Fairness in multi-agent systems

Fairness in Multi-Agent Systems

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus,
Prof. mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op donderdag 4 juni 2009 om 12:00

door

Steven de Jong

Promotores:

Prof. dr. H.J. van den Herik (Universiteit van Tilburg / Universiteit Leiden)
Prof. dr. E.O. Postma (Universiteit van Tilburg)

Copromotor:

Dr. K. Tuyls (Technische Universiteit Eindhoven)

Beoordelingscommissie:

Prof. dr. ir. R.L.M. Peeters (*voorzitter*)
Prof. dr. M. Gyssens (Universiteit Hasselt)
Prof. dr. N.R. Jennings (University of Southampton)
Prof. dr. H.J.M. Peters
Prof. dr. ir. J.A. La Poutré (Technische Universiteit Eindhoven)

SIKS Dissertation Series No. 2009-13.
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Printed by Optima Grafische Communicatie, Rotterdam

ISBN 978-90-8559-514-4

"A man's ethical behavior should be based effectually on sympathy, education, and social ties [...]. Man would indeed be in a poor way if he had to be restrained by fear of punishment and hope of reward [...]."

A. Einstein, *Religion and Science*
New York Times Magazine, November 9, 1930

Preface

Within the field of artificial intelligence (AI), the research area of multi-agent systems investigates societies of autonomous entities, called *agents*, that need to cooperate or compete in order to achieve a certain goal. Example applications include resource distribution, auctions, and load balancing. In many of these applications, taking into account fairness and social welfare would be desirable. Some applications require agents to interact with humans, who are known to care strongly for fairness and social welfare; in other applications, caring for fairness and social welfare is essential for agents to achieve a satisfactory solution.

Our research aims at making up (some of the) shortcomings of AI with respect to social behavior. We look at two specific problems, i.e., (1) fair division of a limited resource, and (2) the tragedy of the commons. Both problems are considered here in a rather abstract form, but they have many potential applications. They are difficult to solve if agents do not take into account the effect of their actions on others. We start from established approaches that are well-known in the field, i.e., game theory, multi-agent systems, and reinforcement learning. The thesis then discusses a great deal of literature describing human social conditioning, kindness, fairness, and the human tendency to punish those who refuse to behave in a social manner. Such descriptive work is translated to computational models, which may be used by learning agents in multi-agent systems.

This thesis, which bears only one name on the front cover, would not have been as it is now without the assistance and encouragement of many people. In recognition of this fact, the preface contains the only part of the thesis that is written in the first person singular. In the remainder of the thesis, I will write in the first person plural.

My supervisors, Jaap van den Herik and Eric Postma, were the first to be involved in my research. I thank them for allowing me to perform my work under their guidance and for carefully reshaping the text of the thesis. My daily advisor, Karl Tuyls, deserves my sincerest gratitude for putting me on the research track presented here, for the many fruitful meetings we had over the years, and for his personal advice and belief in me, which shaped me as a researcher and as a person.

Partially outside the walls of our building, the Women's Prison, I had the pleasure to meet and work with Ida Sprinkhuizen-Kuyper, Katja Verbeeck, and Peter Vrancx, each of whom substantially influenced my research. Inside the Prison, there were also many opportunities for joint work. In recognition of this joint work, I would like to thank the students I had the pleasure to supervise. In particular, I mention Simon Uyttendaele and Rob van de Ven, since parts of our joint work are integrated in the thesis. I also greatly enjoyed working with my colleagues Guillaume Chaslot, Nyree Lemmens, Marc Ponsen, and Nico Roos, on research that is not reported in the thesis, but elsewhere in joint articles.

A pleasant research environment is not only facilitated by numerous opportunities for cooperation, but also by a friendly, constructive and comforting atmosphere. The supportive staff of MICC helped me in many respects; an explicit thank-you is given to Peter Geurts, Joke Hellemons, and Marijke Verheij. My roommates over the years, Michel van Dartel, Joyca Lacroix, Marc Ponsen, and Benjamin Torben-Nielsen, made our room a place that I enjoyed being in; meanwhile, they taught me many things. I am grateful to Sander Bakkes, Jahn-Takeshi Saito, Maarten Schadd, Evgueni Smirnov, Sander Spek, and Andra Waagmeester for many inspiring hours. Also, I appreciated the facilities being made available to me by Katia Sycara while I was visiting Carnegie Mellon University during the last bit of thesis work in January 2009.

In conclusion to these acknowledgments, I particularly would like to thank my parents, Dirk and Marije de Jong, for everything. Moreover, I would like to recognize the Gouder de Beauregard family, as well as Philippos Dainavas, for their encouragement. I am grateful for my friendship with Stephan Adriaens, Vaughan van Dyk, and Ron Leunissen, whose support is very much appreciated.

Finally, love to Myrna.

Steven de Jong
4 June 2009

Contents

1 Introduction

Sharing limited resources with others is a challenge for individuals in human societies as well as for agents in multi-agent systems. Often, there is a conflict of interest between personal benefit and group benefit. This conflict is most prominently present in *social dilemmas*, in which individuals need to consider not only their personal benefit, but also the fairness of their choices. The concept of fairness has recently received a great deal of attention from two different perspectives, i.e., (1) from the perspective of descriptive models, investigating human fairness mechanisms (Fehr and Schmidt, 1999; Gintis, 2001), and (2) from the perspective of computational models that provide fairness mechanisms in, e.g., multi-agent systems (Chevaleyre et al., 2006; Endriss, 2008).

In practice, humans often deal with social dilemmas in a way that leads to satisfactory rewards. However, current computational models of fairness are commonly not aligned with human mechanisms. As a consequence, they may not be able to reach satisfactory rewards in social dilemmas. To address this issue, we aim to develop *computational models of human-inspired fairness*. After discussing the foundations of such models, we propose three different models and apply them to the two most prominent social dilemmas. In our research, we aim at a concluding statement, i.e., "human fairness mechanisms need to be incorporated in computational models of fairness."

This chapter provides an introduction to the topic of human-inspired fairness in multi-agent systems. In §1.1, we introduce descriptive models of human fairness. In §1.2, we describe multi-agent systems and computational models of fairness. In §1.3, we discuss the presence of an undesirable gap between descriptive models of human fairness on the one hand, and computational models of fairness for multi-agent systems on the other hand. Our research aims at reducing or even bridging this gap. In §1.4, we present our problem statement in combination with five research questions, and in §1.5, we describe the research methodology. In §1.6, we provide the structure of the thesis, and in §1.7, we present a summary of the main contributions we are aiming at.

The work reported on in this chapter has partially been published in:

S. de Jong, K. Tuyls, and K. Verbeeck. Fairness in multi-agent systems. *Knowledge Engineering Review*, Vol. 23(2):153-180, 2008.

S. de Jong, K. Tuyls, and I. Sprinkhuizen-Kuyper. Robust and scalable coordination of potential-field driven agents. *Proceedings of the International Conference on Intelligent Agents, Web Technologies and Internet Commerce (IAWTIC)*, pp. 230–242, 2006.

1.1 Descriptive models of human fairness

Humans show a surprising ability to balance personal benefit and fairness. In a variety of complicated problems, they optimize their personal benefit while taking into account the effects of their actions on others. Most prominently, humans usually have little difficulty proposing solutions to a class of problems known as *social dilemmas*. In social dilemmas, individuals have to choose between being selfish (i.e., being individually rational and caring only for their own benefit) and being social (i.e., being driven by fairness considerations, which means also taking into account the benefit of others). The dilemma lies in the fact that being individually rational may lead to a lower benefit than being fair. There are two distinct reasons why this may happen, resulting in two different social dilemmas, which we will refer to as the *agreement dilemma* and the *tragedy of the commons*.

In the agreement dilemma, individuals need to agree with each other, as lack of agreement may lead to frustration, rejection, and failure (in that order). The issue is that humans may reject a positive reward if they consider this reward to be unfair. A stylized example of such a situation is the *Ultimatum Game* (Gueth et al., 1982), in which two players bargain about the division of a reward which is received from an outsider (e.g., €10). The first player is assumed to offer a certain amount to the second player, say €4, which the second player may then accept or reject. In case of rejection, both players end up with nothing. An individually rational first player would offer the second player €0.01, the lowest positive amount possible. In contrast, human first players consistently offer more than the lowest amount. Moreover, low offers are almost always rejected by human second players (Fehr and Schmidt, 1999). Interestingly, approximately 9 out of 10 people playing the Ultimatum Game are immediately able to reach a successful deal with an unknown opponent from the same cultural background (Henrich et al., 2004; Oosterbeek et al., 2004). This may be a 50-50 split, but also, e.g., a 80-20 split – the precise agreement depends on many factors, such as culture (Henrich et al., 2004) and wealth (De Jong et al., 2008).

In the tragedy of the commons, agreement is not sufficient, as awareness of the benefit of others is explicitly required to obtain a satisfactory solution (Hardin, 1968). A stylized example here is the *Public Goods Game* (Binmore, 1991), in which players have to decide whether or not to invest money in a common pool, say €10. The total sum will be multiplied by some factor (e.g., 3). The resulting amount of money will be evenly distributed among all players. Clearly, individual players gain more if they do not invest any money while the others do; however, if all players reason in this way, they will not gain money at all. In contrast to an individually rational player, a typical human player is willing to pay a small fee (e.g., €1) to punish those who refuse to invest (e.g., by reducing their final reward by €4). Such punishment deters the players from refusing to invest and therefore leads to a situation in which everyone gains money (Sigmund et al., 2001; Fehr and Gaechter, 2002).

Researchers in behavioral economics have conducted many experiments with humans in order to unravel the human decision process in social dilemmas (Bowles et al., 1997; Fehr and Schmidt, 1999; Henrich et al., 2004; Oosterbeek et al., 2004; Dannenberg et al., 2007). This decision process has been captured in *descriptive models*, aimed at clarifying why and

how fairness influences the decisions of humans. Most importantly, researchers find that humans respond to unfair interactions using two mechanisms, i.e., (1) by performing *altruistic punishment*, implying that someone who is perceived to act in an unfair way is treated with an immediate negative effect on his final reward (Fehr and Schmidt, 1999), and (2) by *withholding action*, implying that an unfair actor is somehow excluded from future interactions (Hauert et al., 2002).

Thus, both mechanisms pose a clear threat to those willing to act in an unfair manner, and therefore effectively allow humans to enforce fairness. However, in both cases, the decision whether to apply these mechanisms essentially entails a so-called second-order social dilemma (see, e.g., Panchanathan and Boyd, 2004), as applying them may negatively influence any individual reward. Punishing someone generally requires some investment (Fehr and Gaechter, 2002; Boyd et al., 2003) (hence the term 'altruistic punishment'), and refusing to interact with someone in an interaction with (generally) an expected positive reward clearly hurts any expected reward (Hauert et al., 2002). Thus, an individually rational strategy would be to refrain from applying altruistic punishment or from withholding action.

Among others, the most prominent forces assumed to drive humans to applying the mechanisms of altruistic punishment and withholding action, regardless of the fact that doing so is not individually rational, are *inequity aversion* (i.e., resistance to overly large differences), *reciprocal fairness* (i.e., sensitivity to additional information, such as reputation), and *social networks* (i.e., in many circumstances, local interactions are more frequent than distant interactions, and some people interact more than others).

1.2 Multi-agent systems and computational models of fairness

Multi-agent systems are generally accepted as valuable tools for designing and building distributed dynamical systems consisting of several interacting agents, possibly including humans (Jennings et al., 1998; Weiss, 1999; Ferber, 1999; Shoham et al., 2007). While there is no general definition for concepts such as agents and multi-agent systems, we adopt the following definitions (cf. Jennings et al., 1998).

Definition 1.1 An *agent* is an entity, situated in some environment and perceiving this environment through its sensors. In order to meet its objectives, the agent is capable of flexible, autonomous actions.

Definition 1.2 A *multi-agent system* is a loosely coupled network of agents that cooperate or compete with the goal of solving problems that are beyond the individual capabilities of each agent.[1]

[1] In the remainder of this thesis, we use the term 'agent' to refer to both computer agents as well as humans, unless a clear distinction needs to be made. Multi-agent systems may thus be comprised of human agents as well as computer agents. Moreover, when referring to an agent, for brevity, we use 'he' and 'his' wherever 'he or she or it' and 'his or her or its' is meant.

A central problem in multi-agent systems is resource allocation (Chevaleyre et al., 2006; Endriss, 2008). Example domains in which resource allocation is explicitly considered include online auctions or bargaining (Preist and van Tol, 1998; Erev and Roth, 1998; Kalagnanam and Parkes, 2004), electronic institutions (Rodriguez-Aguilar, 2003), developing schedules for air traffic (Mao et al., 2006), and decentralized resource distribution in large storage facilities (Weyns et al., 2005; De Jong et al., 2006c). Moreover, even outside problem domains in which agents deal explicitly with resources, agents are using resources (e.g., bandwidth, memory or computational resources), which usually need to be shared (Endriss, 2008).

Allocating resources is a challenging task, since agents may have their own specific preferences concerning personal benefit, which may be incompatible with other agents' preferences. A satisfying allocation should respect and balance these preferences by considering concepts such as fairness in addition to pure personal benefit. Thus, agents can no longer be assumed to be designed according to the principles of classical game theory, i.e., to be purely self-interested and individually rational. Researchers have therefore proposed to learn from the field of welfare economics, in which concepts such as personal benefit and fairness have been extensively studied, resulting in *computational models* of fairness (Sen, 1970; Chevaleyre et al., 2006; Endriss, 2008).

Multiple operationalizations of the notion of 'computational fairness' have been proposed, coupled with measures of collective utility that tie in with the operationalizations. Examples include utilitarian social welfare (i.e., optimizing the average performance) and egalitarian social welfare (i.e., optimizing the performance of the least-performing agent). It is important to note that these measures are mostly not directly based on human decision-making. As a result, the human mechanisms of altruistic punishment and withholding action are commonly not motivated by such measures.

1.3 The need for human-inspired fairness in multi-agent systems

As has been outlined above, existing computational models of fairness rely on the assumption that agents aim to optimize a certain measure of collective utility. In other words, they assume that all agents participating in an interaction explicitly care about fairness.

There are two problems with this assumption in relation to social dilemmas. First, existing measures of collective utility commonly do not lead to fair solutions in social dilemmas, even if we assume that all agents aim to optimize collective utility.[2] Second, we may actually not always assume that all agents care about fairness. If agents that care about fairness are allowed to interact with agents that do not (e.g., individually rational agents, or agents aiming to exploit the fairness), unfair and suboptimal solutions may result from the fair agents optimizing their collective utility. For instance, with utilitarian social welfare, the fair agents only care about average performance, which implies that they will not mind if a certain agent steals the entire reward of another agent.

[2] We note that we will elaborately discuss this matter in §3.3.1.

Thus, there is an need for the improvement of the performance of computational models of fairness in social dilemmas. We already mentioned that the mechanisms of altruistic punishment and withholding action are effective when we are aiming to drive a population of *human* agents to cooperative, fair solutions in social dilemmas (Fehr and Schmidt, 1999; Gintis, 2001). These mechanisms do not require the assumption that all agents actually explicitly care about fairness (i.e., group benefit) in addition to personal benefit. With altruistic punishment and/or withholding action, we can force agents to care for group benefit, simply by caring for their personal benefit only, as their personal benefit will be harmed if they decide to perform actions that other agents may not consider desirable.

A logical conclusion is therefore that there is a need for the inclusion of human-inspired computational models of fairness in multi-agent systems. As a case in point, we provide two typical applications for multi-agent systems that will benefit from including such models. The first application is taken from our own earlier work; it contains elements of the agreement dilemma. The second application is taken from the work by Verbeeck et al. (2007); it contains a dilemma similar to the tragedy of the commons.

The agreement dilemma

The agreement dilemma is prominently present in applications that are meant to provide a service to humans, for instance, in online auctions or distributed robotics. Clearly, human customers must be satisfied with the given service, as they will otherwise decide to end their customer relationship with the service provider (potentially moving to a competitor that provides better service). In other words, with the agreement dilemma, the issue is that mechanisms such as altruistic punishment and withholding action are already present. This implies that the service provider should keep customers sufficiently satisfied.

An example application is the task of resource gathering. In this task, there are various locations, distributed over a potentially quite large environment, at which resources need to be picked up. We developed two approaches to the task (both using distributed, simulated robotic systems), viz. (1) an approach based on gradient descent, potential-field theory, and planning (De Jong et al., 2006a,b,c), and (2) an approach based on bee behavior (Lemmens et al., 2007, 2008). For an appropriate comparison of these approaches with each other and with existing work, we needed to select a suitable performance measure. This turned out to be difficult, since any chosen performance measure strongly influences what is meant by 'good performance'. For instance, comparing the two approaches, the average delay experienced by customers may be lower for one of them, but the standard deviation may also be notably higher. In other words, some customers may wait intolerably long (Weyns et al., 2005). Thus, to decide which approach is actually better depends on the evaluation criteria considered. We aimed at resource gathering systems that would be working as a service to human customers (e.g., in warehouses). We therefore wished to incorporate a 'human' performance measure that would enable us to measure how satisfied customers would be.

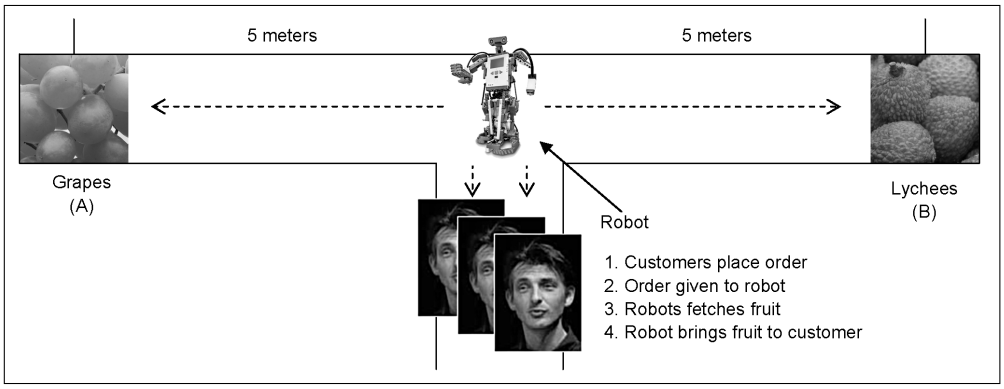


Figure 1.1 This simple fruit shop demonstrates the need for human-inspired fairness in multi-agent systems

To determine which measures would be important to customers, we developed a small, challenging resource gathering task (see Figure 1.1). We visualized a rather primitive linear fruit shop, selling only *grapes* (stored at A) and *lychee's* (stored at B). A robot, initially located somewhere between A and B, needs to fetch fruit for a customer who is waiting in the middle. The robot is assumed to be immediately available to the customer, but does not know in advance which of the two fruits the customer wishes to obtain. However, there is a certain probability $0.5 < p < 1$ that the customer wishes to obtain grapes.

The task was given to an audience of 50 faculty members and students. They were asked to determine the best initial position for the robot on the line between A and B, given certain values for p. There are at least two rational solutions. More precisely, for any $0.5 < p < 1$, placing the robot at A minimizes the expected waiting time experienced by the customer[3], while placing it in the middle minimizes the maximum waiting time.[4] However, 45 of the 50 participants did not choose a rational solution, such as minimizing the expected waiting time by placing the robot at A. Instead, given $p = 0.6$, the robot was placed somewhere between A and the middle (in fact, the participants usually placed the robot near the center of gravity, given that we interpret the two probabilities as masses). When we increased the probability that an item stored at A would be requested, the robot was placed more to the left (i.e., closer to A), but only for very high probabilities was A chosen as the location for the robot. Thus, it seemed that our participants favored the requesters of the more popular item with only a slight advantage over the requesters of the less popular item.

[3] Given that the robot is located at some x between A and B, where A corresponds to $x = 0$, customers requesting grapes have the robot travel a distance of $d = 5 + x$, and customers requesting lychees have the robot travel $d = (10 - x) + 5 = 15 - x$. The expected distance traveled given x and p, the probability that a customer requests grapes, is therefore $E(d|x, p) = p(5 + x) + (1 - p)(15 - x)$. This evaluates to $E(d|x, p) = (2p - 1)x + 15 + 20p$. For $0.5 < p < 1$, we obtain that $2p - 1$ is positive and therefore the expected distance traveled increases with x. Thus, $\min_x E(d|x, p) = 0$.

[4] Given the analysis directly above, we need to find $\min_x \max_x \{5 + x, 15 - x\}$, which is $x = 5$.

This challenging example task demonstrates that there is a gap between human performance measures on the one hand, and rationally optimal performance measures on the other hand. As we explained above, the task is one of the many examples of a social dilemma we named the agreement dilemma. Games such as the Ultimatum Game pose more abstract, but also more risky agreement dilemmas, as insufficient agreement leads to no reward at all.

Thus, in addition to individually rational performance measures and the fairness measures proposed by welfare economics, we need agents that interact with humans to be able to use human (or human-inspired) performance measures.

The tragedy of the commons

In contrast to the agreement dilemma, the tragedy of the commons is also prominently present in applications that do not directly relate to humans. In fact, we may argue that all modern computer applications have to deal with the tragedy of the commons, as they require computational resources that need to be shared with other applications in a balanced way. In multi-agent-systems applications, agents are often relatively independent entities, either working on a task that explicitly concerns resource sharing, or having to share resources to be able to address their actual task (Endriss, 2008).

In the agreement dilemma, as discussed above, we see that mechanisms such as punishment and withholding action are already present, and require agents to consider the impact of their strategies on other agents' benefits. In the tragedy of the commons, the situation is roughly reversed: to obtain a satisfactory benefit, agents need to consider the impact of their strategies on other agents' benefits, and mechanisms such as punishment and withholding action may be introduced to achieve this.

An example application is load balancing (Bourke, 2001). Assume that agents need to perform certain calculation tasks. They may perform their tasks on a relatively slow client computer, or they may move them to a much faster, shared server. It is in everyone's personal interest to use the server, but then, the server becomes overused, while the computational power offered by the clients is not used at all. Using human-inspired fairness models may motivate agents to become more aware of the balance they need to achieve. For instance, we may punish agents that have recently over-used the server by prohibiting them from using it in the near future, or the server may refuse to perform their calculations. Indeed, Verbeeck et al. (2007) applied a reinforcement-learning technique that was (remotely) inspired by a descriptive model of human fairness, i.e., inequity aversion. The authors show how agents may reach a satisfactory load balance if they are willing to give up some of their reward in order to increase the reward of others. However, it is assumed that the agents are cooperative; in this case, we may indeed consider that all agents will be willing to give up reward if they know that this benefits the group.

Games such as the Public Goods Game are abstract instantiations of the tragedy of the commons in which we usually do not assume that agents are cooperative. A mechanism allowing

agents to achieve a satisfactory outcome in the Public Goods Game may also be applied in similar, less abstract situations, for instance, competitive load balancing.

Thus, in addition to individually rational performance measures and the fairness measures proposed by welfare economics, we need agents that have to deal with the tragedy of the commons to be able to use human (or human-inspired) mechanisms.

1.4 Problem statement and research questions

Humans have the ability to deal with social dilemmas adequately. Given the fact that many multi-agent systems are facing such dilemmas regularly, we argue that agents in multi-agent systems need to have this ability as well. Considering the need for human-inspired fairness in multi-agent systems, the problem statement (PS) of this thesis reads as follows.

PS *How can we obtain human-inspired fairness in multi-agent systems?*

Considering this problem statement, it is of obvious importance that we should operationalize what we mean by the term 'human-inspired fairness'. Therefore, part of the research will be directed towards establishing a satisfactory operationalization of human-inspired fairness for our purposes, i.e., an operationalization that is human-inspired as well as computationally applicable in multi-agent systems. To this end, we need to determine how humans are using fairness in their decisions. This leads to the first research question.

RQ1 *How are humans using fairness in their decisions?*

We have already briefly outlined that the main decisions humans make due to fairness are (1) to punish others that are perceived as behaving in an unfair way, and/or (2) to refrain from interacting with such others in the future. In this thesis, these two decisions are further investigated. Once we have established a clear view on human fairness and how it leads to better solutions for various problems, we may start looking at human-inspired computational fairness models. To this end, we first need to discuss two elements, i.e., (1) the requirements for such models, and (2) the common elements of each model we will consider. In the remainder of the thesis, these elements will be referred to as the *foundations* of human-inspired computational fairness. Our second research question therefore reads as follows.

RQ2 *What are the foundations of human-inspired computational fairness?*

Once these foundations have been established, we can use them to develop actual computational models of human-inspired fairness, i.e., models that can be used in multi-agent systems. This leads to the third research question.

RQ3 *How can human-inspired fairness be modeled computationally, taking into account the foundations of human fairness?*

In this thesis, we will present three computational models of human-inspired fairness. These computational models must be extensively validated. To this end, we need to perform both

analytical studies of the models' properties as well as experimental studies, with the goal of determining (1) whether the models are indeed computationally applicable (i.e., whether agents in multi-agent systems are able to behave in a desired manner, according to the models) and (2) whether outcomes obtained by multi-agent systems driven by the models, conform to human outcomes (which are better than outcomes obtained by the existing models). Thus, we formulate RQ4 and RQ5 as follows.

RQ4 *What are the analytical properties of the computational human-inspired fairness models developed in this research?*

RQ5 *What are the (empirical) benefits of incorporating explicitly human-inspired fairness in adaptive multi-agent systems?*

We note the addition of the word 'adaptive' in RQ5. More precisely, we mean 'adaptive' according to the principles of multi-agent reinforcement learning (Sutton and Barto, 1998). In the research presented in this thesis, we restrict ourselves to multi-agent systems in which agents learn to behave using positive and/or negative reinforcement. Our learning agents are confronted with a variety of social dilemmas, modeled in single-stage games known from game theory, such as the Ultimatum Game and the Public Goods Game. In contrast to existing work on reinforcement learning in social dilemmas, which focuses on games with (small) discrete strategy sets, we consider almost exclusively games with a continuous strategy space. We use learning automata (Narendra and Thathachar, 1989) and continuous-action learning automata (Thathachar and Sastry, 2004), which have a proven convergence to (local) optima when the feedback from the environment is sufficiently smooth.

Our research therefore does not propose new multi-agent reinforcement techniques. Instead, we investigate how we may set up reinforcement mechanisms that enable agents to learn fair strategies, given that they are equipped with an existing reinforcement learning algorithm. We propose novel ideas in the application of individual utility functions (e.g., inequity aversion, priority awareness) as well as game dynamics (e.g., networks of interaction, reputation). These utility functions and game dynamics are eventually assembled into computational fairness models that do include important elements of human solutions regarding social dilemmas; thus, we enable multi-agent systems to benefit from the human ability to deal with these interesting and challenging problems.

1.5 Research methodology

In this section, we discuss our research methodology, as related to the problem statement and research questions stated above. The methodology consists of eight steps, which may be equally divided into two parts of four steps each, i.e., step 1–4 concerning descriptive models of human fairness and the corresponding research question RQ1, and step 5–8 concerning computational models of human-inspired fairness and the corresponding research questions RQ2–RQ5. We present the eight steps in the subsections §1.5.1 and §1.5.2 below.

1.5.1 Descriptive models of human fairness

Below, we list the four research steps that we will perform to answer RQ1.

1. Literature study. We start by performing an extensive literature study in the field of behavioral economics and identify three main descriptive models of human fairness. The key ideas of these three models are (i) *inequity aversion*, (ii) *reputation*, and (iii) *social networks*.

2. Analysis. From the results found by means of our literature study, we identify two possible opportunities. First, we analyze altruistic punishment in public-goods interactions, which is not easily explained using existing descriptive models. Second, as a larger opportunity, we will analyze all three models. Slightly anticipating on the outcome of our analysis, we here mention that the three models are missing an important element (which has been present in classical game theory for a long time), viz. bargaining power, also called immediate reputation or (as we call it throughout this thesis) *priority*.

3. Design. To address the missing element (priority), we provide an addition to the existing reputation model, by designing a new model of human fairness, i.e., *priority awareness*. The model stipulates that humans do not necessarily need repeated interactions to be able to classify another person as being nice or nasty; additional (explicit or implicit) information that they may have about this other person immediately influences their strategies. This is also reflected in the outcomes of the small experiment presented in §1.3.

4. Validation. Using analyzes and experiments, we will validate to what extent the priority-awareness model adequately predicts human behavior in various settings. The model attempts to be appropriate when describing immediate as well as emerging reputation.

1.5.2 Computational models of human-inspired fairness

Below, we list the four research steps that we will subsequently perform to answer RQ2–RQ5. For later reference, we continue the numbering of the steps. Note that 'design' and 'analysis' occur twice in the final list of eight steps.

5. Formalization. We have investigated three descriptive models of human fairness, each of which has its own specific way of describing human decision-making in social dilemmas. Since RQ2 asks for the foundations of human-inspired computational fairness models, we formalize the requirements as well as the basic principles of these models.

6. Design. Once we have obtained the foundations, RQ3 asks how we may arrive at actual models. The three descriptive models are translated into three computational models of human-inspired fairness.

7. Analysis. The three computational models of human-inspired fairness need to be analyzed so that we may understand how they drive agents to expected behavior. To this end, we address RQ4. We examine the theoretical outcomes of social dilemma interactions given our computational models.

8. Experiments. The computational models have been grounded theoretically. They then need to be applied in multi-agent systems, as assumed in RQ5. Using multi-agent learning algorithms and our three computational models of fairness, we construct three adaptive multi-agent systems that use a computational instantiation of human fairness. We will perform experiments with these multi-agent systems in order to determine whether they are able to address the problems in which we are interested, i.e., social dilemmas.

Thus, by this last step, we aim at providing an answer to the problem statement and at showing how a mechanism can be designed for obtaining computational, human inspired fairness in adaptive multi-agent systems.

1.6 Structure of the thesis

The thesis is structured as follows. Chapter 1 provides a general introduction. The precise formulations of the problem statement and five research questions are given, as well as the methodology. We complete the chapter by summarizing the goals we are aiming at, which are formulated as two contributions that we envisage as results from our research.

In Chapter 2, we describe the background of the thesis, i.e., game theory and multi-agent reinforcement learning. Moreover, we discuss the social dilemmas that will be studied throughout the remainder of this thesis.

The five research questions are addressed in Chapters 3 to 6. RQ2 constitutes a special case, since it concerns the formulation of the foundations of computational, human-inspired fairness. This research question is therefore addressed before the others, i.e., in Chapter 3.

Chapters 4, 5, and 6 follow a similar structure, as each of them discusses RQ1 concerning descriptive modeling of human fairness and RQ3–RQ5 concerning computational models of human-inspired fairness. In each chapter, we address RQ1 by discussing a specific descriptive model of human fairness (i.e., inequity aversion in Chapter 4, reputation and priority awareness in Chapter 5, and fairness in social networks in Chapter 6). We then create a computational model of fairness, incorporating this specific descriptive model (RQ3), analyze the computational model (RQ4), and use the model in an adaptive multi-agent system that is learning to find good solutions to social dilemmas (RQ5).

In Chapter 7, we provide our conclusions by answering our research questions and problem statement. We also give ideas for future work.

1.7 Contributions of the thesis

In this section, we summarize the two main contributions aimed at in this thesis, and indicate how these contributions relate to the research questions.

The first contribution relates to descriptive models and RQ1. This thesis aims to present a coherent overview of the current state of the art in descriptive modeling of human fairness.

It also aims at addressing any possible important concepts that are missing in current descriptive models of human fairness. To this end, we apply an existing descriptive model to a number of games it has not yet been applied to. Moreover, in particular, we investigate an identified missing concept, viz. the human tendency to use additional information immediately. To address this concept, we introduce our own descriptive model of reputation, named priority awareness.

The second contribution relates to computational models and RQ2 to RQ5. We aim at establishing the foundations of human-inspired computational fairness models (RQ2), i.e., the requirements and the basic principles for such models. Building upon these foundations, we develop computational models of human-inspired fairness (RQ3). Theoretical results for these models will be analyzed (RQ4) and compared to empirical results, as obtained by adaptive multi-agent systems in which agents learn behavior driven by these models (RQ5).

2 Background

Introducing fairness in multi-agent systems requires a thorough understanding of how local interactions between independent agents may lead to desired global outcomes. To this end, background knowledge of research areas such as game theory and reinforcement learning is required. In this chapter, we discuss the relevant background knowledge. Moreover, we further explain our problem domain, i.e., social dilemmas. In §2.1, we give a brief overview of the 'classical' agent model, i.e., the rational solution concepts proposed by game theory. We also look at the limitations of game theory. Next, §2.2 presents a brief overview of multi-agent reinforcement learning, specifically focusing on learning automata, an approach suitable for our purposes. In §2.3, we discuss game-theoretic games known as social dilemmas, which constitute our problem domain. We complete the chapter by a summary (§2.4).

The work reported on in this chapter has partially been published in:

S. de Jong, K. Tuyls, and K. Verbeeck. Fairness in multi-agent systems. *Knowledge Engineering Review*, Vol. 23(2):153-180, 2008.

2.1 Game theory

Game theory is a mathematical formalism used to model and analyze strategic interactions between different agents. It started as an economic discipline in the 1940's, when von Neumann and Morgenstern (1944) published a book that coined the term.[1] In the post-war years, the discipline quickly gained ground, especially after Nash (1950a) contributed his well-known theory on non-cooperative games. Due to its power to explain how certain outcomes result from interactions between individually rational entities, game theory has been extensively applied in economics, politics, and social sciences. It also found its way into multi-agent systems and still has a profound influence on the assumptions usually taken into account when such systems are designed.

In this section, we discuss the fundamentals of game theory and look at the most important game-theoretical solution concepts. First, we discuss normal-form games (§2.1.1). Second, we discuss utility functions, which have a central role in our work (§2.1.2). Third, we discuss the most common solution concepts, i.e., the Nash equilibrium, Pareto-optimality, and the Nash bargaining solution (§2.1.3). Fourth and last, we discuss that the predictions of game theory are often not aligned with real-world observations (§2.1.4).

