id 0, 2 and 5 was successful and some minutes later in this round the same actors are in another successful coalition but with a different initiator, these coalitions are treated as two instances of one coalition.

The same 2-player coalitions were successful on average 1.6 times each round, while the same 3-player coalitions were successful on average 2.0 times per round. What is more, 42% of the successful 2-player coalitions and 56% of the successful 3-player coalitions consisted of coalitions that occurred before. So once players interacted in a successful 3-player coalition, they prefer to cooperate with the same players again in future interactions. We state the following result:

Result 10. *3-player coalitions are more stable than 2-player coalitions, despite their higher defection rate.*

6.4 Concluding Remarks

Our hypothesis that the nature of other actors a?ects the cooperation between them is partially confirmed: the nature of other actors affects the fairness of offers made between them. However, humans seem to only discriminate agents when making proposals and have no preference for either humans or agents regarding other cooperative actions.

We also formulated the hypothesis that a trust relation between the actors affects the cooperation between them. First we looked at positive trust relations between actors. The stability of 3-player coalitions confirms that a positive trust relation, strengthened by previous interactions, positively influences future cooperative actions. However, we have not established that a negative trust relation negatively affects cooperation between actors. Defectors perform as well as non-defectors and there is no clear relation between the defection rate of a player and the successful coalitions he is part of.

Chapter 7

Future Work

7.1 MIPDD

During our analysis we found several results that cannot be explained by our questionnaire or data from the logs. We did offer some hypotheses that need further analysis and testing to confirm them. Here we list the remaining issues and approaches to investigate them:

- Higher fairness in 2-player coalitions than in 3-player coalitions.
  If subjects in 2-player coalitions repeatedly interacted with each other in other 2- and 3-player coalitions, their familiarity could possibly explain the higher level of fairness in 2-player coalitions. We need to look more closely at who cooperated with whom to confirm this. We also mentioned the possibility of initiators of 3-player coalitions being less fair because they are able to make more lucrative offers than initiators of 2-player coalitions. In order to test this we intend to look at coalition formation games where only 3- player coalitions are allowed.

- Negative trust relations negatively affects cooperation between actors.
  In order to establish whether defectors were punished, further analysis is needed of coalition formation over time during the rounds. It would be interesting examine whether defectors were in future interactions less invited by members of the coalition they defected from. Another hypothesis is that defectors were still invited but actors offered them a less fair payoff allocation.

- Joining coalitions is significantly more correlated to performance than initiating coalitions.
  Initiating coalitions does not seem to pay off as well as joining coalitions. It might be that initiating coalitions often results in disappointment because of rejections and defections. Initiating also takes more time than accepting membership. It appears that initiators do not keep a significant larger part of the payoff to compensate for these efforts. Either initiators just want to play fair or the members do not want to reward leadership. Future work on teamwork might shed some light on this.

7.2 BDI Agents

We started out by saying that our objective was to study human-agent interaction in coalition formation scenarios. We first analyzed some of the aspects that influence human behavior towards both agents and humans in these interactions. Now that we have done so, we would like to use our data to create models of the human players and their decision making processes. We can use these models to implement intelligent agents that we expect to do as well as or outperform human players.

To create and implement these models, agents that use a beliefs-desire-intention (BDI) structure are of particular interest. The agents in our environment should be able to perceive and respond to the environment, take initiative in order to satisfy their goals and be capable to interact with other (possibly human) actors. They should be able to reason about a changing environment and dynamically update their goals. Working in a team requires agents to plan, communicate and coordinate with each other. The agents could implement an *aspiration level* [54] that updates the agent's belief about how much payoff it can expect from every other actor, based on our findings on the influence of nature and previous encounters. A BDI architecture lends itself well to implement these requirements in an intuitive yet formal way [33]. The framework CTAPL [50] combines the agent programming language 2APL [19] with Colored Trails in such a way that agents with a BDI decision structure can interact with humans and other agents in a CT environment. In future work, we intend to use this framework to implement agents with social models that can interact successfully in the MIPDD. Additionally, we would like to investigate whether the 'social' agents are more cooperative with humans in this scenario than agents with more algorithmically optimal or rational strategies.

7.3 Pre-Established Payoff Distributions

In this work we looked at how self-interested actors cooperate in a coalition setting. The incentive for cooperating for actors was the significant increase of payoff that successful coalitions generated. Another incentive for cooperation between self-interested actors that we haven't looked at is given by theories of *reciprocity* [26]. Reciprocity theories state that people will act to help (hurt) those who help (hurt) them, or those who are perceived as potentially being helpful (harmful). When A asks B to cooperate as a favor, chances are that B expects A to do something in return. It also works the other way around. Actor A may only be inclined to cooperate after actor B has performed a helpful action during previous encounters. We would like to examine whether the results found in this thesis still hold if the motivation of the actors is not to maximize their payoff in the short run, but when the motivation is reciprocity. Reciprocity could still entail maximization of their own payoff and can therefore be considered as a strategy that appeals to self-interested agents. Imagine the scenario where 2- and 3-player coalitions still obtain a significant larger payoff, but now the payoff is simply rewarded to the initiator and none of it goes to the members. In other words, there is a *pre-established payoff distribution*. On short term, the cost for joining is larger than the gain for joining. In this setup, trust becomes even more important since players have no certainty that other players will behave reciprocally; they could turn out to be freeriders [59].

7.4 Utilitarian Strategy Preferences

People can have utilitarian preferences: the inclination to contribute to society as a whole or to help others that are not in control of the required resources. Utilitarianism determines the utility of an actor solely by its contribution to the happiness or pleasure of all actors. In a small community such as situated in our domain, the motivation for this is most likely interpersonal sympathy. It would be interesting to empirically investigate the tradeoff people make between increasing their personal benefits and contributing to the welfare of all players. This could for example be implemented in our domain by creating packages that can be delivered for the 'greater good': the payoff of these packages will be distributed among all players or will be given to those who need it the most, i.e., those who do not perform as well as the rest. An alternative would be to ensure a fixed partition of the payoff (for example, 30% of the payoff of a LP) is donated to the entire community. An interesting research question would be whether people are willing to contribute more to a society consisting solely of humans than to a society in which both humans and agent interact.