Except where noted otherwise, we refer to Binmore (1991), Osborne and Rubinstein (1994) and Gintis (2001) for a more elaborate discussion.

2.1.1 Normal-form games

In game theory, interactions are modeled in the form of games. More precisely, we assume a collection of n agents participating in an interaction where each agent i has an individual finite set of actions

$$A^i = \left\{a^i_1, \ldots, a^i_{m^i}\right\}.$$

Thus, the number of actions available to the agent i is denoted as m^i, or simply as m if all agents have the same number of actions available.

The agents play a game in which each agent i independently selects an individual action a^i from his private action set A^i. The combination of actions of all agents constitute a joint action a from the joint action set $\mathbb{A} = A^1 \times \ldots \times A^n$. A joint action a is thus a vector in the joint action space $\mathbb{A}$, with components $a^i \in A^i, i : 1 \ldots n$.

For every agent i, the reward function $R^i : \mathbb{A} \to \mathbb{R}$ denotes their reward or payoff, given a joint action $a \in \mathbb{A}$. Considering the entire collective of n agents, the reward distribution $R : \mathbb{A} \to \mathbb{R}^n$ is defined as a vector in $\mathbb{R}^n$, with components $R^i \in \mathbb{R}, i : 1 \ldots n$. In the remainder of the text, we use r^i as a shorthand notation for $R^i(a)$ and r for $R(a)$.

[1] The early history of game theory is rather complicated. A discussion falls outside the scope of this thesis. In summary, we remark that von Neumann already did some work before publishing his book. Moreover, in hindsight, work of others, such as Zermelo (1907), Borel (1921), and Zeuthen (1930), can also be considered to be game-theoretical. See, for instance, Weintraub (1992), Schwalbe and Walker (2001), and Walker (2005).

	Silent	Betray
Silent	$(-1,-1)$	$(-10,0)$
Betray	$(0,-10)$	$(-5,-5)$

Table 2.1 The reward matrix for the Prisoner's Dilemma Game (for details, see text).

If $\sum_i r^i = 0$ for all joint actions a, the game at hand is called a zero-sum game; otherwise, it is a general-sum game.

The tuple $(n, \mathbb{A}, r^{1\ldots n})$ defines a single-stage strategic game, also called a *normal-form game*. In contrast to extensive-form games, which are commonly represented using a tree, normal-form games are commonly represented in a matrix form. The rows and columns represent the actions for agent 1 (the row player) and agent 2 (the column player), respectively. In the matrix the reward distributions for every joint action can be found; the first (second) number in every table cell is the reward for the row (column) agent.

As an example, we consider the Prisoner's Dilemma Game, popularized by Axelrod (1984). In this two-agent strategic normal-form game, two suspects are arrested. The police has insufficient evidence for immediate conviction. Having separated both suspects, they visit each of them and offer the same deal: if one testifies against the other, and the other remains silent, the betrayer goes free and the silent accomplice receives the full 10-year sentence. If both remain silent, both suspects are sentenced to only a year in jail for a minor charge, due to lack of evidence. If each betrays the other, each receives a five-year sentence. Each suspect must make the choice of whether to betray the other or to remain silent. However, neither suspect knows for sure what choice the other suspect will make. The dilemma lies in the fact that betrayal yields a better reward than staying silent, assuming we do not know what the other suspect chooses to do. However, if both suspects reason in this way, they will betray each other (i.e., they will play the joint action (*Betray, Betray*)) and thus have to spend a longer time in prison than if they had both remained silent (i.e., if they had played the joint action (*Silent, Silent*)). The matrix representation of the Prisoner's Dilemma Game is given in Table 2.1. Since spending time in jail can be considered a 'negative' reward, we use negative numbers for the rewards. Commonly, the rewards are transformed in such a way that all rewards are positive.

With regard to action selection, agents can adopt either a pure strategy or a mixed strategy. In a pure strategy, a single specific action is always played, i.e., the action is selected with probability 1. In a mixed strategy, a set of actions is played with a certain probability per action. More formally, agent i's strategy p^i is a vector taken from an m-dimensional strategy space P^i. Each element $p^i_k \in p^i$ indicates the probability that action a^i_k is chosen by agent i, i.e., $p^i_k = Pr(a^i = a^i_k)$. Clearly, the individual components of a strategy p^i should sum to 1. Joint strategies p are elements of the joint strategy space $\mathbb{P} = P^1 \times \ldots \times P^n$, which contains all possible strategies for the n agents.

Game-theoretic games may require agents to select only one action, but they may also be extended to multiple actions per agent. In the first case, games are said to be single-stage as well as single-shot. Extensions may be performed along two different dimensions. First, games may require agents to perform a series of actions. Such games are denoted as *multi-stage* games. Second, games may be played repeatedly instead of only once. Such games are denoted as *iterated* games. Agents are traditionally assumed to know that they are participating in an iterated game. This knowledge may affect the agents' decisions. For instance, in the iterated Prisoner's Dilemma Game, agents may choose to stay silent (Axelrod, 1984), whereas the single-shot game usually leads to agents betraying each other.

2.1.2 Utility functions

In addition to a reward function, game theory also introduces the concept of a utility function. Reward functions are defined by the problem at hand, but do not capture what agents actually want, or how they actually perceive their reward. As our research is concerned with developing computational models of human-inspired fairness, the perception of a reward is (even) more important than the actual reward itself, as this perception may be influenced by other factors than the reward only. In the remainder of this thesis, we will therefore consider that agents are striving to optimize their utility instead of their reward (as is commonly assumed).

Formally, utility functions define a mapping from a reward distribution $R \in \mathbb{R}^n$ to an individual utility value, i.e., $U^i : \mathbb{R}^n \rightarrow \mathbb{R}$. The utility distribution $U : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is defined as a vector in $\mathbb{R}^n$ with components $U^i \in \mathbb{R}, i : 1 \ldots n$. As with rewards, we use the shorthand notations $u^i = U^i(R(a))$ and $u = U(R(a))$, as well as $u > u' \leftrightarrow \forall i : u^i > u'^i$.

Given a certain joint strategy $p \in \mathbb{P}$, agents may need to know their expected utility. Expected utility may be calculated as follows. First, consider a certain joint action

$$a_c = (a^1_{c^1}, \ldots, a^n_{c^n}).$$

The probability that this joint action is executed by the agents, is given by the agents' strategies. Given for the individual actions in the joint action that

$$p^i_{c^i} = Pr\left(a^i = a^i_{c^i}\right),$$

the probability for the entire joint action to be chosen is

$$p^{a_c} = \prod_i p^i_{c^i}.$$

The expected reward for agent i due to a joint action a_c is therefore $p^{a_c} \times R^i(a_c)$. The expected reward for agent i over the entire joint strategy then evaluates to

$$\tilde{R}^i(p) = \sum_{a_c \in \mathbb{A}} p^{a_c} \times R_i(a_c).$$

Then, the expected utility over the joint strategy may be calculated as the utility resulting from the expected reward distribution $\tilde{R}$. We once again introduce a shorthand notation, i.e., $\tilde{u}^i = U^i(\tilde{R}(p))$ and $\tilde{u} = U(\tilde{R}(p))$.

2.1.3 Game-theoretic solution concepts

Game theory aims at providing agents with ways to derive which strategies to choose, i.e., which actions to play, and possibly with which probability per action. More precisely, there may be various joint strategies that adhere to some definition of optimality. In game theory, such joint strategies are known as *equilibria*. We will discuss the four most important equilibrium concepts and solutions, i.e., the Nash equilibrium, Pareto optimality, the Nash bargaining solution, and the generalized Nash bargaining solution.

The Nash equilibrium

The most well-known solution to game-theoretical interactions is the Nash equilibrium.

Definition 2.1 The agents are in a Nash equilibrium when no agent can increase his expected utility by changing his strategy while the other agents remain at their strategies.

In other words, there is no incentive for the agents to play any other strategy than their Nash equilibrium strategy. Formally, with respect to any agent i, we denote with p' any alternative joint strategy where $p'^i \neq p^i$, while $\forall j \neq i : p'^j = p^j$. Then, the current joint strategy p is a Nash equilibrium iff $\forall i \forall p' : U^i(\tilde{R}(p)) \geq U^i(\tilde{R}(p'))$.

Nash (1950a) proved that every strategic game with finitely many actions has at least one mixed Nash equilibrium. A mixed Nash equilibrium consists of playing a mixed strategy, i.e., a set of actions with a certain probability per action, in contrast to a pure Nash equilibrium, in which a single action is always played.

In the Prisoner's Dilemma Game, the Nash equilibrium is a pure one, i.e., joint action (*Betray, Betray*). In other words, both agents should betray each other, yielding a jail sentence of 5 years for both. Interestingly, in contrast to this 'rational' solution, it is observed that on average, 40% of human players actually choose to stay silent (Tversky, 2004).

Pareto optimality

Despite Nash's proof that a Nash equilibrium exists for every strategic game with a finite action set, the problem of describing an optimal solution for strategic games is still not completely addressed. For instance, an equilibrium point is not necessarily unique. If more than one equilibrium point exists, they do not always give the same reward to the agents. Pareto optimality is a second solution concept, introduced in game theory to distinguish between multiple equilibria.

Definition 2.2 The expected outcome of a game is Pareto-optimal if there exists no other expected outcome for which all agents simultaneously perform the same or better.

Formally, we denote by $p \succ p'$ that a joint strategy p Pareto-dominates p'. This means that all agents receive at least the same expected reward given p as given p', i.e., $p \succ p' \rightarrow \forall i : U^i(\tilde{R}(p)) \geq U^i(\tilde{R}(p'))$. A joint strategy p is Pareto-optimal iff $\forall p' : p \succ p'$.

The Prisoner's Dilemma game is actually the classical example of a game in which the Nash equilibrium leads to an inferior solution. As we saw in the previous section, this game has just one pure Nash equilibrium, i.e., the joint action (*Betray, Betray*). However, it is not a Pareto-optimal equilibrium. The joint action (*Silent, Silent*) is obviously superior; it is the only strategy which Pareto-dominates the Nash equilibrium, but it is not a Nash equilibrium itself. Other games in which the individually rational action is not a 'good' one, regardless of what other agents' actions are, include the Traveler's Dilemma (Basu, 2007) and the Public Goods Game (see §2.3).

The Nash bargaining solution

A bargaining solution is a way in which agents agree to divide a certain good. In contrast to the two equilibria presented above, bargaining solutions are specifically aimed at interactions with a *continuous strategy space*. Such interactions allow agents to perform actions taken from a continuous space $\mathbb{A}^i$, instead of an individual finite set of actions A^i (as in interactions with a discrete strategy set). A prime example is bargaining about a divisable good, i.e., a task that is prominently present in social dilemmas (Endriss, 2008).

Nash (1950b) suggested a bargaining solution that has been demonstrated to be equivalent to Zeuthen's solution, published 20 years earlier (Zeuthen, 1930; Harsanyi, 1956). He specifically considered one specific game modeling bargaining interactions, i.e., the Nash Bargaining Game. In this game, agents demand a portion of some good (usually some amount of money). The total good available is denoted by R. The game is traditionally played between two agents, but it may be extended to more agents. Each agent's action $a^i \in [0, R]$ corresponds to a demanded portion of the good. If the agents' proposals sum to no more than the total good, i.e., $\sum_i a^i \leq R$, then all agents obtain their demand, i.e., $\forall i : r^i = a^i$. Otherwise, we denote the reward distribution with r_0, since nobody receives anything, i.e., $\forall i : r_0^i = 0$. In both cases, the reward distribution leads to some utility experienced by the agents, i.e., $u^i(r)$ and $u^i(r_0)$, respectively.

Nash proposed that a rational solution (i.e., the Nash bargaining solution) to this game should satisfy four axioms being: (1) invariance to affine transformations (or invariance to equivalent utility representations), (2) Pareto optimality, (3) independence of irrelevant alternatives, and (4) symmetry. Here, (1) implies that the solution should remain the same if both agents apply the same transformation on their utility; (2) has been explained above; (3) implies that the solution should remain the same if only a subset (which contains this solution) of the possible outcomes is considered; and (4) implies that players with the same utility functions should receive the same utilities.

Under these conditions, rational players will seek to maximize $\prod_i \left| u^i(r) - u^i(r_0) \right|$. With two equal players and equal utility functions, the only rational solution to this game is therefore a 50-50 split.

The generalized Nash bargaining solution

The generalized Nash bargaining solution introduces the concept of bargaining power (Binmore, 1991, 1998), which relates to the relative abilities of parties in a situation to exert influence over each other. For instance, a large multi-national oil company will have more influence on the Nigerian village near which it wishes to drill for oil, than the villagers of said village will have on the oil company. In general, if one of the agents has a higher bargaining power, he may obtain a higher utility than the other(s). Under such circumstances, the symmetry axiom needs to be dropped. Rational agents will now seek to maximize

$$\prod_i \left| u^i(r) - u^i(r_0) \right|^{\alpha^i}$$

with α^i being the bargaining powers. As in the symmetrical case, this leads to a single rational solution.

2.1.4 Limitations of game theory

After introducing the concept of Nash equilibria, Nash participated in a study with human participants to assess their behavior in various games (Kalisch et al., 1952). The study was considered a failure, because the participants failed to find the Nash equilibria. In hindsight, however, it is one of the first studies showing that game theory often does not predict human strategies correctly.

A few years later, Simon coined the term *bounded rationality* (Simon, 1957; Simon and Newell, 1972) to describe the phenomenon that most people are only partly rational, and are in fact emotional or even irrational in the remaining part of their actions. Bounded rationality suggests that people use heuristics to make decisions, rather than a strict, purely rational rule of optimization. They do this because of the complexity of the situations at hand, and the inability to process and compute all alternatives. Simon describes a number of dimensions along which 'classical' models of rationality can be made more realistic, while remaining within the vein of rigorous formalization. These include placing limitations on the utility functions used, and introducing costs for gathering and processing information.

The field of *evolutionary game theory* was proposed by Maynard-Smith and Price (1973) as a new research area, in order to improve the alignment between game theory on the one hand and behavior as observed in nature (more specifically, biology) on the other hand. In biological interactions, it becomes impossible to assess what choices would be most rational. Instead, individuals choose strategies that increase their species' chance of reproduction (and survival). Evolutionary game theory therefore introduces a new equilibrium concept refining the Nash equilibrium, aimed at explaining the evolution of behavior in animals, namely the evolutionarily stable strategy (ESS) (Maynard-Smith and Price, 1973; Maynard-Smith, 1982). This concept accounts for the stability of strategies in a population of agents. Informally speaking, an ESS is a Nash equilibrium which is 'evolutionarily' stable. This means that once it is fixed in a population, natural selection alone is sufficient to prevent alternative (mutant) strategies from successfully invading.

Over the years, applications of evolutionary game theory have broadened to areas other than biology, as for instance social sciences, in which attempts are made to explain cultural evolution and individual learning using the concept of evolutionary stability. The application of evolutionary game theory to multi-agent systems is a recent, interesting development (Tuyls and Parsons, 2007).

2.2 Multi-agent reinforcement learning

As we discussed in §2.1.2 the notion of a utility function u^i is derived from game theory. It represents the preference relation each agent i has on the set of action profiles $\mathbb{A}$ at a certain time t. The goal of an agent participating in a game-theoretic interaction is to optimize his utility u^i.

The value of u^i often does not depend only on the reward r^i of agent i, but also on other factors, such as the rewards obtained by the other agents. What exactly is meant by 'optimize' may therefore vary. As we discussed, researchers have argued for years that an individually rational solution such as the Nash equilibrium is also the optimal one. This view has caused many problems; we mention two of them, i.e., (1) inferior performance in comparison to human strategies, and (2) a lack of alignment with human expectations.

In this thesis, we basically 'shift' the problem from the issue of defining optimality and finding ways to obtain it, to the issue of developing utility functions that may lead to desired (or in some sense optimal) human-like behavior when utility is 'optimized' individually by agents. Thus, using the new utility functions developed in our research, agents may be equiped with straightforward existing learning algorithms, as long as these are suitable for the types of interaction that we are investigating. For instance, our games mostly require agents to perform a single action, taken from a continuous strategy space (see §2.3).

In this section, we provide a brief background on multi-agent learning techniques that can be applied in the game-theoretic interactions that we examine. More precisely, we first briefly discuss reinforcement learning. Next, we discuss learning automata and continuous action learning automata. Both are (rather specific) techniques within the broad field of reinforcement learning. For a much broader overview, we refer to 't Hoen et al. (2006).

We note that we follow the notational conventions introduced in §2.1 throughout this section as well. Thus instead of using concepts such as the 'reward r' (from reinforcement learning) or 'feedback β' (from learning-automaton research), we use the agents' utility value as a basis for learning. Moreover, we recall that the shorthand notation u^i for agent i's utility value at time t actually implies $U^i(R(a))$.

2.2.1 Reinforcement learning

Reinforcement learning (Sutton and Barto, 1998) is a machine-learning method that is inspired by psychology. It allows agents to learn to optimize their actions with respect to a cer-

tain (positive or negative, and potentially long-term) utility. The method has been applied successfully to various problems, including robot control, elevator scheduling, telecommunications, backgammon, and chess (Sutton and Barto, 1998).

Most reinforcement learning algorithms are applicable to problems with multiple states and/or multiple decision moments (usually called time steps). More precisely, in time step t, agent i is assumed to be in a state $s^i(t)$ from a discrete set of states S. In state $s^i(t)$, agent i may choose to execute an action $a^i(t)$ from its (discrete) set of actions A^i (note the addition of the argument (t) to the action and other familiar concepts). This set of actions may be dependent on the current state. Each action $a^i(t)$, as performed in state $s^i(t)$, leads to the agent being in a new state $s^i(t+1)$, in which agent i potentially receives a non-zero utility $u^i(t+1)$. Commonly, reinforcement learning assumes that multiple actions (and state transitions) are needed to arrive at a state in which an agent receives a non-zero utility. A parameter $0 \leq \gamma \leq 1$, called the discount factor, is introduced to model the fact that non-zero utility in the distant future may be less important than non-zero utility in the near future. Agents now need to find a policy $\pi : S \rightarrow A$ which maximizes $\sum_t \gamma^t u^i(t)$.

Reinforcement learning algorithms can be classified as either *value-based* or *policy-based* (Sutton and Barto, 1998). In value-based algorithms, utility is associated with states or state-action pairs, by means of *value estimation*. Agents estimate values by maintaining a set of estimates of the expected utility for a single policy π (usually either the current or the optimal policy). In such approaches an agent attempts to estimate either the expected utility starting from state $s^i(t)$ and following π thereafter, or the expected utility when taking action $a^i(t)$ in state $s^i(t)$ and following π thereafter. The most well-known value-based algorithm is Q-Learning (Watkins, 1989).

In policy-based algorithms, agents learn directly in the policy space, meaning that they associate utility exclusively with their actions, instead of with states or state-action pairs. Policy-based algorithms are less common in reinforcement learning than in other methods of machine learning, such as simulated annealing or evolutionary computation. A notable exception is formed by *learning automata*, which are mainly applied in problems with only one or a few states. In our research, we use this policy-based reinforcement-learning method, as (most of) the problems investigated are stateless.

2.2.2 Finite-action learning automata

Originally, learning automata were developed for learning optimal policies in single-state problems with discrete, finite action spaces (Narendra and Thathachar, 1989). Finite-action learning automata (FALA) are assumed to be situated in a stateless environment, which implies that a similar action will yield a similar utility, regardless of any previous actions by the automaton itself or any other actors.

As in game theory, every automaton i keeps track of its current strategy $p^i(t)$, where an element $p^i_k(t)$ denotes the probability that each possible action a^i_k from its finite set of possible actions A^i is chosen. As most learning automata, FALA learn in iterations, each of which

entail that the problem at hand is addressed once. More precisely, in every iteration t, the automaton i chooses a single action $a_c^i = a^i(t) \in A^i$ according to its strategy $p^i(t)$. The joint action $a(t)$ resulting from the chosen actions of all FALA, is observed by the environment. It leads to a normalized utility $u^i(t) \in [0,1]$ for each automaton i, where $u^i(t) = 1$ implies that the best action has been chosen by automaton i. The automaton uses this utility to update its strategy for the next iteration $t+1$, by applying a so-called update scheme.

Many different update schemes exist, with various ways to incorporate a received reward into the current strategy. Commonly, the *reward-inaction scheme* is used to update the current strategy. Using the reward-inaction update scheme, FALA have been shown to converge to an equilibrium point, e.g., a Nash equilibrium (Narendra and Thathachar, 1989). The agents' strategies $p^i(t)$ are updated for the next iteration $t+1$ as a response to the utility $u^i(t)$ that agent i obtained in the current iteration by performing its chosen action $a_c^i = a^i(t)$. The update is calculated using a learning rate λ and the following formula (cf. Narendra and Thathachar, 1989).

$$p_k^i(t+1) = \begin{cases} p_k^i(t) + \lambda u^i(t)\left[1 - p_k^i(t)\right] & \text{for } k = c \\ p_k^i(t) - \lambda u^i(t) p_k^i(t) & \text{otherwise} \end{cases} \tag{2.1}$$

Note that the automata are not informed about the actions or utilities of other automata.

2.2.3 Continuous-action learning automata

Continuous-action learning automata (CALA; see Thathachar and Sastry, 2004) are learning automata developed for problems with a continuous action (or strategy) space $\mathbb{A}$. CALA have a proven convergence to (local) optima, given that the continuous utility function $u^i(t)$ is sufficiently smooth. The advantage of CALA over other reinforcement techniques, is that it is not necessary to discretize continuous action spaces; actions are simply real numbers.

Essentially, each automaton i maintains a Gaussian distribution N^i from which actions are pulled. In every iteration t, the automaton selects *two* actions from this distribution, viz. (1) an action $\mu^i(t)$, corresponding to the mean $\mu^i(t)$ of its Gaussian distribution in the current iteration t, and (2) an action $x^i(t)$, corresponding to a sample taken from this distribution in the current iteration. The environment evaluates the two resulting joint actions, and gives each automaton i feedback in the form of two utilities, which we denote by $u_\mu^i(t)$ and $u_x^i(t)$, respectively. In turn, this feedback is used to update the $\mu^i(t)$ and $\sigma^i(t)$ of the probability distribution. More precisely, the update formula for CALA can be written as follows (cf. Thathachar and Sastry, 2004).

$$\begin{aligned} \mu^i(t+1) = &\quad \mu^i(t) + \lambda \frac{u_x^i(t) - u_\mu^i(t)}{\Phi\left(\sigma^i(t)\right)} \cdot \frac{x^i(t) - \mu^i(t)}{\Phi\left(\sigma^i(t)\right)} \\ \sigma^i(t+1) = &\quad \sigma^i(t) + \lambda \frac{u_x^i(t) - u_\mu^i(t)}{\Phi\left(\sigma^i(t)\right)} \left[\left(\frac{x^i(t) - \mu^i(t)}{\Phi\left(\sigma^i(t)\right)} \right)^2 - 1 \right] - \lambda K \left(\sigma^i(t) - \sigma_L \right) \end{aligned} \tag{2.2}$$

In this equation, λ represents the learning rate; K represents a large constant driving down σ. The variance σ is kept above a threshold σ_L to keep calculations tractable even in case of (near-)convergence. This is implemented using the function:

$$\Phi(\sigma) = \max(\sigma, \sigma_L). \tag{2.3}$$

The intuition behind the update formula is quite straightforward. First, if the signs of $u_x^i(t) - u_\mu^i(t)$ and $x^i(t) - \mu^i(t)$ match, then $\mu^i(t)$ is increased, otherwise it is decreased. This makes sense, given a sufficiently smooth feedback function: for instance, if $x^i(t) > \mu^i(t)$ but $u_x^i(t) < u_\mu^i(t)$, we may expect that the optimum is located below $\mu^i(t)$. Second, the variance is adapted depending on how far $x^i(t)$ is from $\mu^i(t)$. The term

$$\left(\frac{x^i(t)-\mu^i(t)}{\Phi(\sigma^i)}\right)^2 - 1$$

becomes positive iff $x^i(t)$ has a distance of more than a standard deviation from $\mu^i(t)$. In this case, if $x^i(t)$ is a better action than $\mu^i(t)$, $\sigma^i(t)$ is increased to make the automaton more explorative. Otherwise, $\sigma^i(t)$ is decreased to make the automaton less explorative and to decrease the probability that the automaton will select $x^i(t)$ again. If $x^i(t)$ has a distance of less than a standard deviation from $\mu^i(t)$, this behavior is reversed: a 'bad' action $x^i(t)$ close to $\mu^i(t)$ indicates that the automaton might need to explore more, whereas a 'good' action $x^i(t)$ close to $\mu^i(t)$ indicates that the optimum might be near.

Using this update function, CALA rather quickly converge to a (local) optimum (Thathachar and Sastry, 2004). Once again, we note that the automata are not informed about each others' actions and utilities.

2.3 The social dilemmas under study

Social dilemmas (Messick and Brewer, 1983) are problems in which there is a conflict between personal benefit and group benefit. Although such problems are common in society, they are still difficult to solve. Models based on assumptions considering individual rationality have shown limited value when they are applied to social dilemmas, as they predict that players will focus purely on their personal benefit (Gintis, 2001). This prediction does not correspond to actual human behavior, since the human tendency to consider group benefit (e.g., fairness) in addition to their own personal benefit allows humans to reach better solutions in social dilemmas.

Researchers have identified two mechanisms that may prevent agents from pursuing only their private benefit in social dilemmas, i.e., *altruistic punishment*, and *withholding action* (Rockenbach and Milinski, 2006). These mechanisms differ in their means for 'disciplining' those who refuse to consider group benefit. Altruistic punishment mechanisms require agents to invest a small amount of reward (e.g., money) in order to reduce other agents'

Table 2.2 The social dilemmas under study in Chapter 4, 5 and 6

Social dilemma	Game	Chapter 4	Chapter 5	Chapter 6
Agreement dilemma	Ultimatum Game (UG)	x	x	x
	Nash Bargaining Game (NBG)	x	x	
Tragedy of the commons	Public Goods Game (PGG)	x		x

amounts by a bigger amount. Withholding action implies refusing to interact with other agents; such a refusal may be motivated by information on the reputation of these others (Nowak et al., 2000; Milinski et al., 2002). In the remainder of the thesis, the two mechanisms will be extensively addressed.

As we already discussed in §1.1, we distinguish two types of social dilemmas, viz. (1) the agreement dilemma, in which a common resource needs to be distributed over a group of agents, with the threat of individual agents being able to reject a proposed distribution if they do not agree with it; and (2) the tragedy of the commons, in which agents need to deal with a common resource, as for instance when they can invest a certain amount in order to obtain a larger returned reward (given that everyone invests). In both cases, *defection* is the individually rational solution. In the agreement dilemma, an individually rational agent will keep as much as possible to himself, knowing that the others can then choose between obtaining a very small reward or nothing at all. In the tragedy of the commons, the individually rational solution is to defect by refusing to invest: if a single agent in the group follows this strategy, he can keep his original investment and receives a share in the investments of others. In contrast, cooperation is the optimal solution in both cases. In the agreement dilemma, we see that low offers are considered 'rude' and will be rejected if this is possible, leading to zero reward for everyone. In the tragedy of the commons, there is a large gain for everyone if everyone cooperates.

Thus, since social dilemmas require a balance between individual rationality and awareness of social context, they are a frequent subject for research concerning the emergence of fairness, cooperation, and agreement. Throughout the thesis, we investigate three games modeling social dilemmas. The agreement dilemma is represented by the Ultimatum Game and the Nash Bargaining Game. The tragedy of the commons is represented by the Public Goods Game. We will explain these games below. We introduce abbreviations for the three games in this section, which will be used in the remainder of the thesis. In Table 2.2, we schematically outline which games will be studied in each of the Chapters 4, 5, and 6.

2.3.1 The agreement dilemma

As mentioned above, we discuss two games representing the agreement dilemma, viz. (1) the Ultimatum Game and (2) the Nash Bargaining Game.

The **Ultimatum Game** (UG) (Gueth et al., 1982) is a bargaining game, played by two agents. The first agent proposes how to divide a (rather small, e.g., $10) reward R between him and the second agent. If the second agent accepts the proposed division, the first obtains his demanded reward and the second obtains the rest. However, if the second agent rejects the proposal, neither obtains anything. The game is played only once, and it is assumed that the agents have not previously communicated, i.e., they did not have the opportunity to negotiate with or learn from each other.

The individually rational solution (i.e., the Nash equilibrium) to the UG is for the first agent to leave the smallest positive reward to the other agent. After all, the other agent can then choose between receiving this reward by agreeing, or receiving nothing by rejecting. Clearly, a small positive reward is rationally preferable over no reward at all. However, research with human subjects indicates that humans usually do not choose the individually rational solution. Hardly any first agent proposes offers that lead to large differences in reward between the agents, and hardly any second agent accepts such proposals. Bearden (2001) and Oosterbeek et al. (2004) analyze many available experiments with humans. They find that the average proposal in the two-agent UG is about 40%, with 16% of the proposals being rejected by the other agent. We replicated this finding in our own experiments (see §5.2). Cross-cultural studies performed in 'primitive' hunter-gatherer cultures have shown the same result, although the average proposal differs by culture (Henrich et al., 2004).

Usually, the UG is played with only two agents. As we are interested in a multi-agent perspective, we also consider an UG with more than two agents. There are various extensions of the UG to more agents, e.g., introducing proposer competition (Prasnikar and Roth, 1992) or responder competition (Fischbacher et al., 2003; Halko and Seppala, 2006). With proposer competition, one responder can accept one of the potentially multiple offers provided to him. This leads to true competition between the various proposers; average offers can be expected to increase. In case of responder competition, one proposer makes an offer to more than one responder. Experiments show that the addition of even one competing responder reduces the average offer by approximately a factor two. In our research, we are not aiming at understanding the competition between agents. Instead, we wish for a large group of agents to reach an agreement on how to share a certain reward. We therefore propose a different extension. More precisely, we define a game in which all but one agent (each in turn) take a portion r^i of the reward R (or as much as is still available if this is less than their intended reward). The last agent receives what is left.

As already has been explained in §2.1.3, the **Nash Bargaining Game** (NBG) was proposed by Nash (1950b) as an illustration of the Nash bargaining solution.[2] The game is traditionally played by two agents, but can easily be extended to more agents.

In this game, all agents simultaneously determine how much reward r^i they will claim from a common reward R. Due to the simultaneity, each agent does not have knowledge of the

[2] Nash did not name the game and the solution after himself; as with the Nash equilibrium and the game of Nash, it received its name from admiring colleagues at Princeton (Nash, 2001).

claims of the other agents. If $\sum_i r^i > R$, everyone receives 0. Otherwise, everyone receives an amount equal to their claim. Note that rewards may not sum up to R, i.e., a Pareto-optimal solution is not guaranteed. The game has many Nash equilibria, including one where all agents request the whole R. The common human solution to this game is an even split (Nydegger and Owen, 1974; Roth and Malouf, 1979; Yaari and Bar-Hillel, 1981).

The generalized Nash Bargaining Game (gNBG) introduces the concept of bargaining powers, implying that some agents may have more influence on the outcome than others (Binmore, 1991). As has been outlined in §2.1.3, rational agents will now seek to maximize

$$\prod_i \left| u^i(r) - u^i(r_0) \right|^{\alpha^i}$$

with α^i being the bargaining powers. As in the symmetrical case, this leads to a single rational solution.

2.3.2 The tragedy of the commons

The **Public Goods Game** (PGG) is one of the classical examples of the so-called tragedy of the commons (Hardin, 1968), in which Adam Smith's famous "invisible hand" (Vaughn, 1987) fails to work.[3] The tragedy of the commons basically entails that a common resource, i.e., a resource shared by many agents, may become exhausted if all agents benefit from taking a part of this resource for themselves. Other examples than the PGG include the near-extinction of certain species of consumable fish, overgrazing of a shared field, or (a more computational example) balancing the load on powerful servers by also delegating tasks to less powerful ones (Verbeeck et al., 2002).

The game is typically played by 3 to 10 agents. Every agent receives an amount of money, (part of) which can be invested in a common pool. All agents simultaneously choose how much to invest. Then, everyone receives the money he kept to himself, plus the money in the common pool, multiplied by a certain factor (usually 3) and divided equally among the agents. To gain the most profit, everyone should cooperate by contributing their entire private amount. However, every agent can gain from solely not contributing. Thus, the Nash equilibrium is for every agent to defect by not contributing at all, leading to a monetary reward that is much lower than the optimal reward. Sigmund (2007) compares this game to hunting a mammoth in prehistoric times: every hunter runs the risk of being wounded by the mammoth, but if they all work together, they can easily catch the animal.

Formally, a PGG is often considered as having only two strategies available to the agents, i.e., (1) to defect by refusing to contribute, and (2) to cooperate by contributing one's entire private amount. The game is characterized by three parameters, i.e., n, the number of agents (with n_d the number of defectors and n_c the number of cooperators); x, the initial amount

[3] Smith claimed that, in a free market, an individual pursuing his personal benefit tends to also promote the good of his community as a whole through a principle that he called "the invisible hand". He argued that each individual maximising revenue for himself maximises the total revenue of society as a whole, as this is identical to the total of individual revenue.

of money that every agent possesses (which is usually not considered); and c, the amount that an agent can contribute per game. For a 'successful' PGG, $1 < r < n$ must hold, since for $r < 1$, defection is the only sensible strategy (both individually and on a group level), and for $r \geq n$, the same holds for cooperation.

As a result of the rules of the PGG, we see that defectors obtain $\frac{1}{n} rc \cdot n_c$ and cooperators obtain $\frac{1}{n} rc \cdot n_c - c$. Defectors thus obtain a higher reward, so they have a higher probability of being imitated. In the end, the whole population defects, even though in this game (and other examples of the tragedy of the commons), obtaining a fair, cooperative solution is in the material advantage of everyone.