Chapter 8

Conclusion

Coalition formation spans several disciplines from economics and game theory to social science and computer science. Several experimental studies have shown that (i) fairness considerations greatly affect negotiation processes and human coalition formation and (ii) previous experiences strengthen or weaken a trust relation between coalition members. In this study we extended this research by looking at how interactions between self-interested actors can result in cooperative coalition formation. Creating a coalition involves additional risks but at the same time yields a considerably higher payoff. The usual approach in research on coalition formation has been to compare the t of a number of coalition theories and look at the difference between expectancies and actual formations of coalitions and payoff allocations. Both human and agent coalition formation are generally studied from a game theoretical perspectives.

However, the abstract models of game theory is not always most suitable to model real-world processes. Current work on coalition formation that uses game theoretical approaches makes strong assumptions that constrain and control the domain in which coalition formation takes place. In the present study, we did not stick to these assumptions and try to model coalition formation in a more real-world environment. Moreover, we use this environment to study social behaviors between humans and agents. In particular, this study investigates:

1. To what extent do trust and fairness influence team formation in mixed coalition formation?
2. How does the nature of actors affect the way people relate to their actions?
3. Do actors develop stable relationships over time?

Our approach differs from other approaches in that we used mixed-initiative interactions to study purely the differences in behavior between agents and humans that have to deal with a social dilemma. We have brought together work from a number of areas: iterative games, ultimatum games, dynamic coalition formation, social behavior studies and human-agent interaction. We combined findings from these fields to create a domain that discards restrictions commonly used in coalition formation domains. Other than previous approaches, we use a very dynamic and fast paced domain. The self-interested actors are allowed to use defection to their advantage but they also have a commitment to a group. Additionally, we allow the actors to choose between coalition members of different natures. We have designed a coalition formation experiment for mixed actors in which initiators can choose between forming a 2- or 3-person coalition and performing a task individually.

Our first hypothesis expressed our expectation that nature affects people's decisions. The regularities observed in our experiment support the hypothesis that nature sometimes makes a difference in people's behavior. Although our findings show that people do not prefer humans over agents when they choose their partners, both data from the logs and the questionnaire show that they tend to offer less to agents. We found a significant difference between the fairness of offered splits to humans and agents. This results in a subject preference to make splits to agents, since they keep more payoff to themselves when interacting with them. A reasonable amount of subjects claim to take nature into account when choosing team members, though the logs show no difference in joining agent vs human initiated teams or accepting offers. Our results show that other than the fairness of proposals, there is no difference between human cooperative conducts towards agents and humans.

We also hypothesized that a positive trust relation triggers more cooperative behavior between actors and that a negative trust relation results in less cooperation. Our results show that in general people choose to cooperate by building large coalitions, despite the inherent risk. They prefer to cooperate with players they have cooperated before. Surprisingly, a negative trust relation does not necessarily result in less cooperation since players with a high defection rate were as successful as those with a low defection rate.

Modeling and simulating coalition formation behaviors is still a research field with many open issues. One of these issues is how to design an agent to collaborate with multiple human and agent partners simultaneously. One of our future goals is to analytically capture the dynamics of human coalition formation in mixed-initiative environments. We intend to use this analysis to create a model that can be used to design agents that are able to cooperate effectively with both humans and agents. This will bring us another step closer to a world in which people interact with computer systems in a manner similar to the way that people interact with each other.

Appendix A

Handout

A.1 Outline

The following two pages display a booklet version of the handout. The subjects received a color copy of the handout instead of the here displayed gray scale version. The handout was accompanied with some questions to test subjects' comprehension of the game and it's rules. These questions can be found below.

A.2 Questions

Observe the board shown on page 7 of the handout and the coalition window on the right of it. Answer the following questions.

1. Find one of the large packages on the board and draw two possible paths to the depot location.
2. Is the "me" player in a team right now? How many members are in this team? Who initiated the team?
3. Why did one of the teams fail? Who defected? Can you think of a possible reason why the participant decided to defect?
4. Look at the coalition window on the right. How much did the "me" player earn from one of the successful team? Who were the team members?

The Package Delivery Activity

1 General

You are about to engage in an activity whose goal is to collect packages and deliver them to a central depot location. There are two types of packages, small packages can be delivered to the depot without the assistance of others, while large packages require that a team be formed and used to deliver them to the depot. You will be interacting with other participants. You can invite other participants to join a team, or you can accept an invitation from other participants to join their team.

The successful delivery of small packages depends only on your own actions while the successful delivery of large packages depends on your interaction with your team members.Some teams may succeed, while others will dissolve without delivering their pacakge. Your performance in the interaction depends on the number of small and large packages you deliver independently, as well as in a team.

2 The Board

The interaction is conducted on a grid of colored squares. There are both small and large packages located at random locations on the board. The depot location is located as a goal square in the center of the board. You can move freely on the board using the mouse or arrows on the keyboard.

3 The Participants

There are six participants in the task, three humans and three computer agents. Each participant has one of three colors, and an icon on the board that indicates its type. There is a human and a computer agent associated with each color.

4 Movement

Each participant has an icon on the board. You can move freely on the board using the mouse or by the arrows on the keyboard.

5 Packages

Packages are scattered on the board and appear in random locations. There are two types of packages. Small packages (in white) can be delivered by the individual participants without the assistance of others. When you move into a square that has a small package, you automatically pick it up.

Large packages (in red) can be delivered only by a van driven by teams of two or three participants. When you move into a square that has a large package, you can invite one or two other participants to form a team with you, for the purpose of delivering the package. You need to choose team members whose colors match the some path for the van to take from the location of the large package to the depot.