The effect of introducing costly, altruistic punishment in the PGG has been extensively examined (e.g., Yamagishi, 1986; Fehr and Gaechter, 2002). More precisely, agents can give a sum of money to the experimenter to decrease the reward of a defector. This mechanism introduces two new parameters, i.e., e_p, the effect of punishment received by a defector for each other agent that punishes; and c_p, the cost of punishing one other agent. Commonly, these effects and costs are assumed to be constant, independent of the target and originating agent. Obviously, $c_p < e_p$ must hold to make punishment at least slightly attractive. Typically, $c_p = 1$ and $e_p \in [3, 5]$ are chosen. In the presence of punishment, defectors lose $e_p n_p$ (with n_p the number of punishers, i.e., cooperators that also punish) in comparison to the non-punishing game and punishers lose $c_p n_d$ on their reward. As will be explained in §4.2, punishment is not a dominant strategy, since players may resort to second-order free-riding. Nonetheless, human players consistently apply punishment (successfully) if they are allowed. Many possible explanations exist of why this happens, but none of the existing explanations have been proven to lead to stable cooperative solutions. In §4.2, we present our own explanation, which does lead to stable cooperation.

In addition to using altruistic punishment, reputation may also allow agents to reach satisfactory solutions. More precisely, agents may use the reputation of other agents to decide whether they wish to participate in the game (see §5.1). In case the agent refuses, he obtains a relatively low reward s. In this case, $0 < s << (r-1)c$ must hold, because otherwise, it pays off not to participate. As with punishment, research has indicated that optional participation does not lead to any strategy becoming dominant. Defection, cooperation, and refusal to participate oscillate endlessly (Hauert et al., 2002).

2.4 Chapter summary

In this chapter, we discussed the background of the thesis. First, we described game theory, an essential field for our research, which considers interactions between potentially many agents. We also briefly discussed that game theory often predicts rational behavior that is not observed in humans.

Second, we discussed multi-agent reinforcement learning, focusing on learning automata in single-stage interactions with a discrete strategy set or a continuous strategy space.

Third, we discussed our problem domain, i.e., social dilemmas. We outlined two different types of social dilemma, i.e., (1) the agreement dilemma, represented in the thesis by the Ultimatum Game and the Nash Bargaining Game, and (2) the tragedy of the commons, represented here by the Public Goods Game. We described that humans employ two distinct mechanisms to discipline each other towards satisfactory solutions in social dilemmas, i.e., altruistic punishment and withholding action. While altruistic punishment disciplines by subtracting actual rewards (at a smaller cost to the punisher), withholding action disciplines by disallowing certain agents to enter an interaction with an expected positive reward. With such mechanisms in place, agents must make sure that they are not disciplined, even if they are caring only for their own benefit.

3 The foundations of computational fairness

In this chapter, we concentrate on answering RQ2: "What are the foundations of human-inspired computational fairness?" We present two foundations here, i.e., (1) a set of *requirements* that need to be met by human-inspired computational fairness models, and (2) a *template model*. The template model assumes that agents are equipped with a utility function, and specifies how this function may allow agents to address three questions, i.e., (R3-Q1) to what extent an interaction is fair, (R3-Q2) whether one or more of their peers need(s) to be punished, and (R3-Q3) whether it is desirable to participate in a certain interaction.

In the core chapters of the thesis (i.e., Chapters 4, 5 and 6), the template will be implemented by means of concrete human-inspired fairness principles, in relation to the three questions i.e., for R3-Q1, we provide utility functions modelling human decision-making; for R3-Q2, we provide punishment mechanisms applied by humans; and for R3-Q3, we provide ways of influencing whether or not interaction with certain peers will take place. This leads to computational models of human-inspired fairness.

The chapter is structured as follows. First, in §3.1, we outline the three requirements that need to be met by computational models. Second, in §3.2, we present our template model, which is proposed with these requirements in mind. Third, in §3.3, we look at related work, i.e., existing work that proposes computational models of fairness. We show that these models do not meet our requirements. Fourth, we conclude the chapter by a summary (§3.4).

The work reported on in this chapter has partially been published in:

S. de Jong, K. Tuyls, and K. Verbeeck. Fairness in multi-agent systems. *Knowledge Engineering Review*, Vol. 23(2):153-180, 2008.

S. de Jong, S. Uyttendaele, and K. Tuyls. Learning to reach agreement in continuous strategy spaces. *Journal of Artificial Intelligence Research*, Vol. 33:551-574, 2008.

The author acknowledges Ulle Endriss for permission to use his tutorial given at AAMAS 2008 as a basis for the discussion of welfare economics in §3.3.1 (Endriss, 2008).

3.1 Requirements for human-inspired computational fairness

In this section, we outline the three requirements that need to be met by computational models of human-inspired fairness.

R1 The models should be rooted in a game-theoretic background.

Game theory provides us with well-established, well-defined manners to describe interactions between multiple parties. As such, it is a good basis for (learning in) multi-agent systems (Shoham et al., 2007). As we discussed in §2.1.4, game theory has nothing in particular to say about fairness, let alone human(-inspired) fairness. Our research therefore aims at building human-inspired fairness models upon game-theoretic principles.

R2 The models should be computationally applicable, i.e., in a setting of multi-agent systems addressing social dilemmas.

This requirement may sound obvious; if our models would not be computationally applicable in an actual multi-agent system that is learning to address social dilemmas, they would not be very useful for our purposes. However, many game-theoretic solution concepts (e.g., the Nash equilibrium), as well as many existing models of human fairness (e.g., the Homo Egualis utility function, see §4.1) do not meet this requirement directly, for instance due to tractability issues or insufficiently smooth utility functions.

R3 The models should enable adaptive agents to mimic human fairness mechanisms.

This requirement ensures that our models are indeed human-inspired, as this is one of the explicit goals of our work. As we discussed earlier, humans employ two distinct mechanisms to maintain fairness, viz. *altruistic punishment* and *withholding action* (the latter for instance as a response to reputation information). In order to transfer these mechanisms to our computational models, these models should allow agents to answer three questions.

R3-Q1 "Do I consider the given reward distribution to be fair to me?"

R3-Q2 "Am I willing to pay in order to punish a certain other agent?"

R3-Q3 "Do I want to interact with a certain other agent?"

These three questions directly map to the three elements of our template model. According to the template, the agents will use a *fairness utility function*, which meets the first two requirements, and allows agents to measure the fairness of a reward distribution (R3-Q1). Agents may increase their utility by punishing other agents, for instance, because this punishment effectively reduces differences in obtained reward, as in Chapter 4 (R3-Q2). Moreover, if agents predict that a certain interaction will make them (feel that they are) unfairly treated, they may refuse to participate in this interaction, i.e., to withhold action (R3-Q3).

3.2 A template model for human-inspired computational fairness

We address the requirements of the previous section by means of a *template model*, upon which each computational model will be based. In Figure 3.1, we provide a graphical depic-

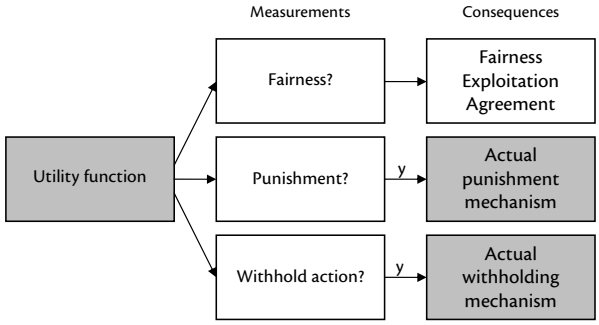


Figure 3.1 The template model for computational human-inspired fairness

tion of our template model for computational human-inspired fairness. Here, white boxes are part of the template, and gray boxes represent elements that need to be added.

From left to right, we first encounter a *utility function*. Progressing one step to the right, we see that, given their utility function, agents may perform three *measurements*, i.e., from top to bottom in the figure, first, they may measure the fairness of an interaction, second, they may determine whether punishment would be desirable, and third, they may determine whether withholding action would be desirable. These three measurements, which are part of the template, are further discussed in the following subsections.

In the right-most column of the figure, we see the *consequences* of the three measurements. From top to bottom, once again, we observe three consequences. First, after having measured the fairness of an interaction, agents 'know' whether this interaction was fair to them (i.e., they may agree with the interaction), or whether they have been exploited. Second, if according to their utility function, they would benefit from punishing, we may provide agents with an actual punishment mechanism, which specifies how punishment actually takes place. Third and last, a similar mechanism may be introduced for withholding action.

Thus, as can be seen from the figure, computational models that are based on the template model need to provide three elements. In the core chapters of the thesis (i.e., Chapter 4, 5 and 6), we typically take the approach of starting from existing descriptive work, and embedding the concepts of this work in our template, i.e., human-inspired utility functions and mechanisms for punishment and withholding action, resulting in computational models of human-inspired fairness. In Table 3.1, we provide a brief summary of the elements used to extend our template model in the following chapters. Explanation will follow in the respective chapters. Below, we discuss the 'white boxes' of the figure.

3.2.1 Determining the fairness of an interaction

As in game theory (see §2.1), reinforcement learning (see §2.2), and welfare economics (see §3.3), our models will be centered around the concept of a utility function, i.e., we measure

Table 3.1 The instantiations of the template model in Chapter 4, 5 and 6

Chapter	Utility function	Punishment	Withholding action
4	Homo Egualis	All agents	No
5	Priority awareness	Specific agents	No
6	Threshold	Specific agents	Rewiring in interaction network

whether agents feel treated fairly by looking at their utility function $u^i = U^i(R(a))$.[1] Thus, their utility function helps agents to answer the question: "Do I consider the given reward distribution to be fair to me?" (i.e., R3-Q1).

Clearly, the actual utility function to be used has to be selected with care, as our three requirements need to be met. We shall refer to suitable utility functions as *fairness utility functions*. A great deal of the work in the thesis is concerned with actually finding and applying suitable fairness utility functions. As an initial example, in §3.2.4, we will present some elementary fairness utility functions that may be used in social dilemmas.

Given a certain utility function U, the baseline fairness utility for agent i may be denoted by $u^i_0 = U^i(r_0)$; it results from the agents not performing any action, and thus yielding a reward distribution r_0. We follow the general principles also applied to, e.g., reward functions, and define that a reward distribution r leading to $U^i(r) < U^i(r_0)$ is clearly not fair, i.e., agents would have preferred not taking any action (yielding r_0) over taking actions that yield r. Similarly, a reward distribution r leading to $U^i(r) = U^i(r_0)$ leads to indifference, and a reward distribution r leading to $U^i(r) > U^i(r_0)$ is fair to a certain degree. Depending on the domain, utility may be only ordinal (i.e., anything above zero is sufficient), or cardinal (i.e., we cannot only express that a reward distribution is fair, but also how fair it is; (Fishburn, 1970)). In the remainder of the thesis, we use cardinal utility.

Using the fairness utility function U^i, we may define the terms 'exploitation', 'agreement' and 'fairness' as follows.

Definition 3.1 Agent i is exploited given a reward distribution r iff $U^i(r) < U^i(r_0)$.

Definition 3.2 Reward distribution r leads to agreement iff $\forall i : U^i(r) \geq U^i(r_0)$.

Definition 3.3 Reward distribution r is fair iff it leads to agreement.

The last definition is rather strict; it would not allow a minority of agents to be exploited at the benefit of a vast majority, and may therefore lead to fairness being impossible to achieve – for instance, one agent may have a truly incompatible view on what is fair and what is not (i.e., his fairness utility function may not be aligned with that of a vast majority). Such outcomes are not necessarily bad; we know from research with humans that they may also have different opinions on what constitutes a 'fair deal' in, for instance, the UG (Fehr and

[1] For convenience of notation, we omit the argument t from a, r and u, which we introduced when discussing reinforcement learning in §2.2.

Schmidt, 1999; Oosterbeek et al., 2004). Typically, around 10% of human players actually play rationally in this game, and therefore do not agree with the remaining 90%. It is therefore safer to look at a more tolerant version of the definition of fairness:

Definition 3.4 An interaction between n agents is fair with an error margin of ϵ iff no more than ϵn agents are exploited, i.e., $|\{i : U^i(r) < U^i(r_0)\}| \leq \epsilon n$.

For instance, in the UG, the average human population would reach fairness with an error margin of approximately $\epsilon = 0.1$. In contrast, a population consisting of only purely rational agents would all agree on offering and accepting the least possible amount and would therefore reach fairness without any error.

3.2.2 Performing altruistic punishment

In addition to allowing agents to measure the fairness of a certain interaction, fairness utility functions may be used to allow agents to answer the question: "Am I willing to pay in order to punish a certain other agent?" (i.e., R3-Q2).

Altruistic punishment is a concept central to human-fairness research. Formally, this concept entails that an agent i may want to invest a portion of his reward r^i to subtract a (larger) portion of the reward r^j of a target agent j. Since punishment is costly, it will most probably reduce agent i's reward if he decides to punish. Agent i therefore cannot base the decision whether or not to punish only on his reward r^i. We denote the cost of performing altruistic punishment with c_p^i, and the effect with e_p^i. Usually, these costs and effects are not considered to depend on the agent performing or receiving punishment. The index i is therefore omitted in the remainder of this text.

Punishment mechanisms may be classified alongside two dimensions, i.e., the *impact* of punishment, and the *scope* of punishment.

First, in some interactions, such as the UG, the only way of performing punishment is by refusing an interaction. In this case, agent i must therefore pay his entire reward r^i to reduce agent j's reward r^j to 0, i.e., $c_p = r^i$ and $e_p = r^j$. In other interactions, such as the PGG, more sophisticated (as well as less drastic) punishment mechanisms may exist, where agent i pays only a (constant) part of his reward in order to punish agent j.

Second, the scope of punishment may be different: in some interactions (such as, once again, the UG), punishment by agent i applies to all agents (possibly even including agent i). We consider the original reward distribution r, as well as the reward distribution resulting from punishment, r', with $r'^i = r^i - c_p$ and $r'^j = r^j - e_p$ (possibly $\forall j \neq i$). If $U^i(r') > U^i(r)$, then agent i should punish. In other interactions (such as the PGG), punishment by agent i is directed at (a) specific other agent(s). Here, we see that agent i needs to punish j iff $U^i(r^i - c_p, r^j - e_p) > U^i(r^i, r^j)$.

3.2.3 Withholding action

Fairness utility functions may also be used to allow agents to answer the question: "Do I want to interact with a certain other agent?" (i.e., R3-Q3).

In a similar way as described above, the fairness utility function may be used to allow agents to refuse a proposed interaction (before it actually takes place) if they have information on, for instance, previous behavior or reputation of the other agent(s). If agent i predicts a reward distribution $\tilde{r}$, he may calculate $U^i(\tilde{r})$. If this yields a lower fairness utility than his baseline fairness utility $U^i(r_0)$, agent i should not participate.

3.2.4 Examples

Below, we illustrate the concept of fairness utility functions by deriving such functions for the UG and the PGG. We are not necessarily presenting functions here that have a proven correspondence with human behavior. However, extending our template by means of these fairness utility functions already leads to simple computational models of fairness.

In the UG, the reward distribution r is identical to the joint action a, given that all agents choose not to reject (i.e., not to perform altruistic punishment). Otherwise, $r = r_0$. If agent i rejects, then $c_p = r^i$ and $\forall j \neq i : e_p = r^j$.

Assume for instance that agent i considers any reward distribution unsatisfactory (i.e., unfair) in which there is an agent that obtains more than twice the amount that i obtains. In this case, we may define the fairness utility function:

$$U^i(r) = r^i - \frac{1}{2}\max_{j \neq i} r^j.$$

We note that in this case, punishing yields $U^i(r') = U^i(r_0) = 0$, since no agent obtained anything, and additionally, no agent obtained more than twice the amount another agent obtained. Thus, if $U^i(r) < 0$, agent i should punish.

In the PGG, a continuous strategy space leads to agents viewing each other as *relative* cooperators and defectors. Everyone likes to play with relative cooperators, but nobody likes relative defectors. We can use a fairness utility function to model this, as well as to enable agents to decide whether or not to perform altruistic punishment. Assume for instance that any reward r^j that is 20% above agent i's reward r^i, is viewed as unacceptable, unfair defection. Then, the fairness utility function U^i will be defined as:

$$U^i\left(r^i, r^j\right) = r^i - \frac{1}{1.2} r^j.$$

In case of punishment, we obtain:

$$U^i\left(r^i - c_p, r^j - e_p\right) = \left(r^i - c_p\right) - \frac{1}{1.2}\left(r^j - e_p\right).$$

Clearly, if $U^i\left(r^i, r^j\right) < U^i\left(r^i - c_p, r^j - e_p\right)$, then agent i should punish agent j.

In Chapter 4, we will discuss a more refined way of incorporating phenomena such as the human tendency to dislike large differences in reward into the fairness utility function. In later chapters, we also provide various ways for agents to perform altruistic punishment or to withhold action.

3.3 Related work: existing computational models of fairness

Existing models of fairness do not adhere to the three requirements presented in §3.1. Most descriptive models do mimic human behavior (R3), but are lacking in game-theoretic background (R1) or may not be (intended to be) explicitly computable (R2). We will discuss this further in the following chapters (see §4.1, §5.1).

In contrast, computational models as presented by welfare economics are usually well-rooted in a game-theoretic background (R1), but even though they are explicitly computational (R2), they may not be directly suited to social dilemmas, because they do not include elements directly inspired by human behavior (R3). We will investigate this further in §3.3.1.

Thus, there is a gap between current descriptive models of fairness on the one hand, and computational models of fairness on the other hand. A possible approach to bridging this gap may be found in research areas such as evolutionary game theory and statistical physics. However, current work in these areas is not actually performed with the explicit goal of bridging the gap. It essentially develops computational implementations of human fairness only as a 'proof of principle', and therefore, certain abstractions may be allowed that we cannot justify in our work. Thus, although such work generally does have a game-theoretic background (R1), and includes human-inspired mechanisms (R3), it may not be fully applicable to the social dilemmas we are interested in (R2). We provide more details in §3.3.2.

3.3.1 Fairness in welfare economics

In welfare economics, the problem of fair division is posed as follows. Given a set of agents, a set of goods (resources), and specific preferences per agent considering the allocation of goods to agents, we can ask (1) what constitutes a good (or fair) allocation?, and (2) how do we find it? We may distinguish fair division problems in which there may be only one good, or several ones (in our work, one, i.e., money), or in which goods are divisable or not (in our work, they are), or in which sharing goods may be possible or not (in our work, they may be – i.e., in the tragedy of the commons, it is the central issue).

In the remainder of this section, we present three distinct topics, i.e., first, we review and compare proposed criteria for "fairness" and the related notion of "efficiency". Second, we

discuss the problem of fair division of a single divisable good, which, in contrast to the good in our research, is assumed to be heterogeneous. Third, we briefly discuss the problem of fair division given a set of indivisable goods (which is a problem less related to our research). After discussing the three topics, we look at the possibilities and limitations of applying concepts from welfare economics to our problems of interest, i.e., social dilemmas.

Fairness and efficiency criteria

In the literature, much attention has been given to defining criteria for deciding what a satisfactory allocation of resources actually entails (Chevaleyre et al., 2006). While different applications may lead to different specific criteria, we may also look at more general criteria, which may be defined in terms of individual agents' preferences. Criteria may then be divided into two categories, i.e., they are either fairness criteria, or efficiency criteria.

As in game theory, welfare economy formalizes agents' preferences by means of an individual utility function U^i and a resulting utility vector $U = (U^1, \ldots, U^n)$, both of which are defined given a certain resource allocation r. We now need to decide on an agreement, i.e., a certain resource allocation r that optimally satisfies the agents' preferences, as captured in the utility vector. Pareto-efficiency (see §2.1.3) with respect to U is often considered a minimum requirement here. However, a Pareto-efficient allocation may still be rather unfair or inefficient (Sen, 1993; Barr, 2004).

Given the utility vector U, we may define a notion of a *social welfare function*, denoted by $S : \mathbb{R}^n \rightarrow \mathbb{R}$, and then aim for an allocation r (and a utility U resulting from it) that maximizes this social welfare function. The social welfare function is generally required to be reflexive, transitive as well as complete. In multi-agent systems, we often encounter the (implicit) assumption that the average utility needs to be maximized in order to obtain maximum efficiency; the social welfare function may then be set to $S(U) = \frac{1}{n}\sum_i U^i$. We discussed already in §1.3 that this assumption is regularly not valid. In accordance, welfare economics states that, while the assumption that only efficiency is important may be reasonable, a more systematic approach to defining social preferences is needed.

One systematic approach is *axiomatic*; we may propose a number of axioms, i.e., properties that we may or may not wish to impose on a social welfare function. We will discuss three axioms here (Moulin, 1988).

1. The first axiom is *zero independence*, which is based on the assumption that agents' utilities may already have been rather different before a certain allocation took place. It is therefore possibly more useful to measure the change of utility every agent experiences as a result of a certain allocation, than to measure the absolute utility. A desirable property of a social welfare ordering may therefore be to be independent from what the individual agents consider 'zero' utility. More formally: given any two allocations r_1 and r_2, as well as two allocations $r'_1 = \{r^1_1 + c, \ldots, r^n_1 + c\}$ and $r'_2 = \{r^1_2 + c, \ldots, r^n_2 + c\}$, and finally the information that $S(U(r_1)) \geq S(U(r_2))$, then the social welfare function S is zero-independent iff $S(U(r'_1)) \geq S(U(r'_2))$ for any c.

2. The second axiom is *scale independence*, which is similar to the first one, except that it enforces the property of being independent from the way agents measure their individual utilities. Formally, once again: given any two allocations r_1 and r_2, as well as two allocations $r'_1 = \{r^1_1 \cdot c, \ldots, r^n_1 \cdot c\}$ and $r'_2 = \{r^1_2 \cdot c, \ldots, r^n_2 \cdot c\}$, and finally the information that $S(U(r_1)) \geq S(U(r_2))$, then the social welfare function S is scale-independent iff $S(U(r'_1)) \geq S(U(r'_2))$ for any c.
3. The third axiom is *independence of the common utility pace*. The underlying idea is that we would like to be able to make social welfare judgements without knowing what kind of tax members of society will have to pay. In essence, we would like to remove the cardinal intensity of $U^i - U^j$, and only preserve the ordinal relation. Formally: given any two allocations r_1 and r_2, as well as two allocations $r'_1 = \{f(r^1_1), \ldots, f(r^n_1)\}$ and $r'_2 = \{f(r^1_2), \ldots, f(r^n_2)\}$, and finally the information that $S(U(r_1)) \geq S(U(r_2))$, then the social welfare function S is independent of the common utility pace iff $S(U(r'_1)) \geq S(U(r'_2))$ for any increasing bijection $f : \mathbb{R} \to \mathbb{R}$.

In addition to axiomatically approaching the social welfare ordering, we may also introduce criteria for an actual allocation to be satisfactory. We mention two criteria here.

1. The first criterium is a rather simple one: *proportionality*. If we define by $U^i_{max} = \max_r U^i(r)$ the utility of agent i for the most attractive allocation possible, then an allocation is proportional iff $U^i(r) \geq \frac{1}{n} U^i_{max}$ for all agents i. Given that each agent may value the resources being allocated differently, a proportional allocation may not be a straightforward equal split.
2. The second criterium is substantially more difficult. It is called *envy-freeness*. Intuitively, an envy-free allocation is one in which no agent would prefer someone else's resources over his own. Clearly, in case there is only one divisable resource (e.g., money), satisfying this criterium would be trivial: we need to give every agent an amount that leads to equal utilities for everyone. In case of multiple (divisable or indivisable) resources, satisfying the criterium is not trivial. An envy-free allocation may not even exist, especially if we also require an allocation that is complete or Pareto-optimal. Therefore, commonly, we need to find an allocation that reduces all agents' envy as much as possible. However, here we run into a recursive problem, as once again, all agents have a certain envy, which raises the question what function, mapping individual agents' envies to a global envy value, is actually applied.

Both in the 'original' problem of finding a suitable social welfare function that maximizes the population's social welfare, as well as in the 'additional' problem of finding a suitable function that minimizes the population's envy, a large number of functions may be used. We mention four of such functions.

1. The first function has already been mentioned, i.e., utilitarian social welfare, where we aim at maximizing total utility, i.e., $S(U) = \frac{1}{n} \sum_i U^i$. Moulin (1988) proved that this is the only social welfare function that satisfies the axiom of zero-independence. It does not satisfy the other two axioms.

2. The second function is known as egalitarian social welfare, where we aim at maximizing the minimal utility, $S(U) = \min_i U^i$. An argument in point for this function is provided by Rawls (1971), i.e., the so-called veil of ignorance. Translated to a multi-agent systems context, the argument may read as follows: "If you were to send a software agent into an artificial society to negotiate on your behalf, what would you consider acceptable principles for that society to operate by?" (Endriss, 2008). The egalitarian social welfare function satisfies the third axiom, i.e., independence of the common utility pace. It does not satisfy the other two axioms.
3. The third function is named the Nash product: $S(U) = \prod_i U^i$. This function, like the utilitarian one, favors increases in overall utility, but also promotes reducing inequalities. As a result, some of the efficiency of the utilitarian function may be lost, but an optimal allocation becomes more fair. This function is the only one that satisfies the second axiom, i.e., scale independence. It does not the satisfy first or third one.
4. The fourth and final function we mention here is called the k-rank dictator function. This function requires the utility vector to be sorted in increasing order, yielding a vector U_{sort}. Now, the function is defined as $S^k(U) = U^k_{sort}$. Varying k allows us to have more (bigger k) or less (smaller k) attention for the best-performing agent. For $k = 1$, this function is equal to the egalitarian social welfare function. For any k, this function satisfies the third axiom, i.e., independence of the common utility pace. It does not satisfy the other two axioms.

Many more social welfare functions, as well as their relevance to multi-agent resource allocation, are addressed by Chevaleyre et al. (2006).

Divisable goods and cake-cutting

Cake-cutting is a well-known mathematical metaphor for the fair division of a single divisable good. Generally, it is assumed that the divisable good is heterogeneous, i.e., some agents prefer a certain part of the good over another, equally large part.[2] Even though the model is simple, there still are many open problems.

More abstractly, the cake may be represented as the unit interval $[0, 1]$. Each agent has a non-negative, additive, and continuous valuation function $v^i(X)$, as defined on intervals $X \in [0, 1]$. If $X = [0, 1]$, then $v^i(X) = 1$ by definition.

The classical approach to cake-cutting with two agents is called Cut-and-Choose. One agent cuts the cake in two pieces (which he considers equally valued), the second agent chooses the piece he prefers. Clearly, the first agent has a strategic disadvantage, as he will obtain a valuation of exactly $\frac{1}{2}$, whereas the second will obtain at least $\frac{1}{2}$. The obtained allocation is proportional and envy-free. Further properties we may be interested in include Pareto-efficiency (which is either trivial or impossible with Cut-and-Choose, the latter for instance

[2] We note that the good our agents bargain about, i.e., money, is not heterogeneous. Our cake-cutting problem is therefore less difficult to solve than the general one.

if one of the agents highly values the middle part of the cake) and equitability, i.e., each agent assigns the same value to their slice (which is already violated by Cut-and-Choose, and for $n > 2$, is in conflict with the more useful property of envy-freeness).

In addition to properties of the resulting division of the cake, we may also be interested in a number of operational properties, i.e., properties of the cake-cutting procedures themselves, such as a minimal number of cuts, whether every agent obtains exactly one contiguous slice, whether an external referee is needed, and whether the procedure is a proper protocol (some procedures may require, e.g., a continuously moving knife, see below). For $n = 2$, Cut-and-Choose is the ideal procedure with respect to each of these properties. It requires only one cut, gives every agent one slice, requires no referee, and is indeed a proper protocol.

For $n > 2$, the cake-cutting problem is much more difficult. A large issue here is that proportionality no longer guarantees envy-freeness. Achieving a proportional division is possible using various elegant procedures (e.g., the Steinhaus procedure for $n = 3$, and the Banach-Knaster Last-Diminisher procedure for arbitrary n; for details, we refer to Steinhaus, 1948). Procedures for envy-freeness are quite complicated already for $n = 3$, as the number of cuts exceeds the minimal number needed (as in the Selfridge-Conway procedure; Brams and Taylor, 1995), or multiple continuously moving knifes are needed (as in the Stromquist procedure; Stromquist, 1980). This implies that agents, and possibly a referee, continuously move their knifes over the cake, until one of the agents feels that a certain piece should be his. For $n = 4$, no known procedure yields continuous pieces, and for $n \geq 5$, no known procedure requires an unbounded number of cuts (Brams and Taylor, 1995).

Indivisable goods

In the previous paragraphs, as well as the remainder of the thesis, we considered the resource that needs to be allocated to be continuous in nature, i.e., it is arbitrarily divisable.[3] Here, we briefly discuss approaches aimed at allocating indivisable goods. Such approaches may be either centralized or distributed.

In a centralized approach, there is an optimization algorithm that computes an allocation meeting the desired requirements concerning fairness and efficiency. This approach gives rise to substantial problems, especially concerning complexity. For instance, checking whether a certain allocation is Pareto-efficient is a Co-NP-complete problem[4], finding an allocation with maximal utilitarian social welfare is an NP-hard one, and checking whether an envy-free allocation exists is an NP-complete one (Chevaleyre et al., 2006). One successful example of a centralized approach is to use the winner-determination problem in combinatorial auctions (Sandholm, 2006), which allows us to approximate empirically an allocation with maximal utilitarian social welfare.

[3] One may argue that money actually is an indivisable good, as there is a smallest *physical* quantity (e.g., €0.01). However, as arbitrarily small fractions of the smallest physical quantity are possible and allowed, money is no less continuous than, e,g, distance.

[4] A Co-NP-complete problem is the complement of an NP-complete problem.

A distributed approach aims at achieving a socially optimal allocation by means of a series of local deals. This approach therefore is closely related to multi-agent negotiation (Endriss et al., 2006). Locally, a rational, reactive agent will only accept deals that improve his individual welfare, i.e., his individual utility. Globally, we may be interested in optimizing, e.g., the average utility of the population of agents as a whole. An important result for our work is that, if we indeed consider average utility to be the quantity of interest, then individual rationality can be shown to be "appropriate". With other social welfare functions, we may require agents to decide based on more than only individual rationality. Most notably, finding an envy-free allocation of indivisable goods by means of rational negotiation is *impossible* (Endriss et al., 2006).

Applying welfare economics to social dilemmas

In general, we may say that welfare economics provide a good starting point for studying and obtaining fairness in multi-agent systems. In order to obtain fair solutions, we define a social welfare function $S : \mathbb{R}^n \to \mathbb{R}$, mapping agents' utilities, as given in a utility vector U, to one single real value $S(U)$, expressing social welfare. We then assume that all agents behave in a manner that maximizes $S(U)$, and obtain a desirable allocation. Thus, an approach based on welfare economics meets our first two requirements (see §3.1). While it clearly does not meet the third requirement (i.e., being explicitly human-inspired), we have not yet demonstrated that this is a problem. We will investigate this here.

In the text above, we discussed one particularly interesting social welfare function (at least for our purposes), i.e., the Nash product, which not only considers the efficiency of a certain resource allocation, but also aims at reducing the inequality (and increasing the fairness) of an allocation. This function may therefore be useful in our problem domain, i.e., social dilemmas. As an example, we apply this function to a two-player two-strategy PGG (without punishment, obviously, as this is a human-inspired mechanism). We assume that agents' utility is calculated in an individually rational manner, e.g., $U^i(r) = r^i$.

Given a PGG with $r = 1.5$ and $c = 10$, the social welfare obtained if both agents defect is simply $S(U) = 0 \cdot 0 = 0$. If one agent cooperates and the other defects, they obtain rewards of -2.5 and 7.5, yielding a social welfare of $S(U) = -2.5 \cdot 7.5 = -18.75$.[5] If both agents cooperate, the social welfare is $S(U) = 5 \cdot 5 = 25$. Clearly, cooperation is best here, as it would also be using egalitarian social welfare. Thus, we may find social welfare functions that allow us to give the highest social welfare value to the most desired solution.

Generally, an approach based on optimizing social welfare does not always work so well, while a human-inspired approach based on punishment and withholding action does. We provide four reasons why.

[5] The complete calculation goes as follows. There is one cooperative and one defective agent, so the common pool contains $1.5 \cdot 10 = 15$. The two agents each obtain 7.5 from this pool. The cooperative one thus gains $7.5 - 10 = -2.5$ and the defective one gains $7.5 - 0 = 7.5$. The Nash product is then $S(U) = -2.5 \cdot 7.5 = -18.75$.

First, selecting a suitable social welfare function requires specific domain knowledge. Even in the rather abstract two-strategy PGG, we could have selected the wrong function. For instance, with utilitarian social welfare, the only inferior outcome would be obtained in case of mutual defection. With the k-rank dictator function, cooperation would be best for $k = 1$, whereas single-sided defection would be best for $k = 2$.

Second, optimizing social welfare in social dilemmas only works in closed multi-agent systems, i.e., in systems where we design or control all agents. Commonly, agents are assumed to optimize their individual utility value U^i instead of a globally-defined quantity such as social welfare $S(U)$. In open systems (Huynh et al., 2006b), we cannot control how every agent calculates his utility. A human-inspired approach is more successful in this respect: if we allow the part of the multi-agent system that we design or control to use mechanisms such as punishment or withholding action, based on their (expected) utility, and set up an environment in which these mechanisms may affect every agent (including ones we do not control), we are able to force even individually rational agents to care for the benefit of others, as failing to do so simply will cost them some of their reward.

Third, calculating an optimal strategy for agents given a certain social welfare function, may be impossible due to complexity issues. For instance, with 2 agents playing a PGG with 2 strategies, we already need to consider 4 different joint strategies, and therefore also 4 different Nash products. With n agents and m strategies, we would need to consider n^m Nash products. In a continuous strategy space (which we use in our social dilemmas), m is actually infinite, which implies that the optimal joint strategy (with respect to the social welfare function) may be at best approximated.

Fourth, given a certain social welfare function, we have not automatically found a mechanism that allows an agent to obtain an optimal resource allocation according to this function. Without punishment (or withholding action) in the PGG, there is no incentive to achieve a cooperative solution, unless we once again assume (as above) that we control all agents, in which case we might use, e.g., egalitarian welfare, or the Nash product, as both social welfare functions are maximized when all agents cooperate.[6]

To address the problem that mechanisms based on a social welfare function do not possess the elegance and efficacy of human mechanisms in social dilemmas, we may turn to recent

[6] Interestingly, the Nash product allows agents (with the restrictions of having a closed system, few agents and few strategies) to determine whether to punish in the PGG if this is possible. For example, in a two-agent, two-strategy game, we calculate $S(U)$, i.e., social welfare given the current allocation of resources (calculated after the agents both decided whether or not to cooperate), as well as $S(U')$, i.e., social welfare given the current allocation of resources and the assumption that agent 1 punishes 2 (or reverse, yielding $S(U'')$). This leads to three possible values for social welfare. If $S(U') > S(U)$, then agent 1 should punish. Similarly, if $S(U'') > S(U)$, then agent 2 should punish. Given a two-agent PGG with $r = 1.5$, $c = 10$, $c_p = 1$ and $e_p = 4$, and given that agent 1 cooperates and agent 2 defects, we find $S(U) = -2.5 \cdot 7.5 = -18.75$. Similarly, $S(U') = (-2.5 - 1) \cdot (7.5 - 4) = -5.25$ and $S(U'') = (-2.5 - 4) \cdot (7.5 - 1) = -42.25$. Clearly, $S(U')$ is the best value possible, and thus, agent 1 should punish agent 2. Given that agent 1 punishes agent 2, agent 2 may be deterred from defecting again, as cooperating with agent 1 gives him a higher reward (i.e., 5) than defecting against him (i.e., 3.5).

work in evolutionary game theory and statistical physics, which has created computational models of fairness that are explicitly based on proposed human mechanisms.

3.3.2 Fairness in evolutionary game theory and statistical physics

In statistical physics and evolutionary game theory, mechanisms facilitating satisfactory outcomes in social dilemmas are currently receiving a great deal of attention. Recent research is reported by, e.g., Dall'Asta et al. (2006a); Santos et al. (2006a); Rockenbach and Milinski (2006); Boyd and Mathew (2007); Hauert et al. (2007). We provide an elaborate discussion in later parts of the thesis, most notably in §5.1 and §6.1. Here, we limit ourselves to explaining why current research in evolutionary game theory and statistical physics does not provide us with complete solutions to our problem statement.

Generally, research in these areas is intended to be descriptive rather than computational, and therefore follows a distinct three-step approach. First, from carefully controlled experiments with human subjects, it is proposed that a certain mechanism, as observed in humans, may be responsible for the emergence of fairness and cooperation in social dilemmas. Second, the proposed mechanism is formalized and implemented in a multi-agent-learning context. Depending on the research background of the authors, this is done in three different ways, i.e., (1) using a learning approach, e.g., evolutionary algorithms (as in Santos and Pacheco, 2005; Santos et al., 2006c), or (2) using an evolutionary-game-theoretic analysis (as in Fehr and Schmidt, 1999; Nowak et al., 2000; Hauert et al., 2002), or (3) using tools 'borrowed' from statistical physics (as in Dall'Asta et al., 2006a). Third, experiments are performed to determine the effect(s) of the proposed mechanisms on learned strategies in (selected) social dilemmas; if these learned strategies indeed become more fair or cooperative, this provides support for the proposed mechanism. Mechanisms such as reputation (Nowak et al., 2000; Milinski et al., 2002), volunteering (Hauert et al., 2002), and rewiring in social networks (Santos et al., 2006c), have each been proposed and supported according to this procedure.

Obviously, it is clear that the goal of the research discussed above (i.e., being descriptive) is different from ours (i.e., being prescriptive). Thus, although such research provides us with many proposed human mechanisms, and has shown by experiments that these mechanisms enhance agents' abilities to reach satisfactory outcomes in social dilemmas, it also usually permits abstractions that we may not wish to permit.

Most notably, current research usually considers 'fairness' to be a synonym of 'cooperation'. The proposed mechanisms are therefore mostly only applied to games with two strategies; one of these strategies is then labeled as being 'fair' (or cooperative) and the other one as being 'selfish' (or defective). This abstraction is not made in our work, because assuming a discrete strategy space with only a few strategies is an over-simplification in many real-world examples of social dilemmas. Many tasks in which fairness plays a role, such as resource allocation, allow agents to choose from a large number of strategies, possibly even from a continuum of strategies. Identifying what is fair and what is not becomes more difficult when

continuous strategy spaces are involved, since it is no longer feasible (or even desirable) to label a single strategy (set) manually as 'fair' (or 'cooperative').

Cooperation may be a fair strategy in a context in which cooperation is indeed expected, whereas it may be simply an ignorant strategy in a context where individual rationality is expected (and thus, cooperators may be exploited). In other words, what exactly constitutes a fair strategy depends on the context or *culture* of the environment in which an agent is situated. We know from numerous studies with human players that the human notions of fairness and cooperation are also dependent on culture, although there is a general tendency to deviate from pure individual rationality, in favor of more socially aware strategies (Henrich et al., 2004; Oosterbeek et al., 2004).

Thus, while current research in the areas of evolutionary game theory and statistical physics meets our first and third requirement (see §3.1), i.e., it is game-theoretical as well as human-inspired, the second requirement is only met to a degree, as mechanisms are applied to social dilemmas that are 'simpler' than the dilemmas we are considering in our research. However, current research is clearly an important source of inspiration for our research, as the first steps on the way to human-inspired fairness in multi-agent systems have already been taken. In subsequent chapters, we will therefore extensively discuss (1) current research, and (2) how we extend this research in such a way that it fits all three requirements.

3.4 Chapter summary

In this chapter, we presented the foundations for human-inspired computational fairness. We discussed that three requirements need to be respected by computational models, i.e., (R1) they should be rooted in a game-theoretic background, (R2) they should be computationally applicable, and (R3) they should be explicitly linked to human fairness mechanisms. With regard to the last requirement, we distinguished that computational models should allow agents to answer three questions, i.e., (R3-Q1) how fair a given reward distribution is, (R3-Q2) whether a certain other agent needs to be punished, and (R3-Q3) whether we want to withhold action, e.g., by refraining from interacting again with a certain agent.

Then, we presented a template model for computational human-inspired fairness, which specifies how agents may answer the three questions of (R3), i.e., by applying a fairness utility function. (In the following chapters, we will present implementations of the template model, i.e., utility functions and appropriate mechanisms allowing agents to employ punishment and to withhold action if their utility value reflects this is necessary.)

At the end of the chapter, we provided an overview of related work in two areas of research. First, we discussed welfare economics, which presents models that meet the first two requirements, but not the third. Second, we addressed evolutionary game theory and statistical physics, which has a goal different from ours (i.e., it is descriptive rather than explicitly computational). As a result, researchers permit abstractions that we do not permit.

4 Inequity aversion

In this chapter, the foundations for computational human-inspired fairness are implemented by the descriptive model of human fairness named *inequity aversion*. This descriptive model stipulates that human decisions are influenced by resistance to inequitable outcomes. It is able to explain a great deal of (irrational) human decisions in interactions where limited resources need to be shared. Even though this is the case, the inequity aversion model has not yet convincingly found its way into multi-agent systems. We address this issue here by developing a computational model of fairness, based on inequity aversion. We show that our computational model allows agents to reach satisfactory, human-inspired solutions in both types of social dilemmas under study.

In §4.1, we discuss the descriptive inequity-aversion model, as developed in behavioral economics. Moreover, we outline work that has used some of the principles of inequity aversion in the context of multi-agent systems. In §4.2, we show how the inequity-aversion model can be applied to explain human strategies in the social dilemmas that we study. In §4.3, we build a computational model of inequity aversion, based on our template model for human-inspired fairness. This results in inequity-averse learning agents. We show that such agents are able to find satisfactory, human-like solutions in the social dilemmas under study. We conclude the chapter by answering RQ1 as well as RQ3 to RQ5 for the inequity-aversion model (see §4.4).

The work reported on in this chapter has partially been published in:

S. de Jong, K. Tuyls, and K. Verbeeck. Fairness in multi-agent systems. *Knowledge Engineering Review*, Vol. 23(2):153-180, 2008.

S. de Jong, K. Tuyls, and K. Verbeeck. Artificial agents learning human fairness. *Proceedings of the 7th International Conference on Adaptive Agents and Multi-Agent Systems (AAMAS)*, pp. 863–870, 2008.

S. de Jong, and K. Tuyls. Learning to cooperate in public-goods interactions. *Presented at the European Workshop on Multi-Agent Systems (EUMAS)*, 2008.

4.1 The inequity-aversion model

The inequity-aversion model is one of the least complicated models that have been proposed as a possible explanation why people choose fair solutions rather than individually rational ones. Research started in the 1970s (e.g., Walster et al., 1978). More recently, it received attention following the work by Fehr and Schmidt (1999). They define inequity aversion as follows: "Inequity aversion means that people resist inequitable outcomes; i.e., they are willing to give up some material reward to move in the direction of more equitable outcomes". As we will see below, inequity aversion assumes that humans resist inequity more when it is in their disadvantage, than when it is in their advantage. Fehr and Schmidt (1999) show that disadvantageous-inequity aversion manifests itself in humans as the "willingness to sacrifice potential gain to block another individual from receiving a superior reward".

From experiments with other primates, it is known that inequity aversion is not a typically human phenomenon. For instance, Brosnan and de Waal (2003) demonstrate that a non-human primate, the brown capuchin monkey (*Cebus apella*), responds negatively to an unequal reward distribution in exchanges with a human experimenter. Monkeys refused to participate if they witnessed a conspecific that obtains a more attractive reward (for instance, a grape instead of a piece of cucumber) for equal effort, an effect amplified when the partner received such a reward without any effort at all. As Brosnan and de Waal (2003) note, these reactions support an early evolutionary origin of inequity aversion.

Below, we first discuss a descriptive model of inequity aversion. Second, we provide an overview of the (rather scarce) work that has already been performed regarding the incorporation of inequity aversion in actual multi-agent systems.

4.1.1 A descriptive model of inequity aversion

One of the first researchers discussing and formalizing the role of inequity aversion in models of human fairness is Rabin (1993). Although Rabin's model requires an explicit representation of fair intentions and is only applicable to two-agent, zero-sum games, it is a clear landmark. In the inequity-averse model developed by Fehr and Schmidt (1999), the drawbacks of Rabin's model were alleviated. Essentially, the model of Fehr and Schmidt (1999) assumes that people are motivated by fairness considerations, in addition to being rational. The model is supported by experiments with human subjects, showing that this motivation is indeed present in many people, even without previous reinforcement.

To model inequity aversion, an extension of Homo Economicus (i.e., the classical game theoretic actor) is introduced, named Homo Egualis (Fehr and Schmidt, 1999; Gintis, 2001). Homo Egualis agents are driven by the following utility function:

$$U^i(r) = r^i - \frac{\alpha^i}{n-1}\sum_j \max\left\{r^j - r^i, 0\right\} - \frac{\beta^i}{n-1}\sum_j \max\left\{r^i - r^j, 0\right\}. \tag{4.1}$$

Here, $u^i = U^i(r)$ is the utility of agent $i \in \{1, 2, \ldots, n\}$. This utility is calculated based on agent i's own reward r^i and two inequity-averse terms related to considerations on how this reward compares to the rewards r^j of other agents j. Every agent i experiences a negative influence on its utility for other agents j that have a higher reward (weighed by a parameter α^i) as well as other agents that have a lower reward (weighed by a parameter β^i). The two resulting terms are subtracted from the utility of agent i. Thus, given its own reward r^i, agent i obtains a maximum utility u^i if $\forall j : r^j = r^i$.

Research with human subjects provides strong evidence that humans care more about inequity when doing worse than when doing better in society (Fehr and Schmidt, 1999). Thus, in general, $\alpha^i > \beta^i$ is chosen. Moreover, the β^i-parameter must be in the interval $[0, 1]$: for $\beta^i < 0$, agents would be striving for inequity, and for $\beta^i > 1$, they would be willing to "burn" some of their reward in order to reduce inequity, since simply reducing their reward (without giving it to someone else) already increases their utility value.

4.1.2 An existing computational model of inequity aversion

Even though the inequity-aversion model is conveniently represented by means of a utility function, it has thus far been used in a computational context only sporadically.[1]

Most notably, Verbeeck et al. (2007) use the inequity-averse Homo Egualis model as an inspiration for achieving a balance between optimality and fairness in multi-agent systems, more precisely, in coordination games. They focus on games with competition between players, where the overall performance is measured on a social level (e.g., performance is as good as that of the poorest-performing player, which corresponds to egalitarian social welfare; see §3.3.1). Similar to Homo Egualis agents, which are willing to give away a small portion of their own reward when they are performing better than other agents, the best-performing agents in the proposed system are willing to play suboptimal actions until another agent has become the best-performing agent. The first inequity-averse term of the Homo Egualis utility function (i.e., concerning disadvantageous inequity) is not explicitly considered.

Although Verbeeck et al. (2007) show that their proposed algorithm can find good solutions in many coordination games, they do not explicitly aim at an algorithm that is based on humans as much as possible. For instance, by discarding the first inequity-averse term of the Homo Egualis utility function, they ignore the most powerful of the two terms. In our own work, we therefore use the Homo Egualis utility function directly and completely, rather than being inspired by the principles behind it. Using the function directly ensures maximal alignment between our multi-agent systems and human expectations.

[1] There is some research in which the ideas underlying inequity aversion are computationally modeled and applied to bargaining (Gerding et al., 2003). The authors do not explicitly mention inequity aversion, but present a number of piecewise linear utility functions aimed to match observed human behavior. Results obtained by bargaining agents endowed with these functions are quite similar to the results we find for the NBG (see §4.3.4).

4.2 Inequity aversion in social dilemmas

The Homo Egualis utility function has been shown to explain an "impressive amount of empirical evidence [in humans]" (Dannenberg et al., 2007). Existing work explains how inequity aversion may be used to describe human decision-making in many games, such as the two-player UG (Fehr and Schmidt, 1999). We add three new explanations, i.e., concerning (1) a multi-player UG, (2) the NBG, and (3) the PGG. Thus, we show that inequity aversion may be used to address both types of social dilemmas under study, i.e., the agreement dilemma (§4.2.1) and the tragedy of the commons (§4.2.2). For the reader's convenience, we briefly repeat the principles of each of the three games in every subsection. A more extensive explanation of the three games is given in §2.3.

4.2.1 The agreement dilemma

In this thesis, the social dilemma we labeled *agreement dilemma* is represented by two games, i.e., the UG (with two players as well as with more than two players) and the NBG. In both games, agents must ensure that they agree on the strategies to follow. Otherwise, they run the risk of returning home empty-handed. Below, we discuss how inequity aversion may explain human strategies in these two games.

The Ultimatum Game

In the UG (see §2.3.1), two agents bargain about the division of a (small) reward R. The first agent proposes a division (e.g., '€8 for me, €2 for you'). The second agent can either accept this proposal (yielding, e.g., a reward of €2) or reject it, yielding a reward of €0 for both players. The Nash equilibrium of this game is for the first agent to offer the smallest amount possible to the second agent; the second agent should accept. In contrast, human players generally neither offer, nor accept the smallest amount (Bearden, 2001; Oosterbeek et al., 2004; De Jong et al., 2008). The Homo Egualis utility function illustrates this behavior.

Using the Homo Egualis utility function in the 2-agent game, Fehr and Schmidt (1999) calculate that the optimal reward for agent 1 depends on two factors, viz. β^1 and α^2. More precisely, in the 2-agent game, we have $n = 2$ and $r^1 + r^2 = R$. Thus, the Homo Egualis utility function can be rewritten for two agents as:

$$U^i\left(r^i, r^j\right) = r^i - \alpha^i \max\left\{R - 2r^i, 0\right\} - \beta^i \max\left\{2r^i - R, 0\right\}. \tag{4.2}$$

Thus, if $\beta^1 > 0.5$, agent 1's utility u^1 will decrease with values of $r^1 > 0.5R$. This implies that agent 1 will give $0.5R$ to agent 2 if $\beta^1 > 0.5$. If $\beta^1 < 0.5$, agent 1's utility is not decreased by increasing his reward r^1. The agent would like to keep everything to himself. However, he must ensure that agent 2 receives a reward which is not rejected. Since rejecting yields a

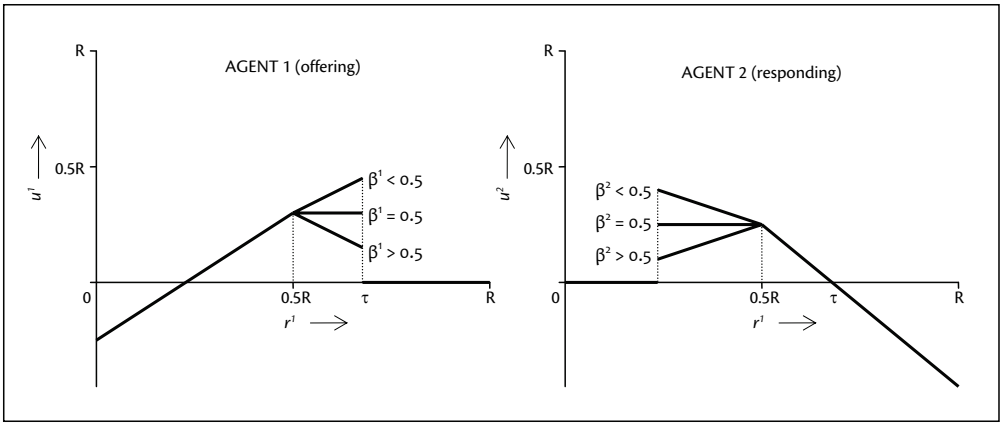


Figure 4.1 The Homo Egualis utility function in the 2-agent Ultimatum Game. We illustrate the functional mapping between the reward that the offering agent (agent 1) keeps to himself (r^1), and the utility experienced by this agent (u^1) as well as the responding agent (u^2).

reward $r^1 = r^2 = 0$, it will also lead to a utility $u^1 = u^2 = 0$. Agent 2 will therefore reject iff $r^2 - \alpha^2\left(R - 2r^2\right) < 0$. Solving this inequality with respect to r^2, we obtain:

$$r^2 \geq \frac{\alpha^2}{1 + 2\alpha^2} \cdot R. \tag{4.3}$$

Thus, agent 1 can keep at most $\tau = \left[1 - \frac{\alpha^2}{1+2\alpha^2}\right]R$ to himself.

We note that $\lim_{\alpha^2 \to \infty} r^2 = 0.5R$. Therefore, agent 2 can expect to obtain at most half of the total reward. For additional clarity, the functional mapping between r^1 and u^1 as well as u^2 is shown in Figure 4.1. As long as $r^1 \leq 0.5R$, the utility u^1 is increasing. Then, depending on β^1, the utility increases, remains constant, or decreases. Finally, after r^1 exceeds the threshold τ, agent 2 rejects (as agent 2 experiences a utility $u^2 < 0$ for $r^1 > \tau$, he is better off rejecting, leading to $u^1 = u^2 = 0$). Due to the threshold, the utility function for agent 1 is not continuous: there is a discontinuity immediately after the maximum.

In brief summary to the above, we may state that strictly positive offers to the second player in the 2-agent UG are explained by inequity-aversion; if the second player is inequity-averse, the first player must ensure that his offer leads to a sufficiently equitable division of the amount at hand.

The Ultimatum Game with more than two agents

Usually, the UG is played with only two agents. As we are interested in a multi-agent perspective, we also analyze the role of inequity aversion in UGs with more than two agents. We define a game in which $n - 1$ agents one by one take a portion of the reward R. The last

agent, n, receives what is left. In this game, we specify that any reward distribution r for which $\forall i : u^i(r) \geq 0$ holds, is acceptable.

We now wish to to calculate the minimal reward needed by any of the agents. To this end, we need to assume that the agent that receives the lowest reward, also receives the lowest utility, independent of how the remaining reward is distributed over the other agents. If we cannot use this assumption, calculating the minimal reward needed to satisfy the agent with the lowest utility becomes a combinatorial problem. Since we are using this calculation to verify the results of our learning agents, we need to assume the same value for α^i (and $0 \leq \beta^i < \alpha^i$) for all agents, in both our analysis (presented here) and our experiments (presented in §4.3.4 and §4.3.5). We note that using 'private' values for α^i would certainly allow agents to learn a solution. We would only be unable to verify this learned solution using a calculated solution.

With this assumption in place, the worst-performing agent's utility value can be calculated as $U^i(r) = r^i - \frac{\alpha^i}{n-1} \sum_j \max\{r^j - r^i, 0\}$, since the term involving β^i evaluates to 0 for the agent with the lowest reward. Thus, assuming a reward of R, we may derive the following.

$$\begin{array}{lll} U^i(r) = r^i - \frac{\alpha^i}{n-1} \sum_j \max\{r^j - r^i, 0\} & \geq & 0 \\ U^i(r) = r^i - \frac{\alpha^i}{n-1} \sum_{r^j > r^i} \left(r^j - r^i\right) & \geq & 0 \\ r^i - \frac{\alpha^i}{n-1} \left[\sum_{r^j > r^i} r^j \; - \; \sum_{r^j > r^i} r^i\right] & \geq & 0 \\ r^i - \frac{\alpha^i}{n-1} \left[\left(R - r^i\right) \; - \; (n-1)\, r^i\right] & \geq & 0 \end{array}$$

This simplifies to:

$$r^i \geq \frac{\alpha^i}{\alpha^i n + n - 1} \cdot R. \tag{4.4}$$

For instance, with three agents and $\forall i : \alpha^i = 0.6$, we obtain that every agent needs at least $0.1578R$ in order to accept the deal at hand. Typically, as in the two-player UG, the last agent will be the worst-performing agent (and therefore the most critical one), as the other agents can actively reduce his reward to the lowest value that still leads to a positive utility.

The Nash Bargaining Game

As has been explained in §2.3.1, the NBG is played by two or more agents. The agents simultaneously request a reward, taken from an available reward R. They obtain what they have asked for only if the total amount requested is at most R. A common human solution to this game is truly fair: an even split (Nydegger and Owen, 1974; Roth and Malouf, 1979; Yaari and Bar-Hillel, 1981; Van Huyck et al., 1995). This solution obviously only works if all agents agree not to request more than an equal share.

Inequity aversion may increase the ability of agents to agree on such a fair solution. To this end, we give agents an additional action, i.e., even if the reward distribution was successful,

agents may compare their reward with that of others. Then, if their reward is too small, they may reject the distribution, once again leading to all agents obtaining a reward of 0. To decide whether their reward is satisfactory, agents use the Homo Egualis utility function, as in the UG. Thus, we may perform the same analysis as in the UG and obtain that any solution conforming to Equation 4.4 is not rejected. For example, with $n = 2$ and $\alpha_1 = \alpha_2 = 0.6$, every agent should obtain at least $0.27R$.

4.2.2 The tragedy of the commons

The tragedy of the commons is often represented by the PGG. As indicated in §2.3.2, the game requires a group of agents to decide whether or not to invest an amount c in a common pool. The invested money will be multiplied by a factor r (usually 3) and subsequently divided equally over all players. Thus, if everyone cooperates by contributing money, an individual agent is better off by not contributing. The Nash equilibrium is therefore for all agents to refuse any contribution, which leads to a monetary gain of 0 for all of them, where they could have gained $(r-1)c$.

In the remainder of this subsection, we first discuss human behavior in this game, and briefly look at research aimed at explaining this behavior.[2] After our brief discussion of existing research here, we present an alternative explanation of human behavior in the PGG using inequity aversion. Finally, we examine the influence of human behavior, as predicted by inequity aversion, on the emergence of cooperative strategies.

Human behavior in the Public Goods Game

In the PGG, many human players initially contribute high amounts; these amounts decrease over time. The initial success obtained by few (relative) defectors is quickly learned by the others, leading to lower and lower contributions and rewards. Introducing the option to punish defectors (at a small cost to the punisher) leads to cooperation being maintained (Yamagishi, 1986; Fehr and Gaechter, 2000, 2002), even if the participants are told that they will never interact with the same other participant(s) again. Thus, punishment in this game is truly altruistic. Interestingly, Henrich et al. (2006) show that punishment and altruism are indeed correlated in human societies: more altruistic societies punish more, *et vice versa*.

Many researchers have attempted to explain this behavior. Sufficient explanations have thus far not been found, mainly because cooperation (i.e., contributing high amounts) and altruistic punishment are not evolutionarily stable strategies (Hauert et al., 2002, 2007). Cooperators cannot invade a population of defectors, since a single cooperator would see his reward reduced to (near-)zero. In addition, we can now have so-called second-order free-riders: a player may choose to cooperate, but at the same time, since punishment is costly, he may refuse to punish defectors. Thus, cooperators easily invade a group of punishers; after that,

[2] As existing research in the PGG mostly concerns more complicated models than inequity aversion (e.g., reputation models), more details concerning existing research may be found in Chapter 5.

defectors can invade the cooperators. Higher-order punishment can be introduced and will work from a computational point of view, but this phenomenon is not observed in humans.

Thus, the question seems to shift: instead of investigating why people cooperate, we need to investigate why people actually perform altruistic, costly punishment. Although concepts such as reputation and volunteering do promote altruistic punishment (see Chapter 5), they have not been proven necessary for human players to punish others. Moreover, none of these concepts leads to stable cooperation being achieved and maintained. Thus, the question why humans punish each other in games such as the PGG, is still only partially answered. In this section, we provide an answer to this question, using the inequity aversion model.

Inequity aversion explains punishment

We are not the first to investigate the role of inequity aversion in the PGG. Dannenberg et al. (2007) examined how the Homo Egualis utility function may explain human strategies in this game. Subjects first played Ultimatum Games, the result of which were used to group them into classes with similar preferences (i.e., similar estimated α^i- and/or β^i-parameters). Next, subjects were matched into pairs in three different ways, i.e., (1) 'fair' pairs with highly inequity-averse subjects, (2) 'selfish' pairs with subjects who did not care strongly about inequity, and (3) 'mixed' pairs in which one subject was 'fair' and the other one 'selfish'. Next, pairs played a standard PGG and one in which punishment could be used. According to Dannenberg et al. (2007), results suggest that the inequity-averse model at least has some explanatory power, especially for the 'fair' pairs.

Our investigation of inequity aversion in the PGG is more formal and analytical than the experimental work performed by Dannenberg et al. (2007). In our analysis, the key idea is that agent i punishes agent j iff i's utility increases because of this punishment. Since agents specifically punish others, we consider that utilities are evaluated on a pair-wise basis, even in interactions in which more than two agents participate. Thus, whether agent i punishes agent j depends only on the rewards these agents will receive, i.e., r^i and r^j. Given this pair-wise comparison, agent i's utility as a consequence of his own reward r^i and agent j's reward r^j can be calculated as:

$$U^i\left(r^i, r^j\right) = r^i - \alpha^i \max\left\{r^j - r^i, 0\right\} - \beta^i \max\left\{r^i - r^j, 0\right\}. \tag{4.5}$$

The term related to β^i can be omitted, since considering to punish an agent j with $r^j < r^i$ would only increase inequity. Thus, when considering punishment, $r^j > r^i$ will hold. Assuming that c_p denotes the cost of punishing agent j, and that e_p denotes the effect of this punishment on agent j, agent i can calculate his utility in case of punishment as follows:[3]

$$U^i\left(r^i, r^j\right) = \left(r^i - c_p\right) - \alpha^i\left(\left(r^j - e_p\right) - \left(r^i - c_p\right)\right). \tag{4.6}$$

[3] Note that we assume that j still has a higher reward than i after i punished j, i.e., $r^j - e_p \geq r^i - c_p$. In this chapter, as in existing work on the PGG, we ensure that this is the case by making strategies sufficiently different.

Sensibly, agent i will punish if his utility as a result of punishing is higher than his utility as a result of not punishing. Thus, considering $r^j > r^i$, agent i punishes iff:

$$\left(r^i - c_p\right) - \alpha^i \left(\left(r^j - e_p\right) - \left(r^i - c_p\right)\right) > r^i - \alpha^i \left(r^j - r^i\right). \tag{4.7}$$

This inequality can be simplified to $-c_p + \alpha^i e_p - \alpha^i c_p > 0$. With respect to α^i, the inequality may be expressed as:

$$\alpha^i > \frac{c_p}{e_p - c_p}. \tag{4.8}$$

Thus, whether agent i punishes agent j does not depend on the agents' actual rewards r^i and r^j; the decision will be taken on the basis of Equation 4.7. If the condition that $r^j > r^i$ is met, and the inequality presented in Equation 4.8 holds, agent i will punish agent j. A common setting for the PGG is $c_p = 1$ and $e_p = 4$; in this setting, $\alpha^i > \frac{1}{3}$ suffices to make agent i a punisher. Experiments with humans reveal that humans often use much higher values of α (Fehr and Schmidt, 1999; Dannenberg et al., 2007). Clearly, if all agents agree on defecting, everyone receives the same reward and therefore does not punish each other according to the inequity-aversion model. However, it is known that most humans are initially cooperative; thus, according to the model, they would punish (rare) defectors, forcing them to be more cooperative as well.

Punishment leads to cooperation

The presence of a punishment mechanism in a population of agents increases the probability that agents choose cooperative actions (Sigmund et al., 2001; Fehr and Gaechter, 2002; Boyd et al., 2003). Usually, in research focused on the PGG, agents are restricted to two possible actions, i.e., a contribution of 0, denoted as a defective action, or a contribution of the maximum amount c, denoted as a cooperative action. As we saw above, using punishment based on inequity aversion, the precise amounts contributed by agents are not relevant; as soon as the amount that agent j contributes is below the amount contributed by agent i, agent j faces a large probability of being punished. Thus, for the sake of clarity, we also consider a PGG with only two possible contributions (i.e., 0 or c).

Considering that most humans are sufficiently inequity-averse to prefer punishing over second-order free-riding, we assume that all cooperators are willing to punish. In this case, given that n_d out of n agents are defectors, we obtain that these defectors receive a reward of $\frac{1}{n}(n - n_d)cr - (n - n_d)e_p$. Cooperators obtain $\frac{1}{n}(n - n_d)cr - c - n_d c_p$. Rationally, cooperation is therefore a dominant strategy iff $(n - n_d)e_p > c + n_d c_p$. When evaluating this inequality, we obtain:

$$n_d < \frac{ne_p - c}{e_p + c_p} \rightarrow \text{cooperation is dominant.} \tag{4.9}$$

Apart from the fact that the formula above validates two obvious conclusions, i.e., (1) a higher punishment e_p leads to more trouble for defectors and (2) a higher cost of punishment c_p leads to more trouble for cooperators, we may also draw a less obvious conclusion from this formula, namely that (3) an increasing number of agents helps cooperators. For instance, for a common setting, such as $n = 5$, $c = 10$, $c_p = 1$, $e_p = 4$, we obtain that cooperation is a dominant strategy for $n_d < 2$ (or 40% of n). For $n = 10$ and identical other parameters, we obtain $n_d < 6$ (or 60% of n), i.e., a higher percentage of defectors is needed in order to make defecting more profitable than cooperation.

4.3 Inequity-averse learning agents

In this section, we instantiate our computational framework using inequity aversion. The agents in our multi-agent systems are learning by means of learning automata. The agents are confronted with both types of social dilemmas, by means of UGs, NBGs, and PGGs. It is important to note that purely rational agents obtain an unsatisfactory reward at least in the UG and the PGG, and that the NBG has not yet been considered in the context of inequity aversion. Below, we wish to determine (1) whether our agents learn to find and maintain satisfactory solutions in all three games, and (2) whether these solutions correspond to solutions found by humans, as reported in literature.

The remainder of this section is organized as follows. In §4.3.1 we discuss how we build upon the foundations of computational fairness, as presented in Chapter 3. In §4.3.2, we outline our general approach, combining the Homo Egualis utility function with learning automata. In §4.3.3, we present some modifications to the learning rule of continuous action learning automata, which are needed to address the discontinuity in the Homo Egualis utility function. In §4.3.4 and §4.3.5, we present a set of experiments and results in the two types of social dilemmas.

4.3.1 Building upon the foundations

The foundations of computational fairness (Chapter 3) consist of two elements, i.e., a set of requirements (§3.1), and a template model (§3.2). Since inequity aversion is represented by the Homo Egualis utility function, it conveniently meets two of our three requirements, i.e., it is (R1) rooted in game theory, as well as (R3) human-inspired. It meets our second requirement only to a degree (i.e., R2, the model should be computationally applicable), since the Homo Egualis utility function is not sufficiently smooth. Below, we discuss how this problem may be addressed, i.e., by slightly modifying the learning rule used by our learning algorithm (CALA). Using the modified learning rule, the learning algorithm is able to find the optimum specified by the Homo Egualis utility function. Thus, the template model may be instantiated by inequity aversion in a rather straightforward way for both types of social dilemmas, i.e., we may use the Homo Egualis utility function directly. The expected

outcomes generated by the resulting computational model of inequity aversion are already known and discussed in §4.2.

In the UG and the NBG, we calculate the agents' utility over the entire reward distribution, i.e., $u^i = U^i(r)$. In the PGG, we use a pairwise utility function $u^i = U^i(r^i, r^j)$, allowing agents to punish specific other agents.

As discussed in Chapter 3, a computational model of fairness allows agents to answer three questions, i.e., (R3-Q1) what is meant by a fair solution, (R3-Q2) whether a certain solution requires agents to punish others, and (R3-Q3) whether agents should withhold their actions or refrain from participating completely, given their expected outcome. Depending on the answers to each of the three questions, there may be certain consequences. We address each of these consequences here.