6 Scoring

The performance of each participant is determined by its score in the activity, which depends on the packages it delivers individually or in teams. Your score is displayed at the bottom part of the main panel.

- Small packages are worth 3 points.
- The score for delivering packages is the number of yellow chips allocated by the team initiator.
  - For two player teams, the inititator is given 60 points to distribute among participants.
  - For three player teams, the inititator is given 180 points to distribute among participants.

6.1 Teams

6.1.1 Creating a Team

Each participant can initiate an invitation to other participants to form a team. For this purpose, the initiator is given yellow chips to distribute among all team members, including the initiator itself. Click on the "proposal" button in the task-bar panel, decide on the size of the team, and make an offer to each of the other potential team members. The yellow chips turn into scoring points only if the team is successful in delivering the package.

Think carefully about who you want to ask to join your team. For each color you need, there are two possible participants to chose from, a computer agent or a human. You can ask the same participant to join a team more than once and vary the amount of points that you offer; you can of course choose to join the team of a participant who was a member of your own team the past.

6.1.2 Joining a Team

If another participants invites you to join a team, you will receive an invitation, which will include an offer of yellow chips which you can accept or reject. If you accept the invitation, the icon of the initiator will appear on the board, on one of the large packages. You will also see the number of participants in the coalition window panel. To join the team, you must move your icon to the location of the large package. Once all of the team members have reached the location of the large package, it is delivered to the goal depot by the van. Note that you can still pick up and deliver individual packages while you belong to a team.

6.1.3 Dissolving a Team

A team will terminate in the following cases:

- Successful delivery of its package.
- An initiator defects by moving off the square with the big package.
- A team member defects by accepting an offer from another initiator.

7 Visibility and Constraints

- In general, you cannot see the location of the other participants on the board. However, once you agree to join a team, you will see your other team-members on the board as long as the team is not dissolved.
- Packages disappear once they have been delivered by an single participant or a team.
- You cannot move on the board if you have a pending offer to join a team.

8 Notifications

- The offers and replies are displayed in the message history window located at the bottom of the screen. Messages will appear in the coalition history window as follows:
  1. If any participant defects from your team.
  2. If a package was successfully delivered.
  3. If another team has delivered the packages you happen to be standing on.

9 Optional: More Detailed Instructions

9.1 Initiating a coalition

If a player wants to create a new coalition, he can do so in the following way:

1. Stand on the large package that you want to deliver.
2. The propose button in the Task-bar becomes active.
3. Press the propose button to create a proposal.
4. Select the player you want to invite to your coalition.
5. Select the total amount of participants you want your coalition to contain.
6. Select the value you are willing to transfer to the player in exchange for joining your coalition. (Reminder: the total value to distribute for this package will be 30. The total amount of points for this package is 30 times the number of participants in your coalition.)
7. Press the "Propose"' button to send the proposal.

9.2 Miscellaneous

- When creating coalitions, participants are enforced to choose between the human or software agent as a coalition partner, since they can only invite 1 player of a specific color.
- The position of other participants will only become visible if they are part of the same coalition.
- Participants are not allowed to invite more participants to their coalition than they have specified in their first proposal.
- While a player has a pending proposal that needs to be answered, the player is unable to move.
- Participants are unable to send proposals to multiple participants at the same time. In stead, they have to wait for a response before they can propose again.
- After the delivery of a large package, participants will be placed on a random position at the board.

10 Snapshots


Game Board – Name 60
Board Display
G
Player Score Display
Player
My Total Score
153
Coalition Window – player 60
Current coalition members:
Player
Name
Initiator:
Reward:
Game Message:
* Package successfully delivered
Previous coalitions:
Member(s)
Success (Y/N)
Initiator
Defector
Reward
Y
50
N
0
N
0
N
0
Action History – Name 5
Proposer
Responder
Proposed Chips
Response
25
accept
10
accept
10
reject

Appendix B

Questionnaire

B.1 Outline

Our questionnaire is an online questionnaire that subjects filled out anonymously as part of the experiment, after they played the game. Our questionnaire consists of questions concerning four main categories:

- general questions
- questions about understanding of the game
- questions about players' strategy
- questions about players' preferences

The order of the questions is such that they flow logically from one to the next and from more general to the more specific ones. Some questions are contingency questions and are only posed when subjects give a particular answer. These questions are more specific and in-depth and aim to learn the motives behind certain strategies or preferences.

We are aware that structured surveys, particularly those with close-ended questions, may have low validity when researching affective variables. We therefore use the questionnaire as an addition to the data from the logs, where correlation between variables is more accurate. To increase the validity of the questionnaire, we use both close-ended and open-ended questions. Keep in mind though that this questionnaire was not designed as a full-blown psychological examination, but merely to support findings from our logs.

The open-ended questions we use are completely unstructured and allow subjects to freely write down their opinions or experiences. We use different kinds of close-ended questions: (i) multiple choice questions and (ii) scaled questions. Multiple choice questions give several answers from which subjects can choose. We ensured that we also defined a neutral option, such as 'other', 'maybe' or 'no preference'. Scaled questions need to be answered on a continuum scale. In most cases we use a five-point Likert-sscale, measuring either a positive or negative response to a statement.

B.2 The Questionnaire

Coalition Formation

Purpose of the research: To examine the interactions between humans and computer agents in coalition formation.

What you will do in this research: If you decide to participate, you will complete one survey. Some of the questions will be about the strategy you used to make decisions about forming or joining coalitions. Others will be about the strategies of other players.

Time required: The survey will take approximately 10 minutes to complete.

Confidentiality: Your responses will be kept anonymous. When research results are reported, responses will be aggregated (added together) and described in summary.

Participation and withdrawal: Your participation is completely voluntary, and you may quite at any time without penalty.