R3-Q1 As explained in Chapter 3, by a fair solution we imply a reward distribution r for which $(1 - \epsilon)n$ agents i experience a utility $u^i > u_0^i$. In this chapter, we use $\epsilon = 0$; thus, we do not allow any agent to experience a utility lower than the baseline utility u_0^i. In all three games under study, this baseline utility $u_0^i = 0$ for all agents i, as no agent gains or loses anything by not playing. Thus, we strive for reward distributions r for which all agents i experience a utility $u^i \geq 0$.

R3-Q2 Considering punishment, agents will punish others iff this increases their utility u^i. We compare the solutions of a multi-agent system that may use punishment to those of a multi-agent system that may not use punishment. Clearly, punishment is effective if the first system learns a better solution than the latter.

R3-Q3 Withholding action implies that agents may decide not to participate if they expect that participating will yield a negative utility u^i. In this chapter, we do not allow agents to do this.

4.3.2 Methodology

In our methodology, we use a combination of the Homo Egualis utility function (see §4.1) and learning automata (CALA in the agreement dilemmas and FALA in the tragedy of the commons; see §2.2.2). We use a four-step process at every iteration t. To avoid unnecessary repetition, we explain this for CALA only. FALA are used in a similar manner, but perform only one action per iteration, in contrast to the two actions performed by CALA (see §2.2.3).

First, every agent i is equipped with a learning automaton, which selects actions μ^i and x^i.

Second, the joint actions μ and x lead to reward distributions $R(\mu)$ and $R(x)$, according to the rules of the game at hand. In the UG, every agent receives what he has asked for, unless there is insufficient reward remaining due to the actions of preceding agents. In this case, the agent receives what is remaining. In the NBG, everyone receives what they have asked for, unless the sum of their requests exceeds the total reward R. In that case, everyone receives 0. In the PGG, actions denote contributions; the contributions of all agents are summed, multiplied by a certain factor, and subsequently equally divided over all the agents.

Third, the reward distributions are mapped to a utility value for each agent i, i.e., $U^i(R(\mu))$ and $U^i(R(x))$, using the Homo Egualis utility function. For any of the two joint actions, if any agent experiences a higher utility by punishing than by refraining from punishment, this agent may be allowed to punish. In the UG and NBG, this punishment nullifies all rewards, leading to a utility of $U^i(r_0)$ for all agents i. In the PGG, punishment reduces the reward of one or more specific agent(s), at a smaller cost to the punisher.

Fourth, the utility values are reported to the learning automata, which subsequently update their strategies. We note that the n-agent UG requires $n-1$ automata, whereas the n-agent NBG and PGG require n automata. In the UG, the last agent's behavior is static: he simply rejects if his utility drops below 0. In the NBG and PGG, all agents are the same.

We use the same parameters in the Homo Egualis utility function (i.e., α^i and β^i) for all agents participating. This makes the analysis and verification of the outcomes easier, in particular when dealing with many agents. Results obtained by giving each agent i private α^i- and β^i-values will be highly similar to our results, but calculating an expected or optimal solution to compare these results with, is difficult, as we explained in §4.2.

4.3.3 Modifications to the learning rule for multiple CALA

For CALA, convergence to a local optimum has been proven in the case of smooth and continuous feedback functions (Thathachar and Sastry, 2004). Howeverm the Homo Egualis utility function is not always sufficiently smooth, especially when multiple agents are learning joint actions. We discuss two problems related to this issue. Both problems need to be addressed without affecting the convergence of multiple CALA. We note that in this chapter, these problems only apply to the UG and NBG, as the PGG is addressed with FALA, which do not require a smooth feedback function.[4]

The first problem is generally present when multiple CALA are learning, and there is a sharp transition from having a 'good' joint action to having a 'bad' joint action. With the Homo Egualis utility function, this problem arises when the automaton i is near the optimum, and either its x^i-action or its μ^i-action is slightly too high. As can be seen from Figure 4.1, one of the actions will then yield (almost) optimal utility, whereas the other action yields a utility of 0. Due to the CALA update function, the μ^i of the underlying Gaussian N^i will therefore shift drastically.[5] As this is a highly undesirable effect, we chose to restrict the terms of the

[4] Note that such modifications are not uncommon in the literature; see, e.g., Selten and Stoecker (1986) on learning direction theory. Grosskopf (2003) successfully applied directional learning to the setting of the UG, focusing on responder competition (which is not further addressed in this thesis).

[5] We illustrate this problem by means of an example. Assume two CALA are learning from each other in the UG. One has $\mu = 5$ and $\sigma = \sigma_L$, the other has $\mu = 6$ and $\sigma = 1$. Their joint actions currently are, e.g., $\mu = (5, 6)$ and $x = (5.00001, 4.5)$. Thus, from the perspective of the first automaton, its μ-action is punished, yielding a reward of 0 according to the rules of the UG; its x-action is not punished, yielding a reward of (approximately) 5. If we use these values in the formula of Equation 2.2, with $\lambda = 0.01$ and $\sigma_L = 10^{-7}$, we obtain a new μ-value of 5×10^7 for the first automaton. This value is far beyond the allowed boundaries.

update function. More precisely, we restrict

$$U^i(R(x)) - U^i(R(\mu)) \in \left[-\Phi\left(\sigma^i\right), \Phi\left(\sigma^i\right)\right]. \tag{4.10}$$

In essence, this addition has the same effect as a variable learning rate, which is not uncommon in literature (e.g., Bowling and Veloso, 2002). In normal cases, i.e., when the automaton is not near the discontinuity, the restriction is hardly, if ever, exceeded. Near the discontinuity, it prevents drastic shifts. This addition to the learning rule does not affect convergence.

The second problem is more specific to the Homo Egualis utility function: there are large areas in the joint action space in which there is no useful feedback at all, as joint actions in these areas lead to zero feedback for all agents. If the automata are currently searching in such an area, it is likely that both their joint μ-actions as well as their joint x-actions yield a utility of 0 – i.e., indeed, the automata receive no useful feedback at all. In this case, due to the CALA update function, the underlying Gaussian's μ^i and σ^i are not changed. Therefore, in the next iteration, there is a high probability that the automata again receive a utility of 0 for both joint actions. In other words, if this happens, the automata are very likely to get stuck. We address this issue by including the knowledge that, if both μ^i and x^i yield a utility of 0, the *lowest* action was nonetheless the best one (recall that we are only considering the UG and the NBG here). Therefore, we set

$$U^i(R(x)) = \max\left(U^i(R(x^i)), \mu^i - x^i\right), \tag{4.11}$$

essentially driving the automaton's μ^i downward. Once again, in normal cases, the update function remains unchanged. In cases where the automaton receives no useful feedback, it can still update the parameters of the underlying Gaussian in such a way that an agreement is more probable. This addition does not hinder convergence.

4.3.4 The agreement dilemma

In this section, we present our experiments and results in both agreement dilemma games (the UG and the NBG). As we perform a large set of experiments, we will first provide a general overview. Subsequently, we will provide a detailed discussion of the experiments, results, and valuable observations.

In every experiment, the agents use CALA for learning and the Homo Egualis utility function is applied to the reward distributions. Except where noted otherwise, the following settings are observed.

1. The CALA parameters are set to $\lambda = 0.01$, $K = 1$ and $\sigma_L = 10^{-6}$.
2. The number of agents is varied between 2, 3, 4, 10, and 100.
3. The Homo Egualis parameters are set to $\alpha^i = 0.6$ and $\beta^i = 0.3$ for all agents i.

4. Punishment (possible or not possible) is indicated per experiment.
5. If necessary, the modifications to the CALA learning rule (see §4.3.3) are used.
6. All experiments last for 10,000 rounds and are run 1000 times.
7. The agents start from an initial solution of equal sharing, i.e., for all n agents' CALA, we set $\mu^i(0) = R \cdot \frac{1}{n}$ and $\sigma^i(0) = 0.1\,\mu^i(0)$. In the experiments, we use $R = 100$.

Experiments and results are detailed in seven parts, i.e., (1) concerning the UG with 2 agents, (2) concerning the NBG with 2 agents, (3) concerning the UG with 3, 4, and 10 agents, (4) concerning the NBG with 3, 4, and 10 agents, (5) concerning the UG with 100 agents, and (6) concerning the NBG with 100 agents, and finally, (7) concerning the NBG in which punishment is disabled.

In every part, we display a table containing the most important results. In the tables, the 'Game' column indicates the game under consideration, possibly adding information on experiments where different settings were used than those detailed above. The number of agents is denoted under 'Ags.'. Whether or not punishment could be used by the agents is indicated in the 'Pun.' column (i.e., with punishment enabled, agents could reject a solution for which they obtained a negative utility). The analytically determined solution, i.e., the solution resulting from playing optimally, which depends on the aforementioned columns, is indicated in the 'Sol.' column (either the exact solution or the conditions that a solution must satisfy, if any). Whether or not the modified learning rule was used in the CALA, is indicated under 'Mod. LR'.

Next, we show experimental results (average reward and standard deviation; the values are separated per agent by a '/'). In every case, we also measure how many times a computationally fair solution, as specified by our computational framework and an error margin $\epsilon = 0$, was found and subsequently maintained over the full 10,000 iterations (results are displayed under 'Maint.'). Finally, we indicate whether the experiment can be considered a success; more precisely, we consider the experiment to be successful (+) if a fair solution, for which all agents experienced a non-zero utility, was found and maintained in all experimental runs. An experiment is a failure (-) if such a solution was not found and/or maintained in any run. Otherwise, an experiment was neither a success, nor a failure (○).

Table 4.1 Experimental settings in the 2-agent Ultimatum Game

Game	Agents	Experiment	Value of α	Value of β	Punishment
UG	2	1	0.6	0.3	No
UG	2	2	0.6	0.3	Yes
UG	2	3	0.6	0.7	Both (2 runs)

Table 4.2 Inequity aversion in the 2-agent Ultimatum Game

Game	Ags.	Pun.	Sol.	Mod. LR.	Avg.	Stdev.	Maint.	Res.
UG	2	no	50.0/50.0	no	50.1/49.9	0.2/0.2	100%	+
UG	2	no	100.0/0.0	no	100.0/0.0	0.0/0.0	100%	+
UG	2	yes	72.7/27.2	no	72.3/27.7	5.5/5.5	100%	+

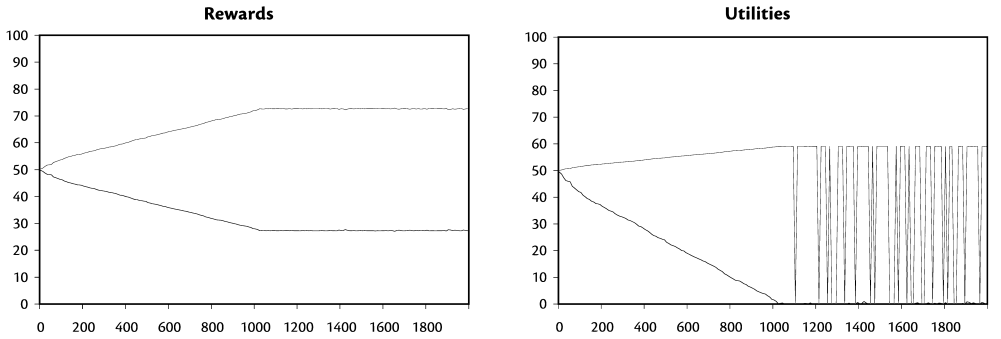


Figure 4.2 Learning to play the 2-agent Ultimatum Game with inequity aversion

(1) The 2-agent Ultimatum Game

In the 2-agent UG, we use only one learning automaton; the second agent's behavior is static. We perform three different experiments. For convenience, the settings for each experiment are given in Table 4.1. Results are reported in Table 4.2.

In the first experiment, we use standard settings for α and β, but we disable the punishment option. In the absence of punishment, the first agent can simply take the whole reward for himself, as predicted also by Fehr and Schmidt (1999).

In the second experiment, we use the same settings, but with punishment enabled, i.e., if the second agent obtains a utility value below 0, he rejects, leading to a reward of 0 for both agents. With $\beta = 0.3$ and $\alpha = 0.6$, Fehr and Schmidt (1999) predict a reward fraction of $\frac{0.6}{1+2\times0.6} \approx 0.27$ being given to the second agent. Results of one particular run are illustrated in Figure 4.2 (first 2000 iterations). The learning process turns out to be robust with respect to the parameters used. As the solution is not rejected by any agent, it is also a fair solution. Thus, we see that our agents are capable of playing the two-player UG in a 'human' way.

In the third experiment, we (exceptionally) set $\beta = 0.7$ and $\alpha = 0.6$. The first setting theoretically ensures that agent 1 gives 50% to agent 2, even in the absence of punishment. We therefore may disable the punishment option. The automaton maintains to offer 50%, without any enforcement (i.e., punishment). With punishment, exactly the same happens (and punishment is never needed).

Table 4.3 Inequity aversion in the 2-agent Nash Bargaining Game

Game	Ags.	Pun.	Sol.	Mod. LR.	Avg.	Stdev.	Maint.	Res.
NBG	2	yes	all≥27.2	no	46.5/46.6	2.9/2.7	0%	-
NBG	2	yes	all≥27.2	yes	48.2/48.2	2.4/2.4	100%	+
NBG	2	no	any	yes	48.3/48.3	2.7/2.6	100%	+

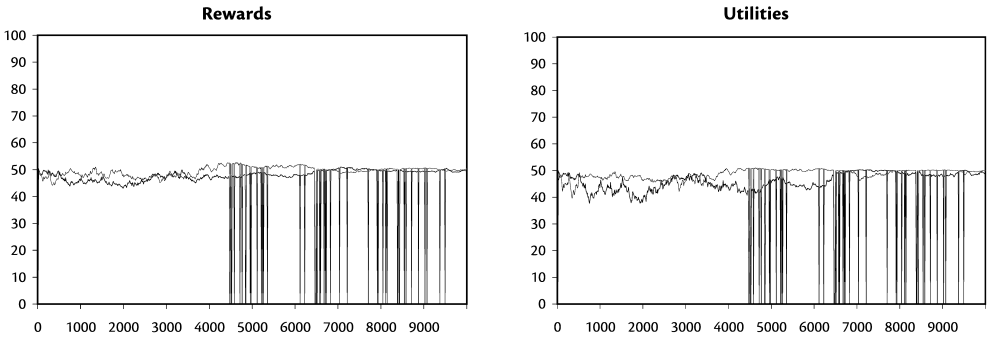


Figure 4.3 Learning to play the 2-agent Nash Bargaining Game with inequity aversion

(2) The 2-agent Nash Bargaining Game

In the two-agent NBG, we need two learning agents and therefore also two CALA. Whenever the joint action of the CALA results in a summed reward higher than R, both agents receive 0. Whenever the summed reward is at most R, the Homo Egualis utility function is applied to determine whether each agent considers their respective reward to be fair. If not, they can choose to give both themselves and the other agent a reward of 0. We perform three experiments, which are reported in Table 4.3.

In the first experiment, we use $\beta = 0.3$ and $\alpha = 0.6$, enabling punishment. Clearly, any solution yielding a reward of at least 27 for both agents is acceptable using the Homo Egualis utility function. As can be seen in Table 4.3, the CALA do *not* learn a solution now.

Therefore, in the second experiment, we introduce the modifications to the learning rule, as outlined in §4.3.3. This time, the CALA find and maintain a correct, fair solution. Note that the solution is nearly Pareto-optimal as well as close to a 50-50 split, as predicted in literature. Typical results obtained here are displayed in Figure 4.3.

In the third experiment, we disable punishment, as we observed that punishment was never used by the agents, even though it was possible. Indeed, results are not substantially different when we disable punishment.

Thus, with the modifications to the CALA learning rule, a fair, 'human' solution to the two-player NBG can be learned, with and without punishment.

Table 4.4 Inequity aversion in the multi-agent Ultimatum Game

Game	Ags.	Pun.	Sol.	Mod. LR.	Avg.	Stdev.	Maint.	Res.
UG	3	yes	all≥15.8	yes	41.0/41.0/18.0	1.6/1.5/1.7	100%	+
UG	4	yes	all≥11.1	yes	29.0/29.0/29.0/13.0	1.5/1.5/1.5/1.6	100%	+
UG	10	yes	all≥4.0	yes	10.5/10.5/.../6.7	1.1/1.1/.../2.0	100%	+

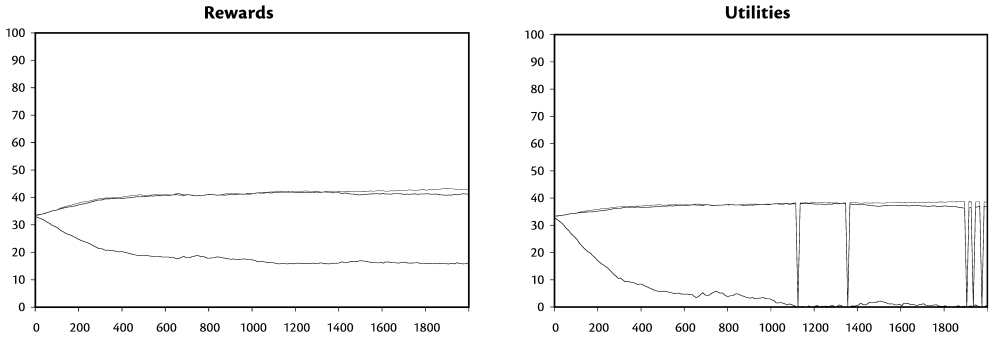


Figure 4.4 Learning to play the 3-agent Ultimatum Game with inequity aversion

(3) The multi-agent Ultimatum Game

As has been outlined before, in the multi-agent UG, multiple agents take turns in taking some of the reward R for themselves. The last agent in the row obtains what is left. We perform three experiments, i.e., with 3, 4, and 10 agents. Results are given in Table 4.4.

Using the standard settings of CALA for the multi-agent UG turns out to lead to invalid solutions. For this reason, we introduce the modifications to the learning rule. In this case, the CALA can indeed learn to obtain and maintain a valid, fair solution. Typical results for a three-player game are shown in Figure 4.4 (first 2000 iterations). We see that the last agent's utility is quickly decreased to a low positive value by keeping approximately 16 for this agent. The other two agents obtain an equal split of the remaining 84. Note that the first agent could have exploited the other agents. More precisely, he could have obtained approximately 64 without the other two agents rejecting. However, since all agents are learning simultaneously, both agent 1 and agent 2 are increasing their rewards at the same time; at a certain point, they thus have reduced agent 3's utility value to 0. Then, if any agent wishes to increase his reward, agent 3 will reject. Thus, agent 1 cannot exploit agent 2 unless agent 2 willingly lowers his reward, which simply will not happen.

Research with humans has shown that only a minority of human subjects actually exploits the other player(s) (Fehr and Schmidt, 1999; Bearden, 2001; Oosterbeek et al., 2004; De Jong et al., 2008). For instance, in our own experiments (see Chapter 5), we saw that people tend to give away 50% even if the stakes are very high (see Figure 5.1). Thus, even though exploiting

Table 4.5 Inequity aversion in the multi-agent Nash Bargaining Game

Game	Ags.	Pun.	Sol.	Mod. LR.	Avg.	Stdev.	Maint.	Res.
NBG	3	yes	all≥15.8	yes	33.2/33.1/33.3	1.7/1.7/1.7	100%	+
NBG	4	yes	all≥11.1	yes	24.5/24.5/24.5/24.5	1.6/1.6/1.6/1.6	100%	+
NBG	10	yes	all≥4.0	yes	9.8/9.8/.../9.8	1.2/1.2/.../1.2	100%	+

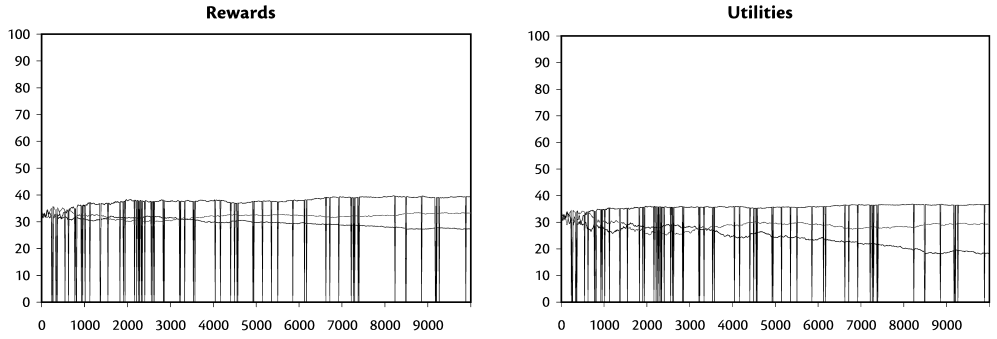


Figure 4.5 Learning to play the 3-agent Nash Bargaining Game with inequity aversion

the others is theoretically possible, the fact that the first agent does not exploit only makes it more 'human'.

Results generalize well over an increasing number of agents: with 4 and 10 agents, a valid, fair solution in which the last agent is 'almost exploited' is found and maintained every time, with the other agents achieving an equal split.

(4) The multi-agent Nash Bargaining Game

For the multi-agent NBG, we scale up the NBG to include more agents. We immediately start with CALA that include the modifications to the learning rule. Typical results with 3 agents are displayed in Figure 4.5. With 3, 4, and 10 agents for whom punishment is possible, a valid, fair solution is always found and maintained. This solution is generally close to a Pareto-optimal equal split, as can be seen in Table 4.5.

(5) The 100-agent Ultimatum Game

Results for the UG and NBG as played by 100 agents are displayed in Table 4.6. A good solution is successfully maintained in only 81% of the experimental runs with standard settings for the CALA's parameters. Since agents now each have to obtain a much smaller portion of the reward R, in particular the learning rate could be lowered to increase convergence. Indeed, with a lower learning rate and a lower σ_L (i.e., ten times lower), as indicated in the table by an asterisk, every experimental run is a success. Note that, in case of success, the

Table 4.6 Inequity aversion in the Ultimatum Game with 100 agents

Game	Ags.	Pun.	Sol.	Mod. LR.	Avg.	Stdev.	Maint.	Res.
UG	100	yes	all≥0.4	yes	0.98/0.98/.../4.4	0.4/0.4/.../3.0	81%	○
UG*	100	yes	all≥0.4	yes	0.98/0.98/.../2.2	0.3/0.4/.../1.7	100%	+

Table 4.7 Inequity aversion in the Nash Bargaining Game with 100 agents

Game	Ags.	Pun.	Sol.	Mod. LR.	Avg.	Stdev.	Maint.	Res.
NBG	100	yes	all≥0.4	yes	0.94/0.93/.../1.0	0.4/0.4/.../0.4	45%	○
NBG*	100	yes	all≥0.4	yes	0.96/0.92/.../1.0	0.3/0.4/.../0.4	100%	+

Table 4.8 Playing the multi-agent Nash Bargaining Game without punishment

Game	Ags.	Pun.	Sol.	Mod. LR.	Avg.	Stdev.	Maint.	Res.
NBG	3	no	any	yes	33.2/33.1/33.1	1.9/1.9/1.9	93%	○
NBG	4	no	any	yes	25.0/25.0/25.0/25.0	1.1/1.1/1.1/1.1	93%	○
NBG	10	no	any	yes	10.0/10.0/.../10.0	0.9/0.9/.../0.9	100%	+
NBG	100	no	any	yes	0.89/0.92/.../0.9	0.3/0.3/.../0.3	25%	○
NBG*	100	no	any	yes	0.92/0.96/.../0.9	0.3/0.3/.../0.3	100%	+

last agent receives a rather high reward, due to the fact that the other 99 agents can only approximate the optimal reward of 1. Thus, we may conclude that a multi-agent UG poses no specific difficulties for our agent architecture. A 'human' solution can always be found.

(6) The 100-agent Nash Bargaining Game

Results for the NBG as played by 100 agents are displayed in Table 4.7. A valid, fair solution is often found (i.e., a near-equal split), but not maintained in about half of the cases. Once again, this is caused by the fact that we did not adapt the learning rate of the CALA. Lowering the learning rate and σ_L by a factor 10, we can achieve success in every experiment (once again denoted by an asterisk). Thus, a multi-agent NBG can be played in a 'human' way by our agent system.

(7) The Nash Bargaining Game without punishment

As the NBG is traditionally played without any agent being able to punish, and we saw that in a 2-agent game, punishment is never used, we assessed the effects of disabling the punishment option in this game also for games with more agents.

Interestingly, the game can be often addressed if agents do not have the possibility to punish. However, when we add the possibility to punish, fair solutions are easier to be found and maintained because agents are slightly more conservative (i.e., less greedy). It is quite easy to see why this happens: an overly greedy agent is always punished, if not by other agents, then

due to the rules of the game. Therefore, regardless of the initial solution, an agent increasing his own reward too much is immediately given negative feedback. As a result, valid, fair solutions are found and maintained only slightly less often with punishment disabled than with punishment enabled. Moreover, it is interesting to note that solutions are on average closer to a Pareto-optimal solution.

Once again, with 100 agents, lowering the learning rate and σ_L by a factor 10 increases the number of experiments that were finished successfully. Thus, we see that the possibility to punish is not really necessary for CALA to learn fair 'human' solutions for the NBG, but it does increase agents' ability to learn such solutions.

4.3.5 The tragedy of the commons

We investigate the tragedy of the commons by means of the PGG. In this chapter, we restrict ourselves to a PGG with only two strategies, i.e., (1) cooperate by contributing one's entire private amount c, and (2) defect by contributing nothing. We may therefore use FALA instead of CALA.

It is easy to see that using a continuous strategy space is problematic, as agents will not be driven to satisfactory solutions in the presence of punishment. The main problem is that our proposed learning algorithms for continous strategy spaces, i.e., CALA, optimize by performing a great deal of local search. For instance, imagine an agent j currently contributing 2 and playing against an agent i that contributes 8. Agent j may try whether contributing more than 2 is a good idea, e.g., by also trying to offer 3 to i. According to the analysis in §4.2.2, agent i will punish j in *both* cases. Therefore, the essential idea underlying punishment (i.e., a reversal of the inverse relation between contribution and reward) fails to work: agent j obtains a higher reward by contributing 2 (and being punished) than by contributing 3 (and also being punished).

We will return to this issue in Chapter 6. For now, we remark that the work presented here fits in with existing work, as all existing descriptive work (or at least all the work we are aware of) in the PGG is also based on a discrete set of (two) strategies (see, e.g., Sigmund et al., 2001; Milinski et al., 2002). If we consider such a two-strategy PGG and restrict it to a two-player interaction as well, it is easy to see that the PGG essentially becomes a Prisoners' Dilemma (see §2.1.1), which also has been extensively studied (see, e.g., Aumann, 1959; Axelrod, 1984; Santos and Pacheco, 2005).

In a two-strategy game, cooperation and defection are initially equally probable. After playing a game, agents compare their rewards on a pairwise basis, and based on inequity aversion, decide whether or not to punish each other (see §4.2.2). The costs and effects of punishments are then subtracted from the agents' rewards. Finally, scaled rewards are provided as feedback to the finite-action learning automata. As indicated in §2.2.2, FALA require feedback that is scaled between 0 and 1. We divide all rewards by a sufficiently large constant to facilitate this requirement. The automata use a reward-inaction update scheme to up-

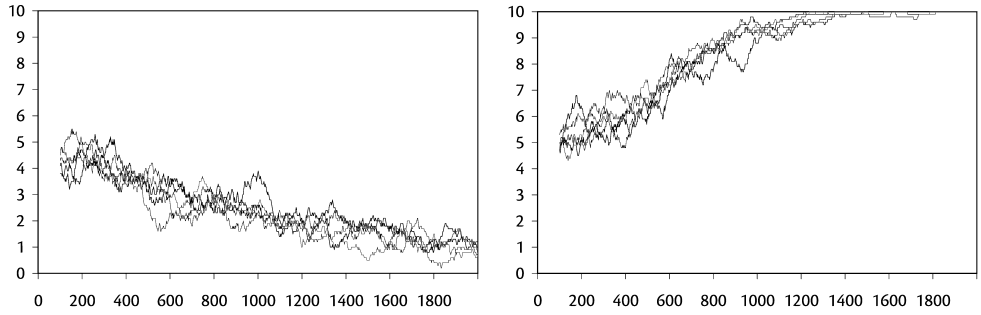


Figure 4.6 Five agents playing the discrete Public Goods Game repeatedly (for details, see text)

date the probability of choosing a certain strategy, since this facilitates convergence to an equilibrium strategy (Narendra and Thathachar, 1989).

We vary the number of agents ($n = 5$ and $n = 10$) and the parameters of the PGG (standard parameters: $r = 3$, $c = 10$, $e_p = 4$, $c_p = 1$; as well as random valid parameter settings). In all cases, we perform an experiment in which we set $\alpha = 0$ for all agents, essentially disabling inequity aversion, as well as an experiment with $\alpha = 1$, i.e., inequity aversion is enabled. In every experiment, we observe that inequity-averse agents converge to a cooperative strategy, whereas rational agents do not. Typical results for five agents and standard game settings are shown in Figure 4.6. Both graphs show a moving average of the agents' contributed amount over the last 100 games. Individually rational agents (left) slowly converge to defection, whereas inequity-averse agents (right) quickly converge to cooperation.

As is apparent from our results, inequity-averse agents find and maintain a satisfactory, profitable solution in a discretized PGG with punishment, whereas individually rational agents do not manage to maintain a joint action which leads to any profit.

4.4 Chapter conclusion

In this chapter, we worked with the first (and least complicated) of three models of human fairness, called inequity aversion. As has been indicated in §1.4, we answered (and will answer) four of our five research questions for each chapter, i.e., RQ1 and RQ3–RQ5. The answers for the current chapter read as follows.

RQ1 *How are humans using fairness in their decisions?*

The inequity-aversion model specifies that humans are willing to give up some of their reward in order to move towards more equitable, fair outcomes. This idea may be expressed using the Homo Egualis utility function. We outlined this utility function in §4.1. In §4.2, we

discussed how this utility function has been used to illustrate results obtained by humans in games such as the UG. Additionally, we use it to illustrate human punishment behavior in a multi-agent UG, the NBG, and the PGG.

RQ3 *How can human-inspired fairness be modeled computationally, taking into account the foundations of human fairness?*

In §4.3, we presented a computational model of inequity aversion and altruistic punishment, combining the Homo Egualis utility function with continuous-action as well as finite-action learning automata.

RQ4 *What are the analytical properties of the computational human-inspired fairness models developed in this research?*

As the Homo Egualis utility function is used directly in our computational model, the analytical properties of the computational model are similar to those of this utility function, as discussed in §4.2. We analyzed there that the model may be used to allow agents to learn satisfactory solutions to the UG, the NBG, and the PGG.

RQ5 *What are the (empirical) benefits of incorporating explicitly human-inspired fairness in adaptive multi-agent systems?*

In §4.3, we showed that our agents learned valid (human) solutions to the UG, the NBG, and the PGG, and that these solutions are fair according to (1) the definition of fairness presented in Chapter 3, with an error of $\epsilon = 0$, and (2) the inequity aversion model. For the PGG, we restricted ourselves to games with only two strategies.

5 Reputation and priority awareness

In this chapter, we discuss that humans are sensitive not only to differences in reward, but also to differences between them, which may be expressed as, e.g., reputation or as bargaining power. Existing descriptive models of fairness generally consider just the first of these two concepts, i.e., reputation, and aim at clarifying how reputation evolves over time, due to mechanisms such as reciprocity, image scoring, and volunteering. We discuss such existing work in §5.1. In the remainder of this chapter, we address the second concept, i.e., bargaining power (or, as we prefer to call it, *priority*), which entails that the human perception of a 'fair deal' may be immediately influenced by the (human) agent they are interacting with.

The concept of priority is most prominently present in actual bargaining interactions. Therefore, in this chapter, we study only those social dilemma games that are closely related to bargaining, i.e., the two agreement dilemmas (UG and NBG). We study human behavior in these games, both by conducting a literature survey, as well as by performing our own experiments with humans. Details follow in §5.2. Then, we aim at an expression of the concept of priority. Therefore, in §5.3 we introduce the priority awareness model, our own extension to the inequity aversion model presented in Chapter 4. We show how priority awareness may be used to explain human bargaining behavior in §5.4. In §5.5, we apply priority awareness to a multi-agent system (in a similar approach as described in §4.3). We conclude the chapter in §5.6 by addressing our research questions RQ1 and RQ3 to RQ5.

The work reported on in this chapter has partially been published in:

S. de Jong, K. Tuyls, K. Verbeeck, and N. Roos. Priority awareness: towards a computational model of human fairness for multi-agent systems. *Adaptive Agents and Multi-Agent Systems III - Lecture Notes in Artificial Intelligence*, Vol. 4865:117–128, 2008.

S. de Jong, K. Tuyls, and K. Verbeeck. Fairness in multi-agent systems. *Knowledge Engineering Review*, Vol. 23(2):153-180, 2008.

S. de Jong, R. van de Ven, and K. Tuyls. The influence of physical appearance on a fair share, *Proceedings of the Belgisch-Nederlandse AI Conferentie (BNAIC)*, pp. 105-112, 2008.

5.1 Reputation and reciprocity

The inequity aversion model presented in the previous chapter already explains a rather impressive range of human behavior in various games. However, the model misses an important element, namely that humans may respond differently to the same action being performed by a different (human or artificial) agent. Differences in response may be caused by additional information the humans received about or exchanged with this agent, either before or during their (possibly repeated) interactions.

We will provide two examples here. First, when interacting with someone familiar, humans respond more generously to people they value and trust than to people they do not (i.e., *reputation* is important). Second, someone who paid for a priority stamp expects his letter to arrive earlier than the letter of an opponent who bought a normal stamp, and someone who is asked to share money will probably offer more to an agent that looks poor than to Bill Gates (i.e., *bargaining power* influences what humans consider to be fair).

Existing research in which fairness is explicitly considered has mostly focused on the first example. The key concept here is *reciprocity* (Bowles et al., 1997; Fehr and Gaechter, 2000; Sigmund et al., 2001). Reciprocity may be direct or indirect. *Direct reciprocity* implies that a person is nice to someone else because he expects something in return from this other person, whereas *indirect reciprocity* implies that a person is nice to someone else because he expects to obtain something from a third person. It turns out that the opposite, i.e., being nasty to someone who was nasty to you (i.e., punishment), has an even greater effect on cooperation (Sigmund et al., 2001). However, being nasty may be costly, and thus, one would expect that humans only punish when they are sure to encounter the object of punishment again. This is not the case: even in one-shot interactions, humans consistently apply punishment if this is allowed. Since this is not of direct benefit to the punisher, this phenomenon is referred to as altruistic punishment (see, e.g., Yamagishi, 1986; Fehr and Gaechter, 2000, 2002), as we also discussed in Chapter 4.

Many researchers argue that altruistic punishment only pays off when the reputation of the players somehow becomes known to everyone (Milinski et al., 2002; Fehr, 2004). There are alternative proposed mechanisms, such as volunteering (Hauert et al., 2002, 2007) or fair intentions (Falk and Fischbacher, 2006). Moreover, researchers have found physical (i.e., neuronal or genetical) explanations for the fact that humans (and other social species, such as ants) are altruistic. Below, we provide an overview of existing work.

We note that many proposed mechanisms were verified by engineering them into adaptive multi-agent systems and studying the dynamics of these systems in various social dilemmas, such as the Prisoners' Dilemma, the UG and the PGG. Thus, there is no clear division in existing work between descriptive models of human fairness on the one hand, and computational models of human-inspired fairness on the other hand. However, in contrast to the work reported on in this thesis, which aims at developing computational models of fairness to obtain explicitly fair multi-agent systems, existing work developed computational models of certain proposed mechanisms purely to demonstrate that humans may use these

mechanisms as well. In other words, in existing work, descriptive models were the goal, and computational models were a means of verification. In our work, computational models are the actual goal.

5.1.1 Reputation

Various researchers (e.g., Nowak et al., 2000; Sigmund et al., 2001; Fehr, 2004; Panchanathan and Boyd, 2004; Falk and Fischbacher, 2006) argue that fairness (or, alternatively, altruistic punishment) is not possible without a representation of reputation.[1] To support this claim theoretically, the behavior of agents driven by reputation is analysed, mostly from the perspective of evolutionary game theory (Gintis, 2001).

Nowak et al. (2000) for instance, convert the UG to a 'mini-game' with only two strategies per agent (offering a high or a low reward, as well as wishing to obtain such offers). The Nash equilibrium of this game is to offer and accept a low reward. In contrast, the fair solution is to offer and accept a high reward. Using replicator equations, which describe population dynamics when successful strategies spread, it is derived that a population of agents playing this game will indeed converge to the Nash equilibrium. Next, the possibility is introduced that agents can obtain information about previous encounters, i.e., previously accepted offers. In this case, offering and accepting the low reward leads to the reputation of being easily satisfied – agents that know of this reputation will then offer the low reward.

Depending on initial conditions, the population is shown either to converge to the Nash equilibrium, or to converge to the fair solution. The same happens in the real UG, i.e., with continuous strategies instead of only two. The more games are played and the more reputation spreads, the faster the system converges to fair solutions. As Nowak et al. (2000) remark, this agrees well with other findings on the emergence of cooperation or of fairness in human bargaining behavior. Similarly, Sigmund et al. (2001) show that reputation is significantly more effective in combination with a mechanism that allows punishing those who have a bad reputation, than with rewarding those who have a good reputation.

5.1.2 Image scoring and good standing

Image scoring (Nowak and Sigmund, 1998; Wedekind and Milinski, 2000; Nowak and Sigmund, 2005) is a practical implementation of the idea of maintaining a reputation value for all individuals in a certain population. If strategies (i.e., whether to help someone) are based on image scoring, one gives help only to those whose score is above a certain threshold (Leimar and Hammerstein, 2001). In practice, an individual's score increases on every occasion he donates aid to a recipient and decreases when there is an opportunity to help someone in need but no help is offered. Analysis and computer simulations show that image scoring may indeed lead to cooperation and reciprocal altruism. In addition, Lotem et al.

[1] Illustrative real-world examples of reputation being used to enforce fairness may indeed be found in many online stores and auction sites (Dellarocas, 2003). Indeed, the benefit of using reputation in negotiation and auction settings is extensively studied (e.g., Huynh et al., 2006a; Reece et al., 2007).

(1999) show that cooperation may actually be increased by adding a small fraction of agents to the population that is physically unable to display cooperative behavior.

Whether image scoring indeed forms a satisfactory implementation of reputation, is debatable. For instance, Leimar and Hammerstein (2001) argue that analytical arguments show that it would not be in an individual's interest to use image-scoring strategies. Image scoring only works in case of strong genetic drift or a small cost of actually being altruistic. As an alternative, they propose the strategy of aiming for *good standing*, which is inspired by Sugden (1986) and demonstrated to be (potentially) evolutionarily stable as well as dominant over image scoring. In contrast to the image-scoring model, the good-standing model initially attributes a high value to all individuals. Individuals may lose standing if they refuse to help others that have a good standing.

5.1.3 Volunteering

An alternative answer to the question why altruistic punishment emerges, is the concept of *volunteering* (Hauert et al., 2002, 2007). For instance, if we see the PGG in analogy with hunting a mammoth (as in §2.3), we may assume that the hunters have volunteered to participate in the hunt. They could also have chosen to stay home safely, or to collect mushrooms. Obviously, the people that collect mushrooms experience a lower food quality (in terms of energy value) than the hunters, but they take this for granted, since picking mushrooms is also less risky than trying to catch a mammoth. Moreover, since the hunters have volunteered to participate, they may expect that all hunters will cooperate, minimizing the risk.

To model this, an alternative third strategy can be introduced (we remind the reader that the first two are 'defect' and 'cooperate', but there is no punishment involved here). This strategy could be labeled 'refuse'. Refusers obtain a reward $0 < s << (r-1)c$. Thus, the reward is higher than what agents obtain when everyone defects (i.e., 0), but at the same time, it is (much) lower than what they obtain when all agents cooperate (i.e., at most $(r-1)c$). Thus, refusers can invade a population of defectors. Then, since s is lower than the reward that would be obtained if everyone cooperates, rare cooperators can invade a population of refusers. Finally, defectors can once again invade the cooperators. These strategies oscillate endlessly (Hauert et al., 2002).

Boyd and Mathew (2007) show that adding punishment here, i.e., a fourth strategy 'punish' (which implies cooperation), makes it possible for punishers to invade the oscillating mixture of cooperators, defectors, and refusers, and once they do they tend to persist. As Boyd and Mathew (2007) note, this means that the population spends most of the time in a happy state in which cooperation and punishment of defectors predominate. Additionally, Hauert et al. (2002) and Nowak et al. (2004) report on interesting results in finite populations. More precisely, if agents do not change their strategies, there will be four absorbing states, representing the four pure strategies. With small mutation rates, a stationary distribution is obtained. With a larger selection strength, punishing is the dominant strategy, which does

not happen if the refusers are removed (in that case, defection is the dominant strategy). This is an interesting example of the problem of irrelevant alternatives.[2]

5.1.4 Intentions

Falk and Fischbacher (2006) present an alternative theory of reciprocity, based on the concept of fair or kind *intentions*. In this theory, a reciprocal (i.e., fair) action is modeled as the behavioral response to an action that is perceived as either kind or unkind.

The two central aspects of the model are the consequences of the action and the actor's underlying intentions. Several experimental studies suggest that fair intentions play a major role for the perception of kindness. For instance, human second players in the UG tend to punish more if the reward is offered willingly by the first player, than if this player is forced to offer a randomly generated reward (e.g., by performing a dice roll). Similarly, if the first player can only choose between the strategies 'give away 80%' or 'give away 20%', offers of 20% are rejected much less often. Inequity aversion alone is not able to explain this, but the kindness theory is. The theory is applied to various other games, and is shown to predict human behavior well (Falk and Fischbacher, 2006).

5.1.5 Physical explanations

In humans, we see that there are many moral and justicial pressure devices. Humans may take the law in their own hands in the absence of those devices. Recent studies show a physical basis for this behavior, i.e., our tendency to be socially aware seems to be explicitly encoded in our neurons. For instance, Sanfey et al. (2003) studied the brain activity of humans playing UGs. They perceived strong activity in brain areas related to both cognition and emotion. Moreover, Knoch et al. (2006) shows that the right prefrontal cortex plays an important role in fair behavior. Disrupting this brain area using (harmless) magnetic stimulation made subjects significantly more selfish.

Apart from a neuronal explanation, researchers also found a genetical explanation for reciprocal fairness. Although genes are commonly regarded as being selfish (i.e., purely focused on successful replication; Dawkins, 1976), a gene that promotes altruism might experience an increase in frequency if the altruism is primarily directed at other (potentially unrelated) individuals who share the same gene (Hamilton, 1964). Obviously, for this kin selection to work in practice, the presence of the gene must be somehow perceivable and recognized. Imagining that altruistic individuals might for instance have a green beard, Dawkins coined the term *green-beard effect*. Interestingly, in analogy to cooperators being vulnerable to exploitation by defectors, green-beard genes are vulnerable to mutant genes arising that produce the perceptible trait (i.e., the green beard) without the associated altruistic behavior.

[2] The problem of irrelevant alternatives is described in many humorous anecdotes, such as the one in which a man in a restaurant is trying to choose between the two dishes of the day (say, souvlaki and spaghetti). He ultimately selects the spaghetti, but is then told by the waitress that there will also be curry today. The man responds by saying: "This changes everything! I'd like souvlaki please."

After being hypothetical for over 20 years, the first green-beard gene was indeed found in nature, i.e., in a type of ant (Keller and Ross, 1998). In humans, similar (groups of) genes may physically control our urge to perform altruistic punishment. These genes apparently have had an evolutionary advantage. Dawkins (2006) provides an interesting discussion.

5.2 Evidence for the human concept of priority

In this section, we discuss experiments performed with humans. These experiments indicate that humans do not only care about reputation that becomes established over time, but also use the concept of *immediate reputation* or *priority*. Due to additional information and/or the activation of stereotypes, humans may respond differently to different opponents under otherwise equal circumstances. Below, we first briefly re-discuss a simple illustration of this phenomenon, i.e., the fruit shop example from §1.3. Then, we present two experiments that were conducted in the UG; in the first, the participants received explicit information on their opponents, and in the second, the information was highly implicit. In the latter case, we show how we may still be able to elicit human decision-making.

5.2.1 Initial experiment: the fruit shop

In §1.3, we discussed a small example task that we gave to a group of human participants, in order to find out how they would 'fairly' address this task. We mention this task only briefly here, as it has been explained in detail in §1.3.

When we give our participants the information that they need to divide a limited resource (i.e., time spent by a service robot fetching fruit) over two groups of customers, with one group being larger, they favor this larger group over the smaller one, while still aiming to keep a low level of inequity between the groups. Thus, the larger group is seen as *slightly more important*. Rationally, we would either consider both groups equally important (given, e.g., egalitarian social welfare, see §3.3.1), or we would strongly favor the largest group (given, e.g., utilitarian social welfare). Using inequity aversion, we would reach a similar outcome (considering both groups equally important). Clearly, an element of human decision-making is missing here.

5.2.2 An Ultimatum Game with variable amounts and wealth

After some initial experiments (such as the one described above), we created a larger, more structured experiment, following the example of many experiments with human fairness in the two-player UG (e.g., Fehr and Schmidt, 1999; Oosterbeek et al., 2004). We asked students and staff members of three faculties of Maastricht University to participate in a survey concerning the UG, which was developed in cooperation with an experimental psychologist.[3] Respondents were asked various control questions and a number of UG dilemmas,

[3] We kindly acknowledge dr. Anton de Vries for his assistance in this matter.

in which they had to indicate how much they would offer as a first player. In the end, 170 surveys were submitted, of which 160 were usable. Of the 160 respondents, 38 were familiar with the UG; the remaining 122 were not.

To introduce the notion of priorities explicitly in the UG, we varied two quantities. First, after playing some standard UGs, participants were told that the other player was ten times poorer or ten times wealthier than they were. In this case, people could either be fair to poorer people (i.e., give them more) or exploit poorer people (i.e., give them less, because they will accept anyway). Second, the (hypothetical) amount of money that had to be divided varied between €10, €1,00 and €100,000, to determine whether this had any effect on people's attitude with respect to poorer, equal or richer people.

Figure 5.1 gives an overview of the results. Below, we discuss three observations that may be inferred from this figure.

Hypothetical money. Since we were asking people to imagine that they had to divide money, instead of giving them real money to divide, we first needed to assess whether this difference had an important impact on observed offering behavior. In our survey, this was not the case, as the behavior we found was in line with behavior found earlier (for an overview, we refer to Bearden, 2001; Oosterbeek et al., 2004). For instance, most people were willing to give away 50%, and some people offered less. We note that there is little research published that compares human behavior in case of real money with behavior in case of hypothetical money. Cameron (1999) has established that there is no statistical difference in offering behavior, and that responders are more tolerant when real money is involved, i.e., responders accept lower offers of real money.

Increasing the amount at stake. With an increasing amount, we see that the first player keeps more to himself. This may seem rather obvious; after all, it is much more difficult for the second player to refuse 10% of €100,000 than 10% of €10. Interestingly, however, results of existing research are inconclusive with respect to this matter. Bearden (2001) provides an overview of experiments concerning the UG, including experiments in which the amount at stake varied. When we analyse research surveyed by Bearden (2001), as well as other recent research, we see opposite conclusions. For instance, Roth et al. (1991), Straub and Murninghan (1995), and Cameron (1999) find that a varying amount at stake has no influence on the amounts offered to the second player, whereas Sonnegard (1996), Slonim and Roth (1998), and Zollman (2008) find a substantial influence. We note that experiments with high amounts at stake were mostly performed in relatively poor countries (e.g., Cameron, 1999; Slonim and Roth, 1998), simply because research institutes cannot afford to let people play games worth €100,000 in Europe or other Western countries. Cultural traits of the people involved may explain why 'high-stake' games have such a fair outcome in general.

Richer or poorer opponents. Players' behavior in the normal UG can already be successfully explained using the Homo Egualis utility function, as outlined in Chapter 4. However, the results of our survey show that priorities, which are not explicitly modeled using this utility

function, strongly matter to human players of the UG. Before confronting participants with UGs in which agents had unequal wealth, we asked them whether they had thus far assumed that the other player was poorer, wealthier, or equally wealthy. Of the participants, 92% assumed that the other player was equally wealthy; the remaining 8% was almost equally divided between the other two options. Subsequently, the participants were confronted with games in which the other player was ten times more or ten times less wealthy than they were. Results indicate that people are actually increasingly fair to poorer people, and expect the same in return from richer people. Poorer opponents were given substantially more money than equal opponents, richer opponents were given substantially less. This indicates that information concerning relative wealth matters to humans.

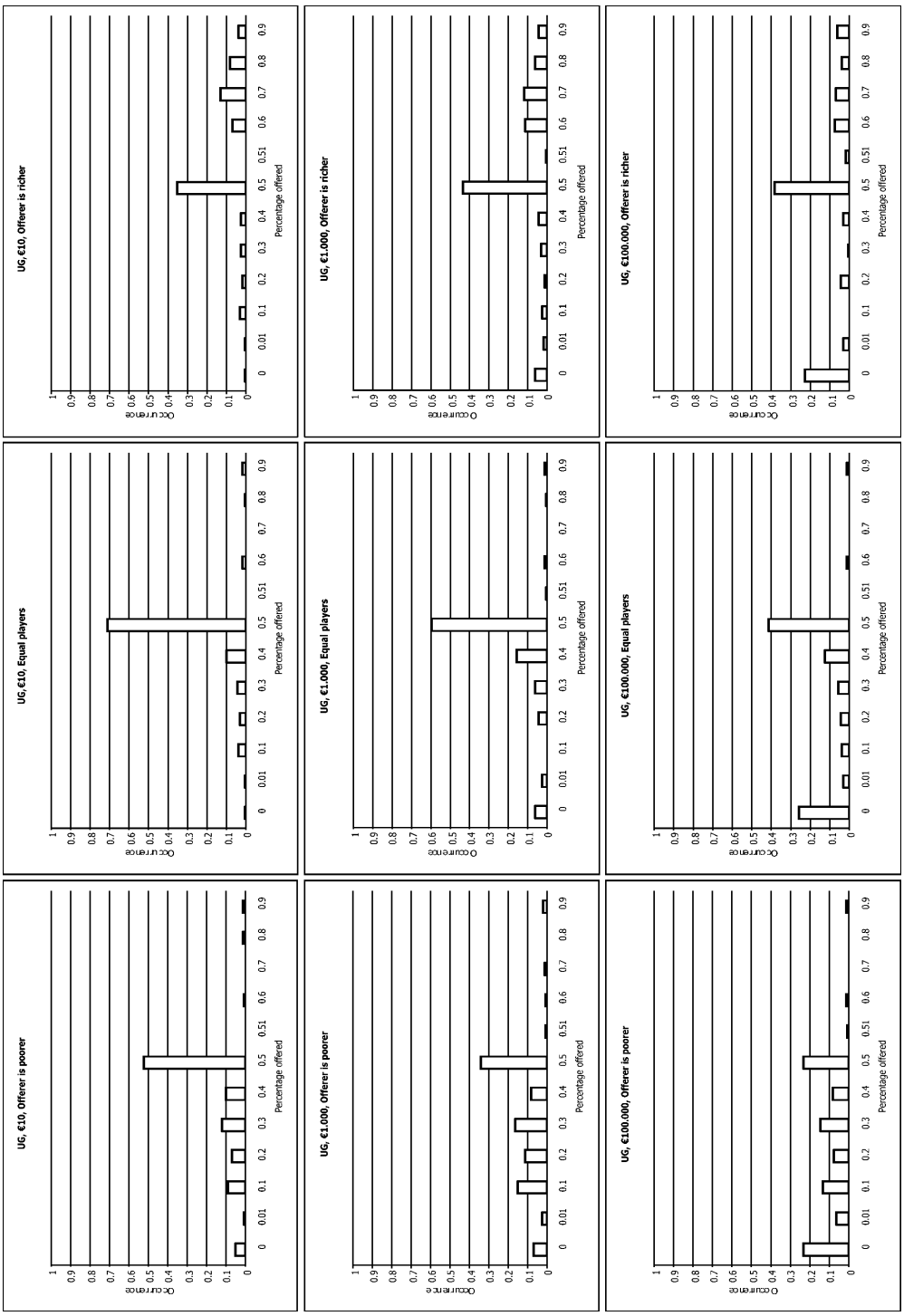
5.2.3 The impact of visual appearance in the Ultimatum Game

As we have discussed above, the human concept of a fair share is influenced by efficiency considerations (i.e., individual rationality), but also by factors such as personal preferences (i.e., some people are more greedy than others) (Fehr and Schmidt, 1999; Gintis, 2001), additional information (i.e., if we know our opponent to be poor, we may be more willing to share, see above), and previous experience (i.e., if we played against our opponent before, we may expect him to behave in a similar manner this time) (Bowles et al., 1997; Milinski et al., 2002; Fehr, 2004). Eliciting such factors is a challenging task, since very often, a carefully designed experiment needs to be conducted with many participants that should not (yet) be acquainted. Often, researchers need to motivate participants by giving them a monetary reward that is somehow related to their performance in the experiment at hand. In the end, obtained data may still be biased or difficult to interpret.

Human computation (Von Ahn, 2006) provides an interesting alternative to expensive laboratory studies and cumbersome analysis. Originally, it was proposed as an alternative to developing complicated algorithms. The task at hand (e.g., tagging images to elicit the objects present in them) is usually presented as a game to attract voluntary participants. Each participant then has to perform a small, carefully selected subtask that is simple for humans, yet complicated for computers (e.g., labeling the objects seen in a few images). The results of the performed subtasks are then aggregated (e.g., into a database with suitable tags for many images).

In this subsection, we show that the usefulness of human computation is not restricted to providing alternatives for complicated algorithms; it may also be successfully applied to elicit implicit factors that influence human decision-making. More precisely, we investigate whether physical appearance triggers different behavior with respect to a fair share, by letting our participants play the UG. As in the previous subsection, the participants in our survey obtain additional information concerning their (fictive) opponents in the UG, but

Figure 5.1 (following page) Offers done by the first agent in the Ultimatum Game, depending on the amount at stake and the relative wealth of the other agent. Offers are binned in 12 bins, of which the *lower bound* is displayed.


UG, €10, Offerer is poorer
UG, €10, Equal players
UG, €10, Offerer is richer
UG, €1.000, Offerer is poorer
UG, €1.000, Equal players
UG, €1.000, Offerer is richer
UG, €100.000, Offerer is poorer
UG, €100.000, Equal players
UG, €100.000, Offerer is richer
Occurrence
Percentage offered
0 0.01 0.1 0.2 0.3 0.4 0.5 0.51 0.6 0.7 0.8 0.9
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

this time the information is *implicit* and represented by means of a photograph. The photos may trigger previously stored stereotypes (Devine, 1989; Wheeler and Petty, 2001) concerning, e.g., gender, race, age, and (apparent) wealth. Such stereotypes are indeed inherently implicit and therefore difficult to make explicit.

Using our survey, we aim at eliciting implicit information such as stereotypes. To this end, all participants have to answer five similar questions. In each of the five questions, we present them with two opponents that are selected in a way that ensures the most (expected) information gain (Janssens, 2008). We then ask which of the two would receive more money (where it is also possible to give the same amount to both).

Clearly, such qualitative questions concerning ordinal preferences pose a simpler task to the participants than having to answer many quantitative questions related to cardinal preferences, such as "how much would you give to this person?". Nonetheless (or actually for this reason), in addition to the five identical questions related to ordinal preferences, we also ask one quantitative question. The quantitative question allows us to determine whether our participants' answers align with results reported in literature (e.g., Oosterbeek et al., 2004; de Jong et al., 2008).

Using the answers obtained from our participants, we construct two rankings of our set of photos (i.e., a qualitative ranking and a quantitative ranking), from the least-earning person to the most-earning person. For the qualitative ranking, this requires additional computation to translate participants' preferences concerning pairs of photos to a global ranking concerning the whole set of photos. Recent research has proposed a method called CollaboRank to perform this task (Janssens, 2008).

Thus, our contribution in this subsection is two-fold. First, we provide additional support for the fact that humans actively use priorities when they are deciding upon sharing with others. Information that leads to priorities may be explicit, but may also be implicit. Especially in the latter case, eliciting information is a difficult task that typically requires many well-controlled experiments. This leads to the second contribution: we show how human computation methods, such as collaborative ranking, may be used to elicit the information on which human decisions are based.

The remainder of this subsection is structured as follows. First, we briefly look at the human-computation algorithms applied here. Second, we present our methodology, which consists of a survey and an analysis of the data provided by our participants. Third, we present the results of our analysis, aimed at assessing the influence of appearance on human behavior.

Human computation and ranking

The CollaboRank method (Janssens, 2008) is inspired by von Ahn (2006), who discusses the power of human computation. The method has been proposed to perform image-based ranking tasks, e.g., for sorting a set of images with respect to the size of the objects presented in these images. Humans are more capable of recognizing (sizes of) objects in images than

computers, and therefore also better at ranking those images. The method may also be used to elicit other (implicit) information in the images, or even to perform ranking tasks that are not related to images at all.

The method enables multiple participants to rank a large set of images in a collaborative, iterative way. To this end, it distributes relatively simple ranking tasks, concerning only a small subset of the set of images that need to be ranked, effectively over the participants. The subsets' rankings are aggregated to a global ranking concerning the entire set. Through an iterative process the global ranking should converge to a stable, high quality ranking, unless there is absolutely no consensus concerning the rankings of subsets.

The key element of the CollaboRank method is the Global Preference Matrix (GPM). Below, we first explain this matrix. Next, we discuss how the matrix is used to create tasks concerning a subset of images, and how results concerning such a subset are integrated in a global ranking.

The Global Preference Matrix (GPM). The GPM fullfills two tasks: first, it forms the basis for formulating the tasks given to participants, and second, it aggregates the personal rankings that the participants have submitted into a global ranking $\mathcal{R}$. The preferences of all the possible combinations of image pairs $\langle i, j \rangle$ for which $i, j \in I'$ (the set of images) are stored in the GPM. The preference relation, as represented by the GPM, on image set I' of size ϕ, is represented by a $\phi \times \phi$ preference matrix $P = (p_{ij})$. A preference $p_{ij} = \frac{1}{2}$ indicates an equal preference for image i and image j (i.e., $i \sim j$). $p_{ij} > \frac{1}{2}$ indicates that i is preferred to j (i.e., $i \succ j$), while $p_{ij} < \frac{1}{2}$ indicates the reverse. In other words, the pair-wise preference p_{ij} is interpreted as the probability of image i being ranked before image j. A property that should permanently hold is $\forall i, j \in P, p_{ij} = 1 - p_{ji}$.

Formulating a task. The formulation of a task for each participant is the first step of the CollaboRank method. Because each participant only ranks a small subset of the images, the algorithm should intelligently select the image subsets (in our case, pairs) that will be given to each participants. This minimizes the number of participants needed to arrive at a sufficient global ranking. CollaboRank uses an entropy function, which ensures that image pairs with the highest level of uncertainty are picked first to be ranked. The uncertainty of the rank probability p_{ij} is calculated using the binary entropy function (MacKay, 2003)

$$H_2(p_{ij}) = -p_{ij} \log_2 p_{ij} - p_{ji} \log_2 p_{ji}. \tag{5.1}$$

The total entropy of an image i is defined as follows:

$$H(i) = \sum_{j \in I : j \neq i} H_2\left(p_{ij}\right). \tag{5.2}$$

This function indicates the uncertainty of the ranked position of an image i. CollaboRank selects an image m having the highest entropy $H(m)$. Then, it selects an image k which has the highest entropy as a pair with image m, i.e., $H_2(p_{km})$. The two images m and k are shown

as a pair to a participant. Note that this method leads to all possible image pairs being ranked initially: unranked pairs have a positive entropy, whereas pairs that are ranked by a single participant have an entropy $H = 0$. After all pairs have been ranked for the first time, the algorithm selects pairs for a second ranking.

After a participant has submitted his personal ranking, it is aggregated into the GPM. The preferences are updated of those pairs of images in the GPM that were presented to the participants. In this way, the GPM is filled with values that are closer to either 0 or 1, eventually creating a global ranking. More precisely, a personal ranking is aggregated into the GPM by assigning the average preference of images i and j, with regard to all the submitted personal rankings, to p_{ij}. For example, when 8 out of 10 participants submit $i \succ j$, then $p_{ij} = 0.8$ and $p_{ji} = 0.2$.

Producing a global ranking. The Greedy-Order algorithm (Cohen et al., 1999) is being used to extract a global ranking from the GPM. To this end, the GPM is interpreted as a directed weighted graph, where initially, the set of vertices V is equal to the set of images I', and each edge $u \to v$ has weight p_{uv}. Each vertex $v \in V$ is assigned a potential value $\pi(v)$, which is the weighted sum of the outgoing edges minus the weighted sum of the ingoing edges:

$$\pi(v) = \sum_{u \in V} p_{vu} - \sum_{u \in V} p_{uv}. \tag{5.3}$$

The vertex t with the maximum potential is selected. The corresponding image is assigned the rank $R(t) = |V|$, which makes it the first image in the ranking. The vertex and its edges are deleted from the graph, after which the potential value π of the remaining vertices are updated. This process is repeated until the graph is empty. The last vertex removed from the graph has rank 1, and will be the last image in the ranking.

Methodology

In order to determine the effect of appearance on the human concept of a fair share, we developed an online survey in which participants are confronted with a small number of opponents, each represented by a photo. The pair shown in each question of this survey is generated by the CollaboRank algorithm in such a way that the information gain per submitted survey is largest, as described above. Ranking is used to aggregate individual participants' answers (concerning this small number of opponents) into a global result (concerning the whole set of photos used). In this section, we take a closer look at the survey.

Control questions. The survey starts with a small number of 'control questions' that were asked to determine whether our participants have strategies similar to those reported in earlier work (see, e.g., Oosterbeek et al., 2004; de Jong et al., 2008). Thus, we asked how much money participants would offer to an unknown opponent. We also varied the amount of money at stake between €10 and €10,000, to verify our earlier results (see above).

You have to divide €10 between you and an opponent. If the opponent accepts your offer, you can keep the money according to this offer. If the opponent rejects the offer, you both receive nothing.

Who would you offer more money?

Opponent A.

Opponent B.

( ) Opponent A.
( ) Opponent B.
( ) I would offer them an equal amount of money.

Figure 5.2 A survey question aimed at deriving a global ranking from pairwise comparisons

Pairwise ranking. The main body of survey questions (i.e., five questions per participant) concern pairwise comparison of two opponents, which are pulled from a set of 24 photographs. One of the 252 possible photograph pairs (and the associated question) is given in Figure 5.2. For the selection of photographs, we used a basis of five properties: gender, age, race, attractiveness, and wealth. We used Google's image search to find many photographs that displayed a large diversity in at least one of these properties. The resulting large set of photographs was then manually reduced to 24, with the aim of keeping sufficient diversity with respect to all properties.

Due to the CollaboRank method, pairs are not presented randomly, but in a way that provides optimal information. For the participants, pairs of photographs should therefore become increasingly difficult to judge, that is, after every pair has been presented to a participant at least once. Since every participant ranks five pairs out of the 252 possible ones, we need 51 participants to have every pair ranked at least once. From the 51^{st} participant onwards, we expect increasingly more participants to select the last answer option, i.e., 'I would offer them an equal amount of money.'

Quality of the ranking. To address the quality of the pairwise ranking, we also ask partici-

pants to quantify how much would be offered to one certain opponent, with a photograph being pulled from the set of 24 photographs randomly (i.e., the photograph shown to a participant for this question may or may not be shown in the pairwise questions). This is clearly a much harder question than having to choose between two opponents. We can create a quantitative ranking by simply sorting the photographs increasingly on the average amount offered. If the pairwise and quantitative rankings are sufficiently similar, this tells us that human preferences may indeed be elicited by (simple) qualitative questions instead of (more difficult) quantitative questions. To measure similarity, we use the Kendall rank correlation coefficicient τ (Kendall, 1938) to determine the similarity of the rankings. A value of $\tau = +1$ indicates that the two rankings are equivalent; $\tau = -1$ indicates that they are reversed, and $\tau = 0$ indicates they are completely uncorrelated.

Results

Below, we discuss results, obtained from 173 participants who were attracted through promotion via various social websites. First, we look at the control questions, and then, we discuss the various rankings. We conclude by providing overall results.

Control questions. As mentioned above, the survey starts with a small number of questions aimed at determining whether our participants have strategies similar to those reported in earlier work. Being confronted with random, invisible opponents, most participants offer as well as accept 50% of an amount of €10 , as also reported in literature (Oosterbeek et al., 2004; De Jong et al., 2008). Increasing the amount of money to €10,000, offering as well as accepting 50% is still the most often-chosen strategy (i.e., it is chosen by roughly 30% of the participants), but a majority of participants would offer somewhat less money, and would accept substantially less (15% already accept only 10% of the amount, i.e., €1,000). This conforms to what we found in our earlier survey (see §5.2.2).

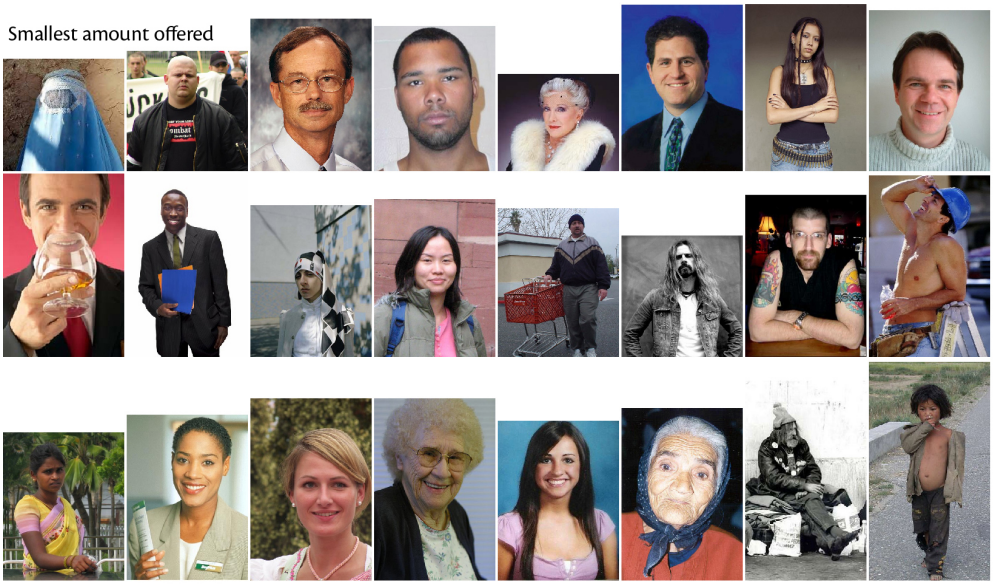
Pairwise ranking. The result of aggregating the 173 pairwise rankings, as submitted by our participants, into a global ranking of the 24 photographs, is given in Figure 5.3(a). While the survey was in progress, we noticed that the photograph pairs presented to the participants became increasingly difficult to distinguish, i.e., the option 'I would offer them an equal amount of money' was chosen increasingly often. Moreover, since 173 participants completed the survey, each performing 5 pairwise rankings, we obtained 865 pairwise rankings. There are 252 different photograph pairs; therefore, if pairs were selected at random, every pair would be ranked a little more than three times. However, due to CollaboRank, we were able to ensure that difficult pairs were ranked more often than easy pairs; some pairs were ranked by more than 15 participants.