Contact: If you have questions or concerns about this research, please contact: Arlette van Wissen Phone: (617) (495-2019) ; Maxwell Dworkin 242, Oxford Street 33, Cambridge, Massachusetts 02138; Email: wissen@eecs.harvard.edu. You may also contact the faculty member supervising this work: Kobi Gal, post-doctoral researcher, Maxwell Dworkin 242, Oxford Street 33, (617) (495-2019), gal@eecs.harvard.edu.

Whom to contact about your rights in this research, for questions, concerns, suggestions, or complaints that are not being addressed by the researcher, or research-related harm: Jane Calhoun, Harvard University Committee on the Use of Human Subjects in Research, 50 Church St., 5th floor, Cambridge, MA 02138. Phone: 617-495-5459. E-mail: jcalhoun@fas.harvard.edu

Please print or save a copy of this page for your records

There are 30 questions in this survey

General

1 What is your sex? *

Please choose **only one** of the following:

- ○ Female
- ○ Male

2 What is the number of the computer you're playing on? (see label on the screen of the computer) *

Please write your answer here:

3 What is your age? *

Please choose **only one** of the following:

- ○ <25
- ○ 25 to 29
- ○ 30 to 34
- ○ 35 to 39
- ○ 40 to 49
- ○ 50 to 59
- ○ 60+

4 What is your current occupation?
(if student, also state your major) *

Please write your answer here:

5 How much experience do you have with multi-player computer games? *

Please choose **only one** of the following:

- ○ none
- ○ little
- ○ some
- ○ much
- ○ very much

The Game

6 How well did you understand the experiment? (e.g., how to collect points, what was the goal of the interaction, how to form teams) *

Please choose **only one** of the following:

- ○ Not at all
- ○ Not very well
- ○ Moderately well
- ○ Fairly well
- ○ Very well
- ○ Completely

7 What did you not understand? *

[Only answer this question if you answered 'Not at all' or 'Not very well' or 'Moderately well' or 'Fairly well' to question 'exp]

Please write your answer here:

Player Strategy

8 What strategy did you use for choosing your team members? *

Please write your answer here:

9 Choosing your team members:
Rank each of the following factors according to their importance for your decision to choose a teammate. *

Please choose the appropriate response for each item:

	1 - not important	2	3	4	5 - very important
color	○	○	○	○	○
nature (human or computer)	○	○	○	○	○
offered points	○	○	○	○	○
previous encounters	○	○	○	○	○

10 What strategy did you use for proposing a split of points? *

Please write your answer here:

11 Splitting points among team members:
Rank each of the following factors according to their importance for your decision on how to split the points. *

Please choose the appropriate response for each item:

	1 - not important	2	3	4	5 - very important
color	○	○	○	○	○
nature (human or computer)	○	○	○	○	○
offered points	○	○	○	○	○
previous encounters	○	○	○	○	○

12 What strategy did you use for accepting proposals? *

Please write your answer here:

13 Joining a team:
Rank each of the following factors according to their importance for your decision to accept a proposal.

*

Please choose the appropriate response for each item:

	1 - not important	2	3	4	5 - very important
color	○	○	○	○	○
nature (human or computer)	○	○	○	○	○

	1 - not important	2	3	4	5 - very important
offered points	○	○	○	○	○
previous encounters	○	○	○	○	○

14 What strategy did you have for defecting as an initiator? *

Please write your answer here:

15 Defecting:
Check the factors that were of importance for your decision to defect as an initiator. *

Please choose the appropriate response for each item:

	1 - not important	2	3	4	5 - very important
outside offer	○	○	○	○	○
behavior of members	○	○	○	○	○
nature of members	○	○	○	○	○
color of members	○	○	○	○	○
appearance of new package	○	○	○	○	○

16 What strategy did you have for defecting as a team member? *

Please write your answer here:

	1 - not important	2	3	4	5 - very important
offered points	○	○	○	○	○
previous encounters	○	○	○	○	○

14 What strategy did you have for defecting as an initiator? *

Please write your answer here:

15 Defecting:
Check the factors that were of importance for your decision to defect as an initiator. *

Please choose the appropriate response for each item:

	1 - not important	2	3	4	5 - very important
outside offer	○	○	○	○	○
behavior of members	○	○	○	○	○
nature of members	○	○	○	○	○
color of members	○	○	○	○	○
appearance of new package	○	○	○	○	○

16 What strategy did you have for defecting as a team member? *

Please write your answer here:

17 Defecting:
Rank the factors according to their importance to your decision of defecting as a team member:
*****

Please choose the appropriate response for each item:

	1 - not important	2	3	4	5 - very important
better outside offer	○	○	○	○	○
behavior of members	○	○	○	○	○
color of members	○	○	○	○	○
nature of members	○	○	○	○	○
appearance of new package	○	○	○	○	○

18 Computer vs. human:
In which scenario's did the nature of the player (human or computer) make a difference?
*****

Please choose the appropriate response for each item:

	preferred humans	preferred computers	no preference
making proposals	○	○	○
accepting proposals	○	○	○
splitting the points	○	○	○
as a team leader	○	○	○

19 How did your preference for humans/computers reflect in your proposing strategy? *

[Only answer this question if you have NOT answered 'no preference' to question 'nature]

Please write your answer here:

17 Defecting:
Rank the factors according to their importance to your decision of defecting as a team member:
*****

Please choose the appropriate response for each item:

	1 - not important	2	3	4	5 - very important
better outside offer	○	○	○	○	○
behavior of members	○	○	○	○	○
color of members	○	○	○	○	○
nature of members	○	○	○	○	○
appearance of new package	○	○	○	○	○

18 Computer vs. human:
In which scenario's did the nature of the player (human or computer) make a difference?
*****

Please choose the appropriate response for each item:

	preferred humans	preferred computers	no preference
making proposals	○	○	○
accepting proposals	○	○	○
splitting the points	○	○	○
as a team leader	○	○	○

19 How did your preference for humans/computers reflect in your proposing strategy? *

[Only answer this question if you have NOT answered 'no preference' to question 'nature]

Please write your answer here:

20 How did your preference for humans/computers reflect in your responding strategy? *

[Only answer this question if you have NOT answered 'no preference' to question 'nature]

Please write your answer here:

21 How did your preference for humans/computers reflect in your strategy for splitting the points?

[Only answer this question if you have NOT answered 'no preference' to question 'nature]

Please write your answer here:

22 How did your preference for humans/computers as a leader reflect in your strategy of joining and defecting from teams? *

[Only answer this question if you have NOT answered 'no preference' to question 'nature]

Please write your answer here:

23 Did your proposing strategy change during a game? *

Please choose **only one** of the following:

- ○ yes
- ○ no
- ○ uncertain

24 How did you change your propose strategy during a game? *

[Only answer this question if you answered 'yes' or 'uncertain' to question 'changepropose]

Please write your answer here:

25 Did your responding strategy change during a game? *

Please choose **only one** of the following:

- ○ yes
- ○ no
- ○ uncertain

26 How did you change your response strategy during a game? *

[Only answer this question if you answered 'yes' or 'uncertain' to question 'changerespond]

Please write your answer here:

Teams

27 What would you say was the ratio between your involvement in 2- and 3-player teams? *

Please choose **only one** of the following:

- mostly 2 player teams
- mostly 3 player teams
- equal
- Other

28 What was the reason for this ratio?

Please write your answer here:

29 What would you say was the ratio between teams in which you were an initiator and in which you were an invited member? *

Please choose **only one** of the following:

- mostly initiator
- mostly invited member
- equal
- Other

30 What was the reason for this ratio? *

Please write your answer here:

Appendix C

Screenshots

The following pages contain screenshots taken during the game. They display how players make proposals, form coalitions and deliver packages.

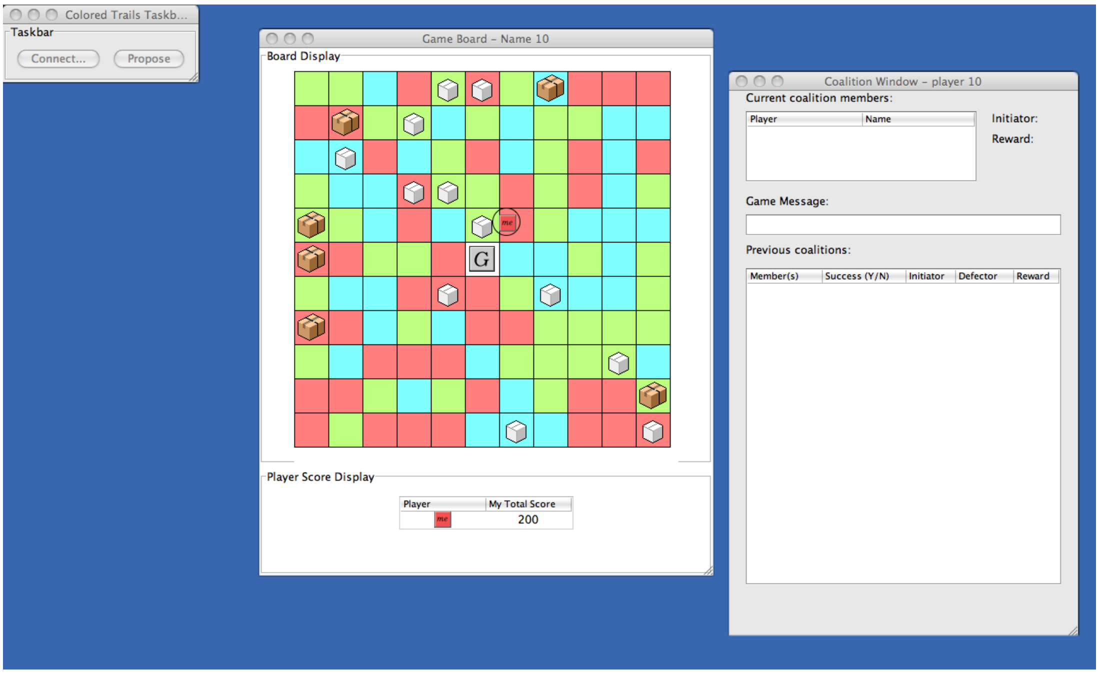


FIGURE C.1: The beginning of the game from the point of view from a red player.

FIGURE C.2: The player picks up a small package.

FIGURE C.3: The player received a proposal from the green human.