Looking at the ranking produced, we see a number of interesting phenomena. For instance, wealth is clearly important: the photographs that seemingly depict poor people are almost all given relatively large amounts (and the reverse). This corresponds to what we found in our earlier survey when we explicitly mentioned the wealth of opponents (see above). Discarding the most-receiving (poorest) people, there also seems to be a rather clear division

(a) Relative ranking

The photographs are presented in increasing order of amount of money offered to the people represented, with the increase going left-to-right and then top-to-bottom.

Smallest amount offered



Largest amount offered

(b) Absolute ranking

We display the absolute rank for the photographs above. Below this rank, we show the average amount given, and the standard deviation on this amount (only for photos that are rated by 5 or more people).

1	8	16	2	12	3	4	14
2.2(1.6)	3.7(1.2)	4.6(0.5)	2.7(1.1)	4.2(0.7)	3.2(1.1)	3.3(—)	4.5(0.7)

5	6	7	18	24	11	15	19
3.4(1.4)	3.5(1.8)	3.6(1.4)	4.7(1.9)	5.6(2.2)	4.0(0.9)	4.5(—)	4.8(1.7)

20	10	9	21	13	17	22	23
4.8(2.0)	3.9(0.6)	3.8(1.2)	5.0(—)	4.2(0.9)	4.7(3.1)	5.0(2.0)	5.0(2.1)

Figure 5.3 Relative and absolute ranking of 24 visible opponents

concerning gender; all remaining photographs on the bottom row (i.e., the most-receiving people) are women, whereas 7 out of the 10 least-receiving people are men. Concerning attractiveness, we can at least note that facial expression seems important – the least-receiving person has no visible facial expression, and many of the people that receive little money are not smiling. Race does not seem to be a decisive factor for our participants, as the various skin colors are not clearly grouped. The same goes for age, if we discard people that are stereotypically wealthy or poor.

Quality of the ranking. Since we asked all 173 participants one quantitative question, we may expect every one of the 24 photographs to be quantitatively ranked an expected $173/24 = 7.2$ times. Due to this low number, many of the photographs displayed no significant difference concerning quantitative information. Other photographs yield clear results, as displayed in Figure 5.3(b). For instance, the least-receiving photo in the qualitative ranking also receives the least in the quantitative ranking (i.e., 2.2 ± 1.6 out of 10). Two of the three most-receiving photographs in the qualitative ranking are to be found in the top-three of the quantitative ranking, with offers as high as 5.6 ± 2.2 out of 10, i.e., significantly more than the least-receiving photograph and also more than 50% of the total amount. This would not be possible if subjects were only motivated by inequity aversion (Fehr and Schmidt, 1999; De Jong et al., 2008). Comparing the two rankings analytically, we observe that the Kendall rank correlation coefficient $\tau = 0.51$, which indicates a significant correspondance. Thus, humans display roughly the same preferences, based on physical appearance stereotypes, in both a pairwise comparison as well as in more difficult quantitative questions.

Overall result. As an overall result, we clearly see that physical appearance matters strongly, as it triggers known stereotypes such as 'poor' or 'rich', but potentially many other stereotypes, some of which have been displayed here, i.e., most notably gender, but also attractiveness (facial expression). Race and age seem less important in this survey.

5.3 The priority-awareness model

In the previous section, we observed that people's concept of a fair deal depends on additional information that they may have about themselves and other people. In real-world examples, the nature of this information may vary. Examples include wealth, probabilities of people belonging to a certain group, or priorities involved in the task at hand. In our work, we assume that the additional information can be expressed with one value per agent, i.e., the *priority*. Henceforth, we assume that the priority values are true and are known by all agents.

Given the experiments above, we see that priorities indeed matter to human agents. This is not sufficiently described by the Homo Egualis utility function of Chapter 4. Most notably, there are two problems. First, as clearly indicated by Fehr and Schmidt (1999, p. 850), the Homo Egualis actor does not model the phenomenon that people actually like to be better than other people. In priority problems, people accept and encourage inequity to a certain extent: agents with a high priority should be allowed to obtain a higher reward than agents with a low priority. Priorities would have to be encoded in an indirect way using the utility function's parameters (most notably α^i), which then would have to be adapted for every possible other agent. Second, the significant number of participants who offer more than half of the amount at stake to, e.g., a poor second agent in prioritized UGs, is not described by the Homo Egualis utility function at all. The priority-aware model, which is also based on a utility function, captures these phenomena in a straightforward way.

We model people's perception of fairness in the presence of additional information using a descriptive model of human fairness, called *priority awareness*. In this model, every agent has a *priority*, and attaches an individual *weight* $w^i \in [0,1]$ to its priority. With $w^i = 0$, agent i does not pay attention to its priority, and with $w^i = 1$, the priority matters strongly. People usually are rather tolerant, i.e., their boundary between verdicts such as 'acceptable' and 'not acceptable' is not immediate. For instance, if someone wishes to receive 50% of the reward, he will probably also accept 49.5%, less probably accept 49%, et cetera. To model this phenomenon, we introduce a weight interval: the value of w^i does not have to be specified exactly, but may be in a range: $w^i \in \left[w^{i^{\min}}, w^{i^{\max}}\right]$. An alternative approach here would be to introduce a probability of acceptance that decreases with a decreasing reward. This approach is further developed in Chapter 6.

Since the inequity-averse Homo-Egualis utility function already describes human behavior in a variety of games in which priorities are not present, we propose an extension to this function for games that do include priorities. More precisely, we introduce the notion of a *perceived reward* $\hat{r}^i$, which is defined as:

$$\hat{r}^i = \frac{w^i}{p^i} r^i + (1 - w^i) r^i. \tag{5.4}$$

Thus, if agent i has a high priority (and $w^i \neq 0$), he has a lower perceived reward $\hat{r}^i$ than the reward r^i he is actually receiving. The bounds on the parameter w^i prevent agents from having an irrealistic perceived reward.

The perceived rewards can be used in the Homo Egualis utility function to determine whether a certain reward distribution $r = \left(r^1, \ldots, r^n\right)$ is acceptable, given the agents' priority and weight parameters, i.e.:

$$u^i(r) = \hat{r}^i - \frac{\alpha^i}{n-1} \max\left\{\hat{r}^j - \hat{r}^i, 0\right\} - \frac{\beta^i}{n-1} \max\left\{\hat{r}^i - \hat{r}^j, 0\right\}. \tag{5.5}$$

We note that this means that agents may need to know or estimate each others' perceived rewards $\hat{r}^i$, in addition to the parameters α^i and β^i. However, with various cryptographic techniques it is possible to allow agents to calculate their utilities without explicitly knowing the values of priorities, weights, and rewards of the other agents (Denning, 1982).

The concept of equity (or inequity) is now viewed from the perspective of a transformed perception of the rewards. Thus, for instance, two agents with $p^1 = 1$, $p^2 = 2$, and $w^1 = w^2 = 1$ perceive an equal split in case agent 2 receives an actual reward that is twice that of agent 1. In an interaction between n agents, a certain reward distribution is acceptable if every agent i can assign a weight w^i from the range $\left[w^{i^{\min}}, w^{i^{\max}}\right]$ to his priority p^i, in such a way that he experiences a positive utility. Thus:

$$\forall i \exists w^i \in \left[w^{i^{\min}}, w^{i^{\max}}\right] : u^i > 0 \rightarrow \text{accept}. \tag{5.6}$$

Otherwise, one or more agents have reason to reject. In this case, the agent with the lowest perceived reward is the first to reject. This is not necessarily the same agent as the agent with the lowest actual reward.

5.4 Explaining human behavior

Unlike the inequity-averse model, the priority-aware model can be used to explain human behavior in the various experiments we performed. We will discuss the experiments below.

5.4.1 The fruit shop

In the fruit shop experiment (discussed in §1.3, also see Figure 1.1), we determined the human response to a very simple priority problem, in which customers of a (rather impractical) shop can be divided into two groups. Each group would like to buy a different item, located at A and B in the shop. A robot that has to fetch items, has to be placed somewhere between A and B such that the delays experienced by the customer groups are balanced. We model each of the groups as a single agent (named after the locations, i.e., agent **A** and **B**), and assign to the two agents a priority value p^i equal to the probability that a customer belongs to their associated group. Let us assume for easier calculation that the distance between locations A and B is 2 and that the customers wait precisely in the middle between these locations. Thus, if the robot is positioned at $a \in [0, 2]$, customers represented by agent **A** will experience a delay $d^A = a + 1$ before the robot delivers the requested item. Customers represented by agent **B** will experience a delay of $d^B = 2-a+1 = 3-a$. Since waiting time can be seen as negative reward, we define the rewards of the agents as $r^A = 3 - a$ and $r^B = a + 1$. Thus, **A** obtains a higher reward with lower values of a, and the other way around for **B**. The goal of the experiment is now to find a position a that satisfies both agents **A** and **B**, i.e., a fair position. More precisely, since both agents should be treated equally well, we would like both agents to have the same utility, i.e., $u^A = u^B$.

In case the group of customers wishing to obtain the item at A is equal in size to the group of customers wishing to obtain the item at B, we obtain $p^A = p^B = 0.5$. With equal priorities for all agents, the priority-aware model is identical to the inequity-averse one, i.e., we can set $\hat{r}^A = r^A, \hat{r}^B = r^B$. Let us assume that we choose an $a \leq 1$ (for $a > 1$, everything can simply be inverted). In this case, we obtain $\hat{r}^A \geq \hat{r}^B$. Now we can calculate the utility values as

$$u^A = \hat{r}^A - \beta^A \left(\hat{r}^A - \hat{r}^B\right) = 3 - a - \beta^A (2 - 2a),$$

$$u^B = \hat{r}^B - \alpha^B \left(\hat{r}^A - \hat{r}^B\right) = 1 + a - \alpha^B (2 - 2a).$$

As has been mentioned above, the fair position satisfies $u^A = u^B$. This is only possible if the robot is placed in the middle, i.e., $a = 1$ (or if, coincidentally and ignorably, $\beta^A + \alpha^B = 1$). This result corresponds exactly to human behavior.

For a situation in which the priorities differ, priority awareness is no longer identical to inequity aversion. The latter model would predict the same outcome as with equal priorities, i.e., $a = 1$. However, with priority awareness, we can use the same analysis as above: we simply have to find the fair position for which $u^A = u^B$. For instance, assume that there is a 60% chance for customers to request the item at A. Thus, we set the priorities to $p^A = 0.6$ and $p^B = 0.4$. Now, for $w^A = w^B = 1$, the perceived rewards are $\hat{r}^A = \frac{3-a}{0.6}$ and $\hat{r}^B = \frac{1+a}{0.4}$.

For $u^A = u^B$ to hold, we must have $\hat{r}^A = \hat{r}^B$ (or once again a coincidental, ignorable relation between the various parameters), and for this, we must have $a = 0.6$.

For lower priority weights w^A and w^B (which are possible in our model), a can become larger, up to $a = 1$ for $w^A = w^B = 0$. This outcome corresponds nicely to what our human subjects did: we saw that most human subjects placed the robot at $a \in [0.6, 0.8]$, with 0.6 being the most frequent choice.

5.4.2 The prioritized Ultimatum Game

The priority awareness model can also be used to explain or predict human strategies in prioritized UGs. As has been mentioned before, in the absence of priorities, the model is equivalent to the inequity-averse model. Thus, human strategies emerging in regular UGs are explained in the same (successful) way by both models. In Figure 5.1, we see the median offer is 50% in this case, with the average varying from 45% (sharing €10) to around 25% (sharing €100,000). Offers of more than 50% are rare. This is indeed predicted by the inequity-averse model, based on the Homo Egualis utility function (as well as the priority-aware model), as we have seen earlier. Using data gathered in these UGs, we can estimate the participants' α^i- and β^i-parameters for every amount at stake (in a similar way as described in Dannenberg et al. (2007)). Clearly, they decrease with an increasing amount. In conclusion, people become increasingly selfish with an increasing amount, and are not punished for that.

Introducing priorities, i.e., richer or poorer players, has no effect on the predictions of the inequity-averse model, assuming that the α^i- and β^i-parameters of individual participants do not change. In Figure 5.1 however, we do see different strategies emerge. With the first (i.e., offering) player being poorer, the median offer is still 50%, but the average drops: around 30% (€10) to 15% (€100,000). Once again, players accept their own offers. This outcome can be predicted by the priority-aware model: assuming that $p^1 > p^2$, the model predicts that the first player wishes (and is allowed) to keep more to himself, which indeed happens. With the first player being richer, the difference between the inequity-averse and priority-aware model becomes even more clear. Many players are now willing to give more than half of the amount to the second player. The median offer is still 50%, but the average increases to around 65% (€10) and 45% (€100,000). This cannot be explained by inequity aversion alone, but priority awareness helps: using this model, we obtain that the *perceived* amount that player 1 gives to player 2 cannot be more than 50%, but the actual amount can. For instance, assume that player 1 has a low priority compared to player 2 (say, $p^1 = 1, p^2 = 4$). In this case, with the priority weights set to 1, a perceived 50%-50% reward distribution corresponds to an actual reward distribution of 20%-80%. Thus, it is possible to give away more than half the actual amount.

5.4.3 Visual appearance

In our survey concerning the impact of visual appearance on human behavior in the UG, we observed substantial differences in the (average) offers being given to the 24 opponents, as can be seen in Figure 5.3. Such differences can clearly not be explained by inequity aversion. Using priority awareness, we may attribute priority values to each opponent, as well as to specific participants in the survey.

Estimating priority values allows us not only to model existing observations, but also to predict how much the people depicted and the participants would be giving *each other*. We note that we have not performed any experiments to determine the validity of this idea. For the idea to be valid, priority values must be sufficiently consistent as well as transitive.

5.5 Priority-aware learning agents

In this section, we apply priority awareness to an actual multi-agent system. The approach is similar to the approach detailed in Chapter 4, since the priority awareness utility function is a rather straightforward extension of the inequity-averse Homo Egualis utility function presented there. Instead of perceiving their actual rewards, agents perceive a reward that is (negatively) influenced by their priority. We can therefore suffice by indicating how we build upon the foundations of Chapter 3 (§5.5.1, briefly summarizing our methodology §5.5.2, discussing our experimental setup, which is slightly different from the one in the previous chapter §5.5.3, and by performing a small set of experiments, aimed only at establishing whether priority awareness allows agents to distinguish between different opponents §5.5.4.

5.5.1 Building upon the foundations

Creating a computational model based on priority awareness, respecting the requirements and the template model discussed in §3.2, may be done in a similar manner as described in Chapter 4 for inequity aversion. Once again, we need to take into account the fact that the utility function used is not smooth. The modifications to our learning algorithm, as introduced in Chapter 4, are also introduced here, with a similar effect.

As discussed in Chapter 3, a computational model of fairness allows agents to answer three questions, i.e., (R3-Q1) what is meant by a fair solution, (R3-Q2) whether a certain solution requires agents to punish others, and (R3-Q3) whether agents should withhold their actions or refrain from participating completely, given their expected outcome. Depending on the answers to each of the three questions, there may be certain consequences. We address each of these consequences here.

R3-Q1 By a fair solution, we imply a reward distribution r for which $(1-\epsilon)n$ agents i experience a utility $u^i > U^i(r_0)$. In this chapter, we (again) use $\epsilon = 0$; thus, we do not allow any agent to experience a utility lower than the baseline utility $U^i(r_0)$. In the games under study in the current chapter, this baseline utility $U^i(r_0) = 0$ for all agents i, as no agent gains or loses anything by not playing. Thus, we strive for reward distributions r for which all agents i experience a utility $U^i(r) \geq 0$.

R3-Q2 Agents will punish others iff this increases their utility u^i. In the previous chapter, we already saw that punishment, based on making a complete interaction fail, works well in the UG as well as the NBG. In this chapter, we therefore always enable punishment. More precisely, we use specific punishment of those others that have performed an offensive action, instead of making the entire interaction fail, as in the previous chapter. Details follow in the next subsection.

R3-Q3 Withholding action implies that agents may decide not to participate if they expect that participating will yield a negative utility u^i. In this chapter, as in the previous chapter, we do not allow agents to do this.

5.5.2 Methodology

Most of our methodology in this chapter is similar to the methodology described in §4.3.2. For clarity, we first provide a short summary. Next, we focus on the two differences between the two methodologies, i.e., targeted punishment, and refined modifications to the CALA learning rule.

Summary

Our agents are each equipped with a CALA, which represents their current strategy, i.e., as a proposer as well as a responder in the UG, and as a player in the NBG.[4] In every iteration, all agents simultaneously select an action a^i. The rules of the game at hand translate the resulting action vector a to a reward vector r. The reward vector r is then translated to a utility value u^i for each agent. If $u^i < 0$, then agent i is willing to punish. The utility value for each agent is updated due to the effects of possible punishment. This utility value is then used as a basis for the agent's CALA to learn a strategy. As in Chapter 4 (§4.3.3), we introduce modifications to the CALA learning rule to facilitate learning.

Targeted punishment

The main difference with the methodology of Chapter 4 is the punishment mechanism. In this chapter, we allow every agent to select (a) specific target(s) for punishment. The punishment mechanism is thus more refined than the one in the previous chapter. It works as

[4] In both cases, we set $\lambda = 0.02$, $\sigma_L = 10^{-7}$, and $K = 1$, after some initial experiments. Changing these parameters does not strongly influence the outcome, although K should not be too small (e.g., 0.01) or too big (e.g., 10), in order to facilitate convergence to a good strategy.

follows. First, agent i calculates his utility $U^i(r)$. If $U^i(r) < 0$, it pays off to punish (since punishment will lead to $U^i(r_0) = 0$). In this case, agent i calculates his pairwise utility $U^i(r^i, r^j)$, for all other agents j. He then punishes those agents j for which the pairwise utility is $U^i(r^i, r^j) < 0$.

In practice, punishment in both the UG as well as the NBG entails that the rewards of agent i as well as agent j are set to 0. Given $U^i(r^i, r^j) < 0$, we may safely assume that $r^i < r^j$. Thus, agent i pays less for punishing agent j than agent j suffers from the punishment.

Refining the modifications to the learning rule for CALA

In the previous chapter, we noted that agents need to start from a reward distribution that is initially valid. We address this issue here. The problem is that our games (especially the NBG) require agents to deal with ambiguous feedback.

Imagine agent i participating in a NBG, in two different situations. First, agent i and the other agents currently request an amount below R, so their request is granted. However, agent i experiences a negative utility (i.e., he receives an amount that is too small in comparison to the amount received by the other agents). He therefore decides to punish, yielding a reward as well as a utility of 0. Second, the agents' currently requested amount exceeds R, so due to the rules of the game, every agent receives a reward as well as a utility of 0.

Even though the agent's utility is the same in both cases, he needs to respond differently. In the first case, for the interaction to become successful, agent i should *increase* his request; in the second case, he should do the opposite. With the modified learning rule of §4.3.3, we always drive agents' strategies down in case of a utility of 0. Thus, if an agent is in the first of the two situations, he will decrease his request, making it even less probable that he will obtain a positive utility in the next game. In effect, when we start with a solution that is initially not accepted by all agents, we will obtain an outcome with all agents requesting 0.

We address this problem in a straightforward way. An agent that punishes is receiving too little, and therefore should not decrease his request, as specified by the modification to the learning rule given in Equation 4.11 on page 57. Thus, if an agent punishes, we do not apply the modification to the learning rule. If an agent is punished, we do apply the modification, which allows the agent to decrease his current request, leading to a more probable agreement with the other agents.

5.5.3 Experimental setup

Our experiments are performed in a similar way to those of Chapter 4. Once again, the agents have to bargain about sharing $R = 100$, starting from an equal split. However, there are two additions. First, we obviously add priorities. Second, we study what happens when add the possibility for priorities to change. Both additions are detailed here.

Adding priorities. We add priorities to our setup by giving agent $i \in [1, n]$ a priority of

$$p^i = C^{\frac{i-1}{n-1}},$$

with C a constant. Due to the formula, the first agent's priority is 1, the last agent's priority is C independent of the number of agents, and the other agents' priorities are (increasing) between 1 and C. In our experiments, we set $C = 2$. The priority weights w^i are set to 1 for all agents. In order to be able to study the effects of priorities properly, we use only low numbers of agents, i.e., $n \in \{2, 3, 5\}$.

Changing priorities. In every experiment, we reverse the priorities of the agents after every $10,000n$ games (also after the first game), i.e., at the first reversal, agent 1 obtains a priority of 2, agent n obtains a priority of 1, and the other agents' priorities are decreasing between 2 and 1. In total, $50,000n$ games are played. Thus, the agents change priorities 5 times, with the first change being immediate.

The remaining parameters of the priority awareness utility function are set as follows. In the UG, we use $\alpha = 0.6$ and $\beta = 0.3$ for all agents, as in the majority of the experiments in the previous chapter. In the NBG, we aim at visualizing the effects of priorities and punishment as well as possible. In the previous chapter, we saw that agents hardly, if ever, punish each other. To increase the probability of agents punishing, we need to decrease the tolerance of the priority awareness utility function against disadvantageous differences, by setting $\alpha = 100$. Thus, agents will quickly punish others if these others obtain too much.

5.5.4 Experiments and results

In this subsection, we present an overview of experiments and results. For each experiment, we show the results of one particular run of $50,000n$ games, as the difference between multiple runs is extremely small. Our result figures display the rewards that agents obtain over time, as well as the average number of times punishment happened in the last 100 games (i.e., a running average), normalized by the number of agents. Thus, 100% (denoted by '1') implies that all agents punished each other, which can obviously not happen.

Two agents. Results for two agents are displayed in Figure 5.4. In the UG, we may calculate, given $\alpha = 0.6$, $p^1 = 1$ and $p^2 = 2$ and Equation 5.5, that the second agent needs to receive at least 40.2 to accept the first agent's proposal. Given $p^1 = 2$ and $p^2 = 1$, the second agent needs to receive at least 15.8. Indeed, in the graph that displays the agents' rewards, we see that initially (i.e., $p^1 = 2$ and $p^2 = 1$, after the reversal of priorities immediately in the beginning) the agents are going towards a situation where the second agent receives decreasingly less. The learning process is interrupted by a priority change before the agents converged to the second agent receiving only 15.8 (which would require the second agent to punish); instead, it receives approximately 21, without the second agent ever punishing. After the priority change, i.e., $p^1 = 1$ and $p^2 = 2$, the punishment executed by the second agent quickly drives the first agent to offering 58.8, leaving 40.2 for the second agent.

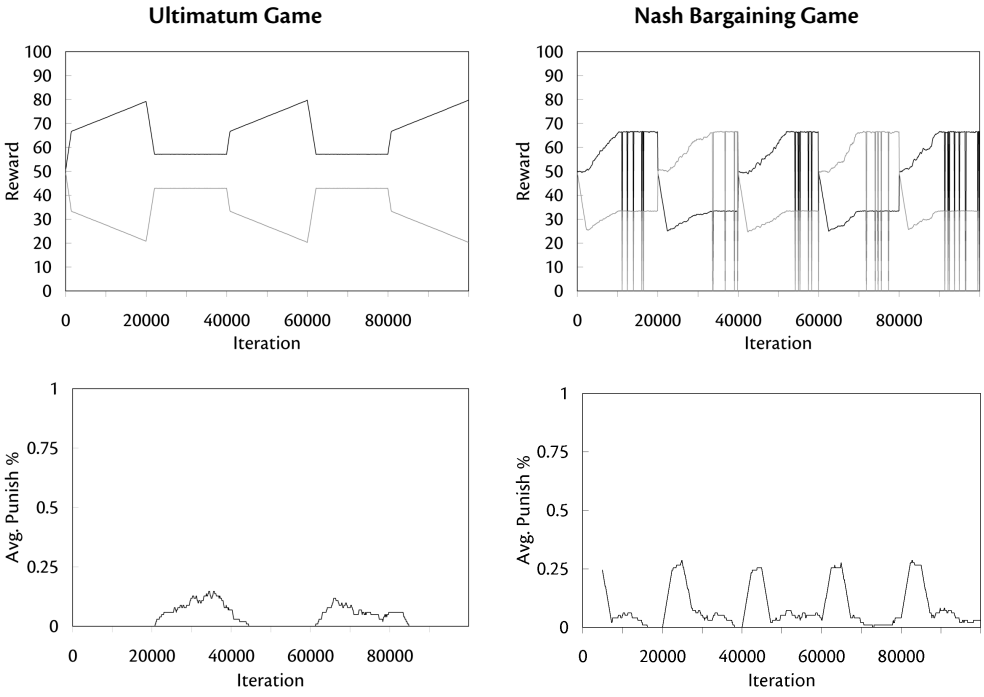


Figure 5.4 Two agents play the Ultimatum and Nash Bargaining Game with priorities

In the NBG, the high value for α implies that the only reward distribution that will be accepted by both agents is a split according to priority. Thus, in the first part of the graph, the agents need to obtain a $\frac{2}{3} : \frac{1}{3}$ split. They indeed effectively achieve this outcome, with punishment being performed quite frequently initially, and with a decreasing frequency over time. After the priority switch, the reverse outcome should be found, and is indeed found, in a similar manner as before the switch. It is interesting to note that, as in Chapter 4, the outcome found is (nearly) Pareto-optimal every time.

Three agents. Results for three agents are displayed in Figure 5.5. The priorities are now initially $p^1 = 1$, $p^2 = \sqrt{2}$, and $p^3 = 2$. As in the two-agent case, these priorities are immediately reversed after the first game. As in the previous chapter, performing calculations with multiple agents is quite difficult, especially in the UG. In the previous chapter, we addressed this issue by setting all agents' α^i- and β^i-parameters to an identical value. Here, we cannot do this, as this would require also setting all priorities to identical values, which would violate the whole idea of the work presented here. Thus, we do not analytically determine the required outcome in the UG here. Clearly, as CALA have a proven convergence to a (local) optimum, any joint action that is repeatedly and stably established by the CALA, may be assumed to be the required outcome.

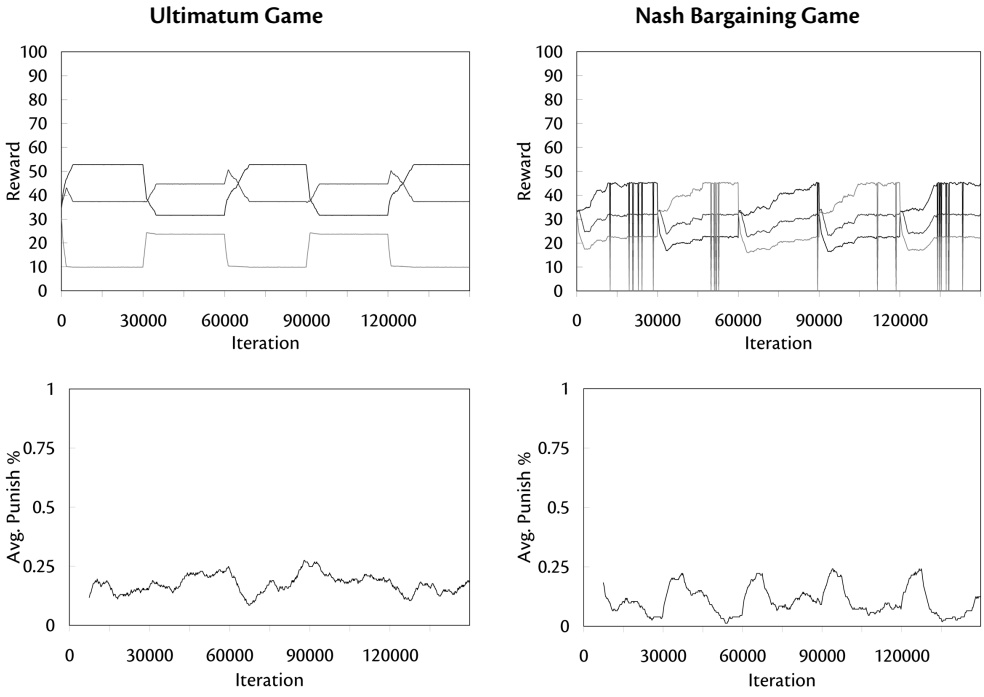


Figure 5.5 Three agents play the Ultimatum and Nash Bargaining Game with priorities

In the UG after the initial reversal of priorities, the three agents obtain approximately 52.7, 37.3, and 9.9. After another reversal, these numbers quite quickly change to 31.6, 44.7, and 23.7. Punishment is often performed by at least one of the three agents.

In the NBG, given the high value of α, we can calculate an approximate expected solution, which is a split according to the priorities. Thus, with $p^1 = 1$, $p^2 = \sqrt{2}$ and $p^3 = 2$, we expect a distribution of approximately 22.7, 32.5 and 45.5. Indeed, the distributions obtained by the agents are very close to this expectation, as well as very close to Pareto-optimality. Punishment behavior is similar to the two-agent case, with agents initially applying more punishment, and gradually reducing their frequency of punishment.

Five agents. Results for five agents are displayed in Figure 5.6. Clearly, these results are very similar to those obtained with a lower number of agents. In case of the NBG with five agents, we see that the last of the episodes does not converge. To investigate whether this is a coincidence, we analyzed the outcomes of $1,000$ experiments with five agents, running for $50,000$ games each. Of these $1,000$ experiments, 93% had the same outcome (i.e., the outcome also obtained in most episodes here). The other 7% did not converge properly.

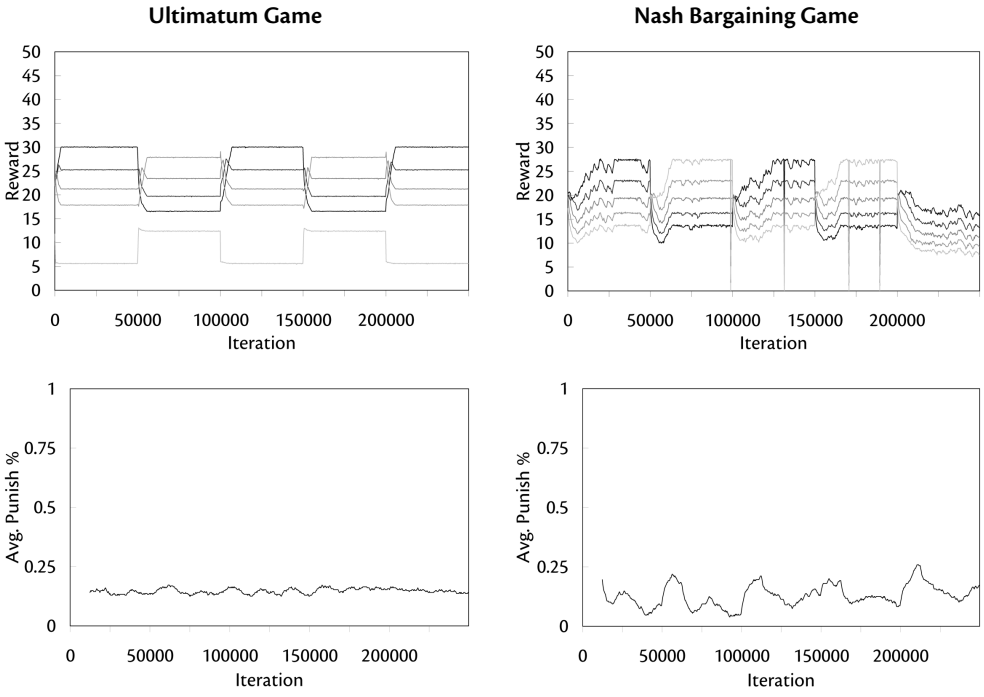


Figure 5.6 Five agents play the Ultimatum and Nash Bargaining Game with priorities

5.6 Chapter conclusion

In this chapter, we discussed reputation and priority awareness. We will now answer the research questions RQ1 and RQ3 to RQ5 for this chapter.

RQ1 *How are humans using fairness in their decisions?*

Existing research, as described in §5.1, discovered the importance of concepts such as reputation and reciprocity in human decision-making. We showed that this existing research misses an important point, i.e., that reputation may not only be established over repeated interactions, but may also be immediately available, due to additional information or the activation of stereotypes. We discussed a number of examples of this phenomenon in §5.2, and proposed a model called priority awareness, which is an extension of inequity aversion, in §5.3. We showed how the model may explain human decision-making in a number of examples in §5.4.

RQ3 *How can human-inspired fairness be modeled computationally, taking into account the foundations of human fairness?*

In §5.5, we presented a computational model of priority awareness, combining the priority awareness utility function with continuous-action learning automata.

RQ4 *What are the analytical properties of the computational human-inspired fairness models developed in this research?*

The priority awareness model is an extension of the inequity aversion model of Chapter 4. Thus, in the absence of priorities, the models have identical analytical properties. We analyze that, with priorities being present, the priority awareness model allows certain agents to obtain more (or less) than what they would obtain with inequity aversion. We calculated expected solutions for the games presented to our learning agents, i.e., the UG and the NBG.

RQ5 *What are the (empirical) benefits of incorporating explicitly human-inspired fairness in adaptive multi-agent systems?*

In §5.5, we showed that our agents learn valid solutions to the UG and the NBG with additional information, i.e., priorities (for the NBG they may be seen as an equivalent to introducing bargaining powers; see §2.1.3 and §2.3.1). These solutions are fair according to the definitions of fairness we presented in Chapter 3.

6 Fairness in social networks

In this chapter, we construct a computational model of fairness that includes interaction in social networks and the dissemination of strategies (or *opinions*, in terms of related work) through these networks. In contrast to the models presented in the previous chapters, the social-network model essentially provides an 'external' motivation for fairness.

Existing work considering interaction networks commonly studies how strategies disseminate in limited games, e.g., with only two strategies, one of which is labeled as 'cooperative' and the other as 'defective'. As has been discussed in §3.3.2, this abstraction may be taken into account when the goal of research is attempting to find support for proposed human mechanisms, but not when the goal is explicitly developing computational models for multi-agent systems. As we are pursuing the latter goal, we extend existing work to the domain of games with continuous strategy spaces. We study whether, and if so, how, mechanisms proposed in existing work may be used here, and add a number of new mechanisms. In §6.1, we discuss existing research concerning opinion dynamics and social networks, which is generally limited to a discrete set of strategies (or opinions). In §6.2, we discuss our methodology, aimed at extending existing work to continuous strategy spaces, as well as making it explicitly computational. We discuss our experimental setup in §6.3, and in §6.4, we present a large set of experiments and results. We review a number of alternative, less successful mechanisms in §6.5. Finally, we conclude the chapter in §6.6.