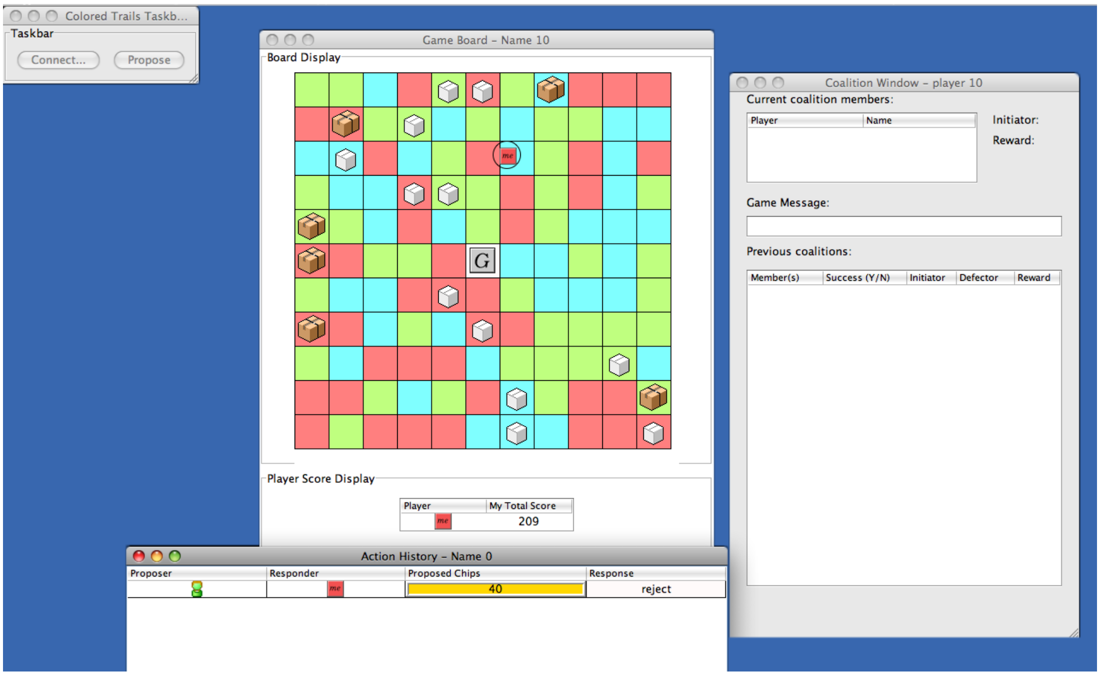


FIGURE C.4: The Action History Window (bottom) shows that the red player rejected the proposal.

FIGURE C.5: Making a proposal: choosing a member, the coalition size and the payoff to transfer.

FIGURE C.6: The player sent his offer of 55 to join a 3-player coalition to the green computer.

FIGURE C.7: As shown by the Action History WIndow, the green computer accepted the proposal. The Coalition Window (right) shows that the red 'me'-player and the green computer are now part of the same coalition. The green computer's locations is now visible (top left on the board).

FIGURE C.8: The red player invites the 2nd member to the 3-player coalition. Notice that the payoff that the initiator can give (125) consists of the total payoff for this coalition (180) minus the the payoff that has already been promised to the first member (55).

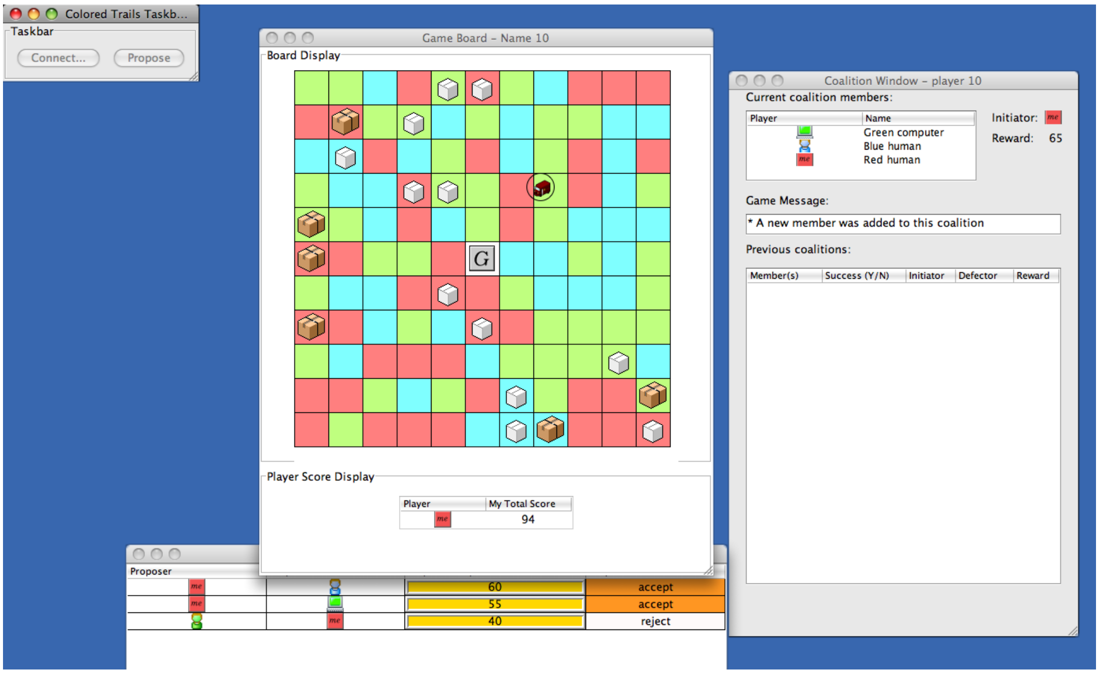


FIGURE C.9: When both members have acceepted and moved to the location of the initiator, the players turn into a van that delivers the package to the goal.

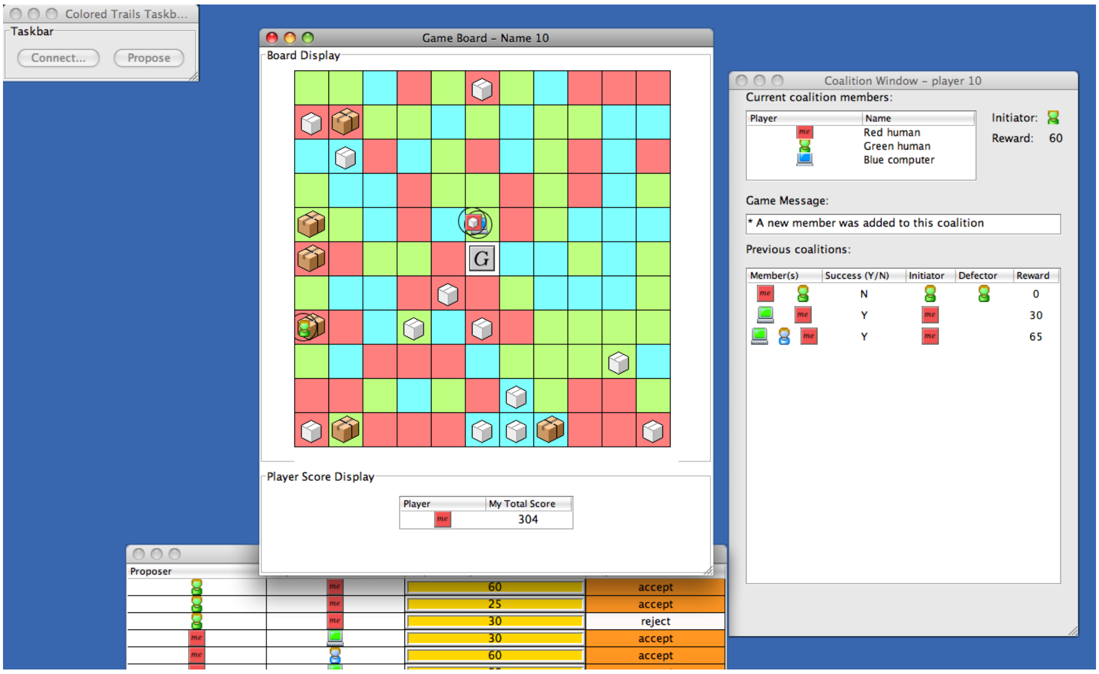


FIGURE C.10: As the game progresses, players can keep track of their history of successful and unsuccessful coalitions in the Coalition Window.

Bibliography

[1] J. Adams. Towards an understanding of inequity. *Journal of Abnormal and Social Psychology*, 67:422–436, 1963.

[2] J. Baron. *Thinking and Deciding.* Cambridge University Press, 1994.

[3] I. van Beest. The social psychology of coalition formation. *Proceedings of ECPR Conference*, 2002.

[4] I. van Beest and E. van Dijk and H. Wilke. The interplay of self-interest and equity in coalition formation. *European Journal of Social Psychology*, 34:547–565, 2004.

[5] I. van Beest and R. Andeweg and C. Koning and P. van Lange. Do groups exclude others more readily than individuals in coalition formation? *Group Processes Intergroup Relations*, 2008.

[6] G. Bernstein and M. Ben-Yossef. Cooperation in intergroup and single-group social dilemmas. *Journal of Experimental Social Psychology*, 30:52–67, 1994.

[7] K. Binmore. *Fun and Games: a Text on Game Theory.* D.C. Heath and Company, 1992.

[8] K. Binmore. *Game Theory: A Very Short Introduction.* Oxford University Press, 2007.

[9] S. Blount. When social outcomes aren't fair. *Organizational Behavior and Human Decision Processes*, 63(2):131–144, 1995.

[10] G. Bolton and J. Brosig. How do coalitions get built: Evidence from an extensive form coalition game with renegotiation and externalities. In *Working Paper Series in Economics*, volume 30. University of Cologne, 2007.

[11] J. Bradshaw. Making agents acceptable to people (abstract of a key-note speech). In *CEEMAS*, pages 1–3, 2003.

[12] S. Breban and J. Vassileva. Using inter-agent trust relationships for efficient coalition formation. In *Proceedings of the 13th Canadian Conference on AI*, 2002.

[13] C. Brooks and E. Durfee. Congregating and market information. In *AAMAS*, 2002.

[14] C. Camerer. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton University Press, 2003.

[15] L. Cameron. Raising the stakes in the ultimatum game: Experimental evidence from indonesia. 1995.

[16] C. Castelfranchi and R. Falcone. Principles of trust for mas: Cognitive anatomy, social importance, and quantication. *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS-98)*, page 7279, 1998.

[17] A. Chavez and S. Kimbrough. A model of human behavior in coalition formation games. In *ICCM*, pages 70–75, 2004.

[18] K. Christoffersen and D. Woods. How to make automated systems team players?, 2008.

[19] M. Dastani. 2apl: a practical agent programming language. *Autonomous agents and multi-agent systems*, 16(3):214–248, 2008.

[20] R. Dawes. Social dilemmas. *Annual Review of Psychology*, 31, 1980.

[21] M. Deutsch. Trust, trustworthiness, and the f scale. *Journal of Abnormal and Social Psychology*, 61(1):138–140, 1960.

[22] R. Falcone and C. Castelfranchi. Social trust: A cognitive approach. In *Trust and Deception in Virtual Societies*, page 5590. Kluwer Academic Publishers, 2001.

[23] X. Fan and J. Yen. Modeling and simulating human teamwork behaviors using intelligent agents. *Physics of Life Reviews*, 2004.

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

[25] Y. Gal and A. Pfeffer. Predicting people's bidding behavior in negotiation. *AAMAS*, 2006.

[26] Y. Gal and A. Pfeffer. Modeling reciprocity in human bilateral negotiation. *AAAI-07*, 2007.

[27] Y. Gal, A. Pfeffer, F. Marzo, and B. Grosz. Learning social preferences in games. *AAAI*, 2004.

[28] G. Gambarelli. Transforming games from characteristic into normal form. *International Game Theory Review*, 09(01):87–104, 2007.

[29] W. Gamson. A theory of coalition formation. *American Sociological Review*, 26(3):373–382, 1961.

[30] M. Gentry. Coalition formation and processes. *Social Work with Groups*, 10(3):3954, 1987.

[31] N. Griffiths. Task delegation using experience-based multi-dimensional trust. *AAMAS*, 2005.

[32] N. Griffiths and M. Luck. Coalition formation through motivation and trust. 2003.

[33] B. Grosz and S. Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, 86(3):269–357, 1996.

[34] B. Grosz, S. Kraus, S. Talman, B. Stossel, and M. Havlin. The influence of social dependencies on decision-making: Initial investigations with a new game. *AAMAS*, 2004.

[35] W. Guth, R. Schmittberger, and B. Schwarz. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, (3):367–388, 1982.

[36] B. Hinckley. Coalitions and time. *A Special Issue of American Behavioral Scientist*, 18, 1975.

[37] B. Hinckley. *Coalitions and Politics*. Harcourt Brace Jovanovich, Inc., New York, 1981.

[38] P. Hinds, K. Carley, D. Krackhardt, and D. Wholey. Choosing work group members: Balancing similarity, competence, and familiarity. *Organizational Behavior and Human Decision Processes*, 81(2):226–251, March 2000.

[39] E. Hoffman, K. McCabe, and V. Smith. On expectations and monetary, stakes in ultimatum games. *International Journal of Game Theory*, 1996.

[40] L. Hogg and N. Jennings. Socially intelligent reasoning for autonomous agents. *IEEE Trans on Systems, Man and Cybernatics - Part A*, pages 381–399, 2001.

[41] C. Jones, K. Fullam, and S. Barber. Exploiting untrustworthy agents in team formation. 2007.

[42] S. de Jong and K. Tuyls and K. Verbeeck. Fairness in multi-agent systems. *The Knowledge Engineering Review*, 2008.

[43] C. Jonker, J. Schalken, J. Theeuwes, and J. Treur. Human experiments in trust dynamics. In *LNCS: iTrust*, pages 206–220. Springer-Verlag Berlin Heidelberg, 2004.

[44] C. Jonker and J. Treur. Formal analysis of models for the dynamics of trust based on experiences. *LNCS*, 1999.

[45] M. Joseph and R. Willis. An experiment analog to two party bargaining. *Behavioral Science*, 8:17–1127, 1963.

[46] J. Kahan and A. Rapoport. Test of the bargaining set and kernel models in three-person games. In A. Rapoport, editor, *Game Theory as a Theory of Conflict Resolution*, pages 119–160. D. Reidel, 1974.

[47] J. Kahan and A. Rapoport. When you don't need to join: the effects of guaranteed payoffs on bargaining in three-person cooperative games. *Theory and Decision*, 8:97–126, 1977.

[48] J. Kahan and A. Rapoport. *Theories of Coalition Formation*. Lawrence Erlbaum Associates, 1984.

[49] E. Kamar and E. Horvitz. Generating shared transportation under varying preferences: Ridesharing models and mechanisms. Technical report, Microsoft Research, April 2009.

[50] B. Kamphorst, A. van Wissen, and V. Dignum. Incorporating bdi agents into human-agent decision making research (forthcoming), 2009.

[51] S. Kraus. *Strategic Negotiation in Multiagent Environments*. MIT Press, 2001.

[52] R. Lin, S. Kraus, D. Tykhonov, K. Hendriks, and C. Jonker. Supporting the design of general automated negotiators. In *Proceedings of the Second International Workshop on Agent-based Complex Automated Negotiations (ACAN'09)*, 2009.

[53] G. Loewenstein, L. Thompson, and M. Bazerman. Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology*, 1989.

[54] M. Macy and A. Flache. Learning dynamics in social dilemmas. volume 99, 2002.

[55] C. Nass, B. Fogg, and Y. Moon. Can computers be teammates? *International Journal of Human-Computer Studies*, 1996.

[56] C. Nass and Y. Moon. Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56:81–103, 2000.

[57] J. von Neumann and O. Morgenstern. Theories of games and economic behavior. *Bulletin (new series) of the American Mathematical Society*, 37, 1944.

[58] A. Okada and A. Riedl. Inefficiency and social exclusion in a coalition formation game: Experimental evidence. *Games and Economic Behavior*, 2004.

[59] M. Olson. *The Logic of Collective Action: Public Goods and the Theory of Groups.* Harvard University Press, 1965.

[60] L. Phlips. *The economics of imperfect information.* Cambridge University Press, 1988.

[61] D. Pruitt and P. Carnevale. *Negotiation in Social Confict.* Buckingham: Open University Press, 1993.

[62] B. Reeves and C. Nass. *The Media Equation : How People Treat Computers, Television, and New Media Like Real People and Places (CSLI Lecture Notes S.).* Center for the Study of Language and Inf, 2003.

[63] A. Riedl and J. Vyrastekova. Responder behavior in three-person ultimatum game experiments. Working Paper, 2003.

[64] J. S. Rosenschein and G. Zlotkin. *Rules of Encounter: Designing Conventions for Automated Negotiation Among Computers.* MIT Press, Cambridge, Massachusetts, 1994.

[65] M. Ross and G. Fletcher. Attribution and social perception. In G. Lindzey and E. Aronson, editors, *Handbook of social psychology.* Random House, 1985.

[66] A. Sanfey, J. Rilling, J. Aronson, L. Nystrom, and J. Cohen. The neural basis of economic decision-making in the ultimatum game. *Science*, (300):1755–1758, 2003.

[67] S. Sen. Reciprocity: a foundational principle for promoting cooperative behavior among self-interested agents. In *Proceedings of the Second International Conference on Multiagent Systems*, pages 315–321. AAAI Press, 1996.

[68] S. Sen and P. Dutta. The evolution and stability of cooperative traits. In *AAMAS*, 2002.

[69] L. Shapley. A value for n-person games. In *Contributions to the Theory of Games 1, Annals of Mathematical Studies 28*, pages 307–317. Princeton University Press, Princeton, NJ, 1953.

[70] Y. Shoham and K. Leyton-Brown. *MultiAgent Systems; Algorithmic, Game-Theoretic, and Logical Foundations.* Cambridge University Press, 2008.

[71] R. Slonim and A. Roth. Financial incentives and learning in ultimatum and market games: An experiment in the slovak republic. *Econometrica*, 1997.

[72] M. van de Vijsel and J. Anderson. Coalition formation in multi-agent systems under real-world conditions. *AAAI*, 2004.

[73] E. Walster, G. Walster, and E. Berscheid. *Equity: Theory and Research.* Allyn and Bacon, Boston, 1978.

[74] A. van Wissen and J. van Diggelen and V. Dignum. The effects of cooperative agent behavior on human cooperativeness. *AAMAS*, 2009.

[75] M. Wooldridge. *An Introduction to MultiAgent Systems.* John Wiley & Sons, 2002.

[76] N. Yorke-Smith, S. Saadati, K. Myers, and D. Morley. Like an intuitive and courteous butler: A proactive personal agent for task management. 2009.