The work reported on in this chapter has partially been published in:

S. de Jong, K. Tuyls, and K. Verbeeck. Fairness in multi-agent systems. *Knowledge Engineering Review*, Vol. 23(2):153-180, 2008.

S. de Jong, S. Uyttendaele, and K. Tuyls. Learning to reach agreement in continuous strategy spaces. *Journal of Artificial Intelligence Research*, Vol. 33:551-574, 2008.

S. de Jong, and K. Tuyls. Learning to cooperate in a continuous tragedy of the commons. *Proceedings of the 8th International Conference on Adaptive Agents and Multi-Agent Systems (AAMAS)*, (in press), 2009.

6.1 Opinion dynamics and social networks

Research concerning social networks is commonly performed from the perspective of statistical physics, which studies the emergence of global complex properties from purely local rules (Reichl, 1980). In statistical physics, a wide variety of phenomena with an inherently stochastic nature is described, for instance in biology, chemistry, and neurology. In addition, techniques have been used to describe phenomena that are observed in sociology. In this case, we may use the term "sociophysics" (Arnopoulos, 1993; Galam, 2004; Castellano et al., 2008). Sociophysics aims at simple models which may allow to understand fundamental aspects of social phenomena, such as opinion dynamics, i.e., the dynamics of opinion formation and opinion spreading.

Opinion dynamics models (Hegselmann and Krause, 2002; Fortunato, 2005) aim at describing (1) how certain opinions may be formed and (2) how opinions are subsequently spread over the entire population. Moreover, they allow us to predict (3) whether agreement will be reached or not. Various models have been proposed and even quantitatively validated, for instance in predicting voting behavior in national elections (Fortunato and Castellano, 2007) and in the formation of language (Dall'Asta et al., 2006b; Barrat et al., 2007).

The Axelrod model is one of the most well-known opinion dynamics models. Axelrod (1997) notes that interactions between humans generally tend to favor homogeneization. However, he then puts forward the question: "If people tend to become more alike in their beliefs, attitudes and behaviors when they interact, why do not all differences disappear?" Axelrod identifies two mechanisms why this happens, i.e., first, *social influence* and second, *homophily*. These two mechanisms may be defined as follows.

Definition 6.1 Social influence in the Axelrod model is the phenomenon that interactions make individuals more similar.

Definition 6.2 Homophily in the Axelrod model is the phenomenon that the likeliness of interaction grows with similarity.

Most initial models of opinion dynamics either used a mean-field approach which assumed that every possible pair of agents has an equal probablity of exchanging opinions (as in the previous chapters), or an approach which assumed that agents' networks of interaction could be described as regular lattices, as for instance in the Ising model (Ising, 1925). Recent progress in network science has revealed that human networks of interaction, as many real-world networks, are commonly complex (Barrat et al., 2007). The most notable complex network structure is the *scale-free network structure*, which entails that many agents interact with only a few other agents, whereas few agents interact with many other agents (Barabasi and Albert, 1999). In other words, some agents are *hubs* that are connected to a large number of others, whereas most agents are connected to only a few others. The mechanisms of social influence and homophily, as stipulated by the Axelrod model described above, may explain why such complex network structures have emerged.

Given that humans commonly interact in scale-free networks, researchers have studied whether such scale-free networks may indeed be an explanation for the human tendency

to cooperate (for instance, in social-dilemma interactions). In research closely related to the research presented in this chapter, Santos et al. (2006a,b) examine the effect on cooperation of local interactions in complex networks using a generalized class of discrete-strategy social dilemmas, characterized by a reward matrix for both players that is parametrized using two variables $s \in [-1, 1]$ (sucker's payoff) and $t \in [0, 2]$ (temptation), see Figure 6.1. We note that this research therefore introduces the abstraction of limiting agents to two strategies, whereas typical applications of multi-agent systems require us to allow agents to select strategies from a continuous strategy space. For a more elaborate discussion, see §3.3.2.

Many values of s and t, i.e., many social dilemmas, are played using agents that are optimizing their strategy by means of an evolutionary algorithm. Agents may imitate the strategies of other agents with whom they played, with a probability similar to the difference in reward (i.e., a successful agent is more likely to be imitated; this corresponds to social influence in the Axelrod model). The agents are paired based on various network topologies, including single-scale networks and scale-free networks, as can be generated using the well-known algorithm by Barabasi and Albert (1999) (see §6.2.5).

Santos et al. (2006b) show that cooperation is the dominant strategy in scale-free networks for many more values of s and t than in other networks; thus, the heterogeneity of the network with which agents interact, increases the probability of cooperation becoming the dominant strategy. More precisely, cooperation prevails once (1) there is a majority of cooperators, (2) the hubs (densely connected agents) in the network are interconnected, and (3) the network is sparse. However, in any other case, cooperation is (still) doomed.

Santos et al. (2006c) therefore continued their studies and subsequently argued that we should not assume that connections in social networks are static; after an unsatisfactory interaction, agents may choose to rewire their connection to someone else (i.e., following the idea of homophily in the Axelrod model). Basically, everyone would like to connect to a cooperator (since this leads to a higher reward for both a defector as well as a cooperator). The probability that such rewiring is allowed, may be chosen to depend on the fitness (average reward) of the agent wanting to be rewired.

Santos et al. (2006c) showed that, with a sufficiently high rewiring probability, full cooperation can be reached in all social dilemma games experimented on. The resulting networks are realistic, with a high average connectivity and associated single-scale to broad-scale heterogeneity. Thus, the authors demonstrated how the Axelrod model of social influence and homophily may be implemented, and how it leads to network structures similar to real-

	a_{21}	a_{22}
a_{11}	$(1, 1)$	(s, t)
a_{12}	(t, s)	$(0, 0)$

Figure 6.1 A generalized discrete social dilemma with two strategies, as presented by Santos et al. (2006b). Various dilemmas can be constructed using two variables $s \in [-1, 1]$ (sucker's payoff) and $t \in [0, 2]$ (temptation).

world human interaction network structures. We note that the idea of rewiring is closely related to the concept of volunteering, as introduced by, e.g., Hauert et al. (2002, 2007); agents can choose whether or not to interact with certain other agents (see Chapter 5).

6.2 Methodology

In this chapter, we extend the work of the previous chapters (most notably Chapter 4) to social networks. As in the previous chapters, we use CALA (with modifications to the learning rule, as used before) as a means for agents to learn an optimal joint strategy. In comparison to our previous work, there are two main differences. First, agents play pairwise games, based on a scale-free network structure.[1] We will show that a structured multi-agent system may efficiently be populated with a much larger number of agents (e.g., thousands) than an unstructured one. Second, we do not use the Homo Egualis utility function or the priority awareness utility function. Instead, agents feel fairly treated as long as other agents perform similar (or more cooperative) strategies. The desired, human-inspired outcome offered by the utility functions' parameters in the previous chapters, is replaced here by (potentially) including agents that always play according to a certain fixed strategy (i.e., simulated human players). We play only pairwise games, in which agents consider those opponents acceptable that have a strategy that is at least as cooperative as their own strategy.

The remainder of this section is divided in eight subsections. In §6.2.1, we address the foundations specified in Chapter 3. In §6.2.2, we summarize our basic setting for the sake of clarity. We continue in §6.2.3 by briefly repeating the working of continuous learning automata, as they are central to our methodology. In §6.2.4, we provide details concerning punishment in the PGG, which, as we saw in the previous chapters, is not a sufficient mechanism to drive our agents to satisfactory outcomes in continuous strategy spaces. We propose an extension to the punishment mechanism that does lead to satisfactory outcomes. In §6.2.5, we discuss the structure and topology of the networks of interaction we use. In §6.2.6, we discuss the agent types and initial strategies of agents. In §6.2.7, we elaborate on how we provide the additional possibility of rewiring connections between agents.

6.2.1 Building upon the foundations

The foundations of §3.2 specify that any computational model of fairness is required to be (R1) founded in game theory, (R2) computationally applicable, and (R3) human-inspired. Since we use a rather simple utility model in this chapter (i.e., agents experience a positive utility when they interact with another agent that is roughly at least as cooperative as they are), the first two requirements are met.

For the model to be explicitly human-inspired, agents must be able to answer three questions, i.e., (R3-Q1) what is meant by a fair solution, (R3-Q2) whether a certain solution re-

[1] In this chapter, we consider the agreement dilemma by means of the UG, and the tragedy of the commons by means of the PGG. The NBG is not considered here.

quires agents to punish others, and (R3-Q3) whether agents should withhold their actions or refrain from participating completely, given their expected outcome. We address each of these questions here.

R3-Q1 In this chapter, we use a simpler fairness utility function than in the previous two chapters, i.e., a pairwise function $u^i(r^i, r^j) = r^j - \tau r^i$, both in the UG and the PGG. The constant τ is a tolerance value, which is usually set to 0, unless noted otherwise. As in previous chapters, $u^i(r_0) = 0$, as refraining from participation yields no reward at all. This chapter is not based on the assumption that all agents need to agree, i.e., $\epsilon = 0$ should not necessarily hold in a satisfactory outcome. Instead, the agents should try to minimize ϵ. In the remainder of the chapter, we show ϵ by means of *performance*. A performance of 0.9, for instance, implies that agents find a fair solution with 9 out of 10 neighbors. Thus, with a performance of 0.9, we have $\epsilon = 0.1$.

R3-Q2 Punishment is an important mechanism in this chapter, as in the previous two chapters. In the UG, agents may have different strategies, which may result in a responder wanting to punish a relatively defective proposer. Given the utility function, this will indeed increase the responder's utility. In the PGG, we use the analysis presented in §4.2.2 to justify the assumption that agents are *always* willing to punish those who contribute less than they themselves; therefore, the utility function above may be used here as well. We introduce a punishment mechanism in §6.2.4 that effectively allows relative cooperators to discourage relative defectors from defecting again.

R3-Q3 Withholding action implies that agents may decide not to participate if they expect that participating will yield a negative utility u^i. In this chapter, withholding action is facilitated by (potentially) using rewiring in the network of interaction. After every interaction with a certain other agent j, agent i may decide to break the link between him and j, and create a new link to a random neighbor of j. Clearly, if $u^i(r^i, r^j) < 0$, agent i will increase his expected future utility by abandoning j.

6.2.2 The basic setting

We study a large group of adaptive agents, driven by continuous action learning automata, playing the UG or the PGG in pairwise interactions. Pairs are chosen according to a (scale-free) network of interaction. In the UG, every agent is randomly assigned the role of proposer or responder.

Agents start with different strategies. For a good performance in the UG, most of them need to converge to agreement by playing many pairwise games, i.e., they need to learn a common strategy. In the PGG, the agents need to converge to the most contributive strategy for optimal performance. Some agents may be fixed in their strategies; these agents represent an external strategy that the adaptive agents need to converge to (for instance, a preference imposed by humans in the UG, or the desired cooperative strategy in the PGG).

Additionally, we study the influence of adding the option for agents to rewire in their network of interaction as a response to an agent that behaved in a defecting manner.

6.2.3 Continuous action learning automata

CALA (Thathachar and Sastry, 2004) have been discussed in §2.2.3. We remind the reader that although CALA have a proven convergence to (local) optima, multiple CALA learning a joint strategy may suddenly have a drastic 'jump' in the mean value of their underlying Gaussian distribution. Therefore, we need to modify the CALA's update function, as detailed in §4.3.3 (Equation 4.10).

In the UG, we also use the second modification given in §4.3.3 (Equation 4.11), i.e., we slightly lower the CALA's mean strategy in case of zero or identical feedback. In the PGG, this second modification is not needed, as our learning approach (detailed immediately below) prevents the CALA from receiving zero or identical feedback.

6.2.4 Probabilistic punishment in the Public Goods Game

Even if the analysis in §4.2.2 may be used to motivate why agents punish, we need an additional concept to make punishment work in a learning-agents setting with a continuous strategy space. As discussed in §2.2.3, CALA (as well as many other learning algorithms) perform a great deal of local search, and therefore, the essential idea underlying punishment in the PGG, i.e., a reversal of the inverse relation between contribution and reward, fails to work if agents get punished equally for two actions that are not equally defective.

To solve this problem, we propose the mechanism of *probabilistic punishment*, which is inspired by common-sense human behavior, i.e., the probability that an agent i punishes an agent j depends on the actual rewards r^i and r^j. Punishment is more often performed for higher differences between these rewards. We formalize this in a way that ensures that punishment allows agents to learn to contribute higher amounts. Assume that agent i contributes an amount c^i, and j contributes a lower amount, $c^j = c^i - \Delta$. It is fairly easy to calculate that agent j will gain an amount of $0.5r(2c^i - \Delta)$, minus an investment of $c^i - \Delta$. Given that agent i punishes j with a probability $P^i(\Delta)$, we find that agent j obtains an expected reward $\tilde{r}^j$ of:

$$\tilde{r}^j = 0.5r\left(2c^i - \Delta\right) - (c^i - \Delta) - P^i(\Delta)\, e_p \tag{6.1}$$

Clearly, it pays off for agent j to decrease Δ (i.e., increase c^j) if $\tilde{r}^j$ increases with a decreasing Δ. Once again, it is easy to calculate that this requires the following for $P^i(\Delta)$:

$$P^i(\Delta) > \frac{(1 - 0.5r)\,\Delta}{e_p}. \tag{6.2}$$

As the difference Δ between two contributions c^i and c^j is at most c (i.e., the maximum amount allowed), we must ensure that $e_p > (1 - 0.5r)\,c$. Otherwise, punishment may need to be performed with a probability greater than 1. In our case, we set $c = 10$ and $r = 1.5$, so that $e_p > 2.5$ should hold. For $e_p = (1 - 0.5r)\,c$, Equation 6.1 yields $P^i(\Delta) > \frac{\Delta}{c}$, and thus, for larger values of e_p, we may conveniently set $P^i(\Delta) = \frac{\Delta}{c}$. In this chapter, we use $e_p = 3$.

6.2.5 The network of interaction

A scale-free network (Barabasi and Albert, 1999) is a network of which the degree distribution follows a power law. More precisely, the fraction $P(k)$ of nodes in the network having k connections to other nodes goes for large values of k as $P(k) \sim k^{-\gamma}$. The value of the constant γ is typically in the range $2 < \gamma < 3$. Scale-free networks are noteworthy because many empirically observed networks appear to be scale-free, including the world wide web, protein networks, citation networks, and also social networks. The mechanism of preferential attachment has been proposed to explain power-law degree distributions in some networks. Preferential attachment implies that nodes prefer attaching themselves to nodes that already have a large number of neighbors, over nodes that have a small number of neighbors.

Previous research has indicated that scale-free networks contribute to the emergence of cooperation (Santos et al., 2006b). We wish to determine whether this phenomenon still occurs in continuous strategy spaces and therefore use a scale-free topology for our interaction network, using the Barabasi-Albert model. More precisely, the probability p^i that a newly introduced node is connected to an existing node i with degree k^i is equal to:

$$p^i = \frac{k^i}{\sum_j k^j}. \tag{6.3}$$

When we construct the network, the two first nodes are linked to each other, after which the other nodes are introduced sequentially and connected to one or more existing nodes, using p^i. In this way, the newly introduced node will more probably connect to a heavily linked hub than to one having only a few connections.

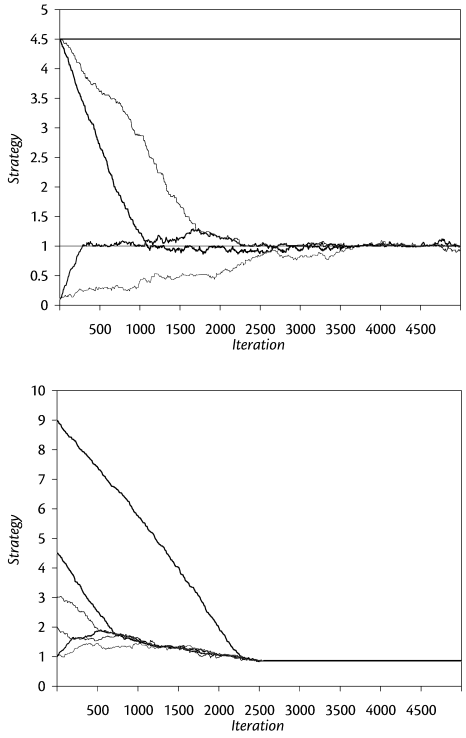
In our simulations, we connect every new node to one, two, or three existing ones (uniform probabilities). This yields networks of interaction that are more realistic than the acyclic ones obtained by always connecting new nodes to only one existing node. For example, if the network is modeling a friendship network, avoiding cycles means assuming that all friends of a certain person are never friends of each other.

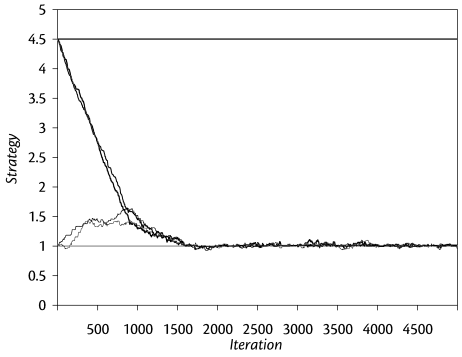
6.2.6 Agent types and strategies

To study how certain strategies emerge in the UG and the PGG, we make our agents start from a situation in which they have different strategies. Moreover, we study the effects of adding 'example' agents, as well as agents that use a different, potentially relatively defective strategy. This leads to two types of agents.

Two types of agents

We introduce two types of agents, i.e., *dynamic strategy* (DS) agents and *fixed strategy* (FS) agents. DS agents are the learning agents. They start with a certain predefined strategy and are allowed to adapt their strategy continuously, according to the learning mechanism of their learning automaton. FS agents are (optional) 'good examples': they model an example





(top-left) Two DS agents, one starting at offering and accepting 4.5, and one starting at offering and accepting 0.01, learn to play optimally against each other and a FS agent offering 4.5 and accepting 1. The DS agents rather quickly learn to offer and accept 1.

(top-right) Two DS agents, both starting at offering 4.5 and accepting 1, learn to play optimally against each other and a FS agent also offering 4.5 and accepting 1. Again the DS agents learn to offer and accept 1.

(bottom-left) Three DS agents, starting at different initial strategies (i.e. offering 9, 4.5, and 1, and accepting 3, 2 and 1, respectively), quickly learn a single, similar strategy.

Figure 6.2 Evolving strategies in a fully connected network of three agents. Proposal strategies are indicated by a bold line, response strategies by a thin line. Agents converge to a situation in which their two initial strategies become similar.

strategy that needs to be learned by the (other) agents in our system, and therefore refuse to adapt this strategy.

One or two CALA per agent in the Ultimatum Game?

In the previous chapters, we equipped agents that played the UG with only one CALA. Clearly, each agent needs to be able to perform two different roles in the UG, i.e., playing as the proposer as well as playing as the responder. In other words, an agent is in one of two distinct states, and each state requires it to learn a different strategy. As CALA are stateless learners, each agent would therefore actually require two CALA. Nonetheless, in the remainder of this chapter, as in previous chapters, we equip every DS agent with only one CALA, representing both the agent's proposer strategy as well as its responder strategy.

Our choice for one CALA is motivated by two observations, i.e., (1) human behavior, and (2) initial experiments. First, human strategies are often consistent, implying that they generally accept their own offers, but reject offers that are lower (Bearden, 2001; Oosterbeek et al., 2004), even with high amounts at stake (Slonim and Roth, 1998; Cameron, 1999;

Table 6.1 Agents' initial strategies in the Ultimatum Game and the Public Goods Game

	UG		PGG	
	Avg	Std	Avg	Std
FS	4.50	0	10.00	1
DSh	4.50	1	7.00	1
DSr	0.01	1	0.01	1

De Jong et al., 2008). Second, in a set of initial experiments, we observed that agents using two CALA will generally converge to one single strategy anyway. As an illustration, three learning curves obtained in a fully connected network of three agents playing the UG are displayed in Figure 6.2. It is clearly visible that agents' proposer strategies (bold lines) are strongly attracted to other agents' responder strategies (thin lines), and especially to the lowest of these responder strategies. In the presence of a FS agent that offers 4.5 and accepts at least 1, the first strategy is immediately ignored in favor of the (lower) second one. With only DS agents, once again all strategies are attracted to the lowest responder strategy present.[2]

We use the observation to justify an abstraction, i.e. we limit the complexity of our agents by equiping them with only one CALA. This CALA then represents the agent's proposer strategy as well as its responder strategy. It is updated when the agent plays as a proposer as well as when it plays as a responder, according to the CALA update formula presented in §2.2.3 and the modifications presented in §4.3.3. Thus, the agents' single CALA receive twice as much feedback as two separate CALA would. This abstraction therefore increases the efficiency of the learning process.

Agents' strategies

In our simulations, we use two types of DS agents (DSr and DSh) and one type of FS agent. More precisely, DSr agents are learning agents that start at a rational solution of offering $X \sim N(0.01, 1)$ in the UG as well as the PGG, and also accepting their own amount or more in the UG. DSh agents start with the more human, fair solution of offering $X \sim N(4.5, 1)$ in the UG (and also accepting their own amount or more), and of offering $X \sim N(7, 1)$ in the PGG. Since FS agents are examples of a desired solution, we equip them with a cooperative solution to see whether the other agents are able to adapt to this solution. The FS agents always offer 4.5 in the UG (and accept any offer of 4.5 or more), and 10 in the PGG. For convenience, these settings are given in Table 6.1.

[2] As is visible in Figure 6.2, the agents more quickly adapt their strategies downward than they do upward. As a result, with multiple (e.g., 10) DS agents learning together (i.e., without any FS agents), we observe that their strategy usually converges to 0. This behavior is due to an artifact of the learning process; two CALA trying to learn each others' current strategy tend to be driven downward in the UG.

All agents are limited to strategies taken from a continuous interval $c = [0, 10]$, where 10 is chosen as the upper bound (instead of the more common 1) because it is a common amount of money that needs to be shared in the UG or contributed in the PGG. If any agent's strategy falls outside the interval c, we round off the strategy to the nearest value within the interval.

6.2.7 Rewiring

Agents play together based on their connectiveness in the interaction network. Thus, in order to avoid playing with a certain undesirable neighbor j, agent i may decide to break the connection between him and j and create a new link to a random neighbor of j (Santos et al., 2006c).[3] For rewiring, we use a heuristic proposed by Santos et al. (2006c): agents want to disconnect themselves from (relative) defectors, as they prefer to play with relative cooperators. Thus, the probability that agent i unwires from agent j, is calculated as:

$$p^r = \frac{\mu^i - \mu^j}{c}. \tag{6.4}$$

Here, μ^i and μ^j are the agents' current strategies (in the UG, agent i's responder strategy and agent j's proposer strategy), and c is the amount at stake in the game at hand, i.e., 10 in both the UG and the PGG. Even if agents determine that they want to unwire because of this probability, they may still not be allowed to, if this breaks the last link for one of them. If unwiring takes place, agent i creates a new wire to a random neighbor of agent j.

6.3 Experimental setup

Using the aforementioned types of agents, we need to determine whether our proposed methodology possesses the traits that we would like to see.

In the UG, our population can be said to have established a successful agreement if it manages to reach a common strategy that incorporates the preferences of the good examples, while at the same time discouraging those agents that try to exploit the dominant strategy. Thus, in a population consisting of only DS agents, any strategy that is shared by most (or all) agents leads to good performance, since all agents agree in all games, yielding an average reward of 5 per game per agent – our architecture should be able to find such a common strategy. When using DS as well as FS agents, the FS agents impose the strategy that the DS agents should converge to, regardless of whether they start as DSr or as DSh agents.

In the PGG, we wish to obtain the outcome that all agents contribute their entire private amount c to the common pool. Given our setting of $r = 1.5$, this means that every agent

[3] Note that we may also choose to allow an agent i to create a new connection to specific other agents instead of only random neighbors of their neighbor j. However, especially in combination with reputation (see §6.5.1), this allows (relative) defectors to identify (relative) cooperators quickly, with which they may then connect themselves in an attempt to exploit. Preliminary experiments have shown that this behavior leads to the interaction network losing its scale-freeness, which may seriously impair the emergence of agreement.

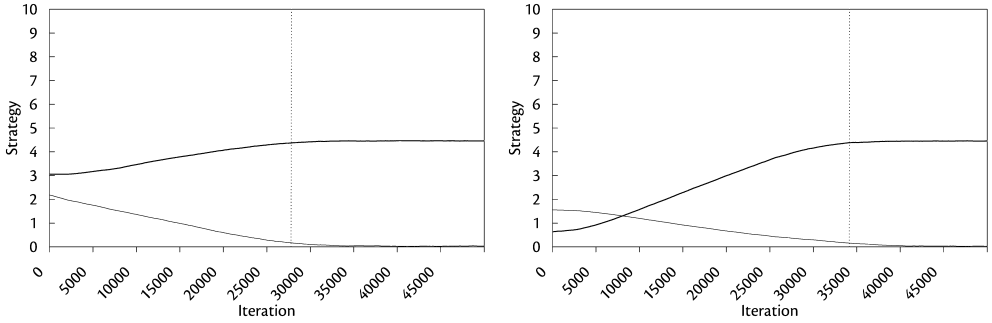


Figure 6.3 Two examples of the convergence point of a single run. In both graphs, we display the average strategy of the population (bold line) as well as the standard deviation on this average (thin line). The dotted vertical line denotes the convergence point, as found by the analysis detailed in the text.

gains 5 per pairwise game. Once again, the initial heterogeneity of the population's strategies leads to some agents (e.g., DSr agents) trying to exploit others (e.g., DSh and FS agents).

In order to measure whether the agents achieved a satisfactory outcome, we study four quantitities related to the learning process and the final outcome, viz. (1) the point of convergence, (2) the learned strategy, (3) the performance, and (4) the resulting network structure. We will briefly explain these four quantities below. In general, we remark that every simulation for the UG lasts for $3,000$ iterations per agent, i.e., $3,000n$ iterations for n agents. For the PGG, we use $6,000n$ iterations. In both cases, we repeat every simulation 50 times to obtain reliable estimates of the quantities of interest.

Point of convergence. The most important quantity concerning the agents' learning process is the point of convergence, which, if present, tells us how many games the agents needed to play in order to establish an agreement. To determine the point of convergence, we calculate and save the average population strategy $avg(t)$ after each pairwise game (i.e., each iteration of the learning process). After T iterations, we obtain an ordered set of T averages, i.e., $\{avg(1),\ldots,avg(T)\}$. Initially, the average population strategy changes over time, as the agents are learning. At a certain point in time t, the agents stop learning, and as a result, the average population strategy $avg(t)$ does not change much anymore. To estimate this point t, i.e., the point of convergence, we find the lowest t for which the standard deviation on the subset $\{avg(t),\ldots,avg(T)\}$ is at most 10^{-3}. Subsequently, we report the number of games per agent played at iteration t, i.e., $\frac{t}{n}$. In our experiments, every simulation is repeated 50 times, resulting in 50 convergence points. We will use a box plot to visualize the distribution of these 50 convergence points.[4]

[4] Note that, in our box plots, we report the average instead of the median, as the average is a more informative quantity, for instance when comparing our results with earlier results. This may allow the box plots' mid point to be located outside the box.

As an example, in Figure 6.3 (left), we see how 16 DSr, 17 DSh agents and 17 FS agents converge to agreement in the UG, using rewiring. Only the first $50,000$ games are shown. In addition to a bold line denoting the average population strategy, as described above, we also plot a thinner line, denoting the standard deviation on this average. Using the method outlined above, the point of convergence is determined to be around $27,500$ games, i.e., approximately 550 games per agent were necessary. In Figure 6.3 (right), we show similar results for 40 DSr agents and 10 FS agents, once again using rewiring. Here, the point of convergence is around $34,000$ games, i.e., approximately 680 games per agent were necessary, which means that learning to reach agreement was more difficult.

Learned strategy. Once we established at which iteration t the agents have converged, we can state that the average learned strategy is precisely $avg(t)$. We repeat every simulation 50 times to obtain a reliable estimate of this average. Once again, in our results, we use a box plot to visualize the distribution of the average learned strategy.

Performance. To measure performance, we first allow our agents to play $3,000n$ UGs or $6,000n$ PGGs. Then, we fix the strategies of all DS agents. In the UG, we subsequently let every agent play as a proposer against all its neighbors (one by one), and count the number of games that were successful. In the PGG, we determine the number of neighbors that have an identical strategy.[5] We divide this number through the total number of games played (i.e., twice the number of edges in the interaction network). The resulting number p denotes the performance, which lies between 0 (for utterly catastrophic) and 1 (for complete agreement). With regard to our computational framework of Chapter 3, we note that $p = 1 - \epsilon$. Human players of the UG typically achieve a performance of 0.8 to 0.9 (Fehr and Schmidt, 1999; Oosterbeek et al., 2004), or equivalently, $\epsilon = 0.1$ to 0.2. Once again, the 50 repetitions lead to 50 measures of performance, which are displayed in a box plot in our results.

Resulting network structure. Since the network of interaction may be rewired by agents that are not satisfied about their neighbors, we are interested in the network structure resulting from the agents' learning processes. We examine the network structure by looking at the degree distribution of the nodes in the network (i.e., the number of neighbors of the agents). With 50 repeated simulations, we may draw a single box plot expressing the degree distribution, averaged over 50 different networks.

6.4 Experiments and results

In this section, we present our experiments and results. In two subsections, we show the results for the agreement dilemma (UG) and for the tragedy of the commons (PGG).

[5] Note that the CALA update formula prevents agents from converging to an exact strategy, as the standard deviation of the CALA's Gaussian is kept artificially strictly positive. Therefore, there is some noise on the strategies to which agents have converged. To counter this noise while measuring performance, we apply a tolerance interval of 99%. Thus, an agent having a strategy of 4 will accept any offer of 3.96 or more in the UG, and will not punish a neighbor with a contribution of 3.96 in the PGG.

6.4.1 The agreement dilemma

We present our experiments and results concerning the UG in two parts. First, we vary the population size; we study the behavior of various population sizes, given a setup without rewiring and a setup with rewiring, while keeping the proportion of DSr, DSh and FS agents constant and equal (i.e., 33% for each type of agent). Second, we vary the proportion of good examples (i.e., FS agents), i.e., we study the same two setups, this time varying the proportion of FS agents, where the remainder of the population is half DSr and half DSh.

In general, we remark that every experiment reports results that are averaged over 50 simulations. In every simulation, we allowed the agents to play $3,000n$ random games, where n denotes the number of agents.

Varying the population size

In many multi-agent systems, increasing the number of agents causes difficulties. Many mechanisms that work with a relatively low number of agents stop working well with a high number of agents, for instance due to computational complexity or undesired emergent properties. According to previous research, this issue of *scalability* also applies to the task of learning social dilemmas. Indeed, previous research using evolutionary algorithms with discrete strategy sets mentions that the number of games needed to converge to an agreement (i.e., on cooperation) may be "prohibitively large" (Santos et al., 2006c).[6]

Since our agents are learning a similar task, we may expect a scalability issue as well. To determine whether our proposed methodology has such an issue, we vary the population size between 10 and 10,000 (with some steps in between), while keeping the proportion of DSr, DSh and FS agents constant at one-third each. We study a setup without rewiring as well as a setup with rewiring, and determine (1) the point of convergence, i.e., the number of games per agent needed to reach convergence; (2) the average learned strategy the agents converged to; (3) the final performance of the system; and finally (4) the resulting network structure. Especially the first and third of these quantities give an indication of the scalability of our methodology.

Point of convergence (Figure 6.4). A setup without rewiring (left) tends to require more games per agent as the total number of agents increases. At a certain point, i.e., around a population size of 200 agents, this tendency stops, mainly because the average number of games per agent approaches the maximum, i.e., 3,000 games per agent. A setup with rewiring (right) convincingly outperforms one without rewiring, as increasing the population size hardly affects the number of games per agent required to reach convergence. Independent of the population size, the setup requires approximately 500 games per agent to converge. Note the difference with previous research (i.e. Santos et al., 2006c), which reports requiring 10^5 games per agent (or more).

[6] In order to limit the time taken for learning, Santos et al. (2006c) terminate the learning process after 10^8 iterations, while using at most 10^3 agents, leading to an average of (more than) 10^5 games per agent being available. With this high number of games per agent, agents occasionally still do not converge.

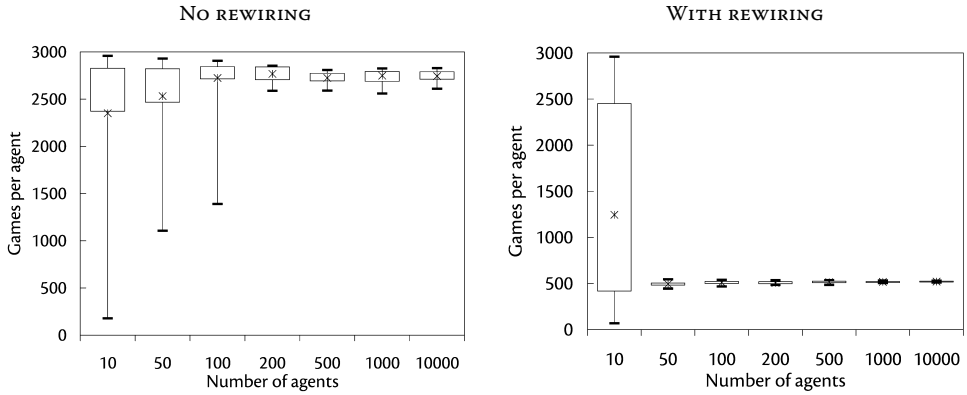


Figure 6.4 Games per agent until convergence; without rewiring (left) and with rewiring (right)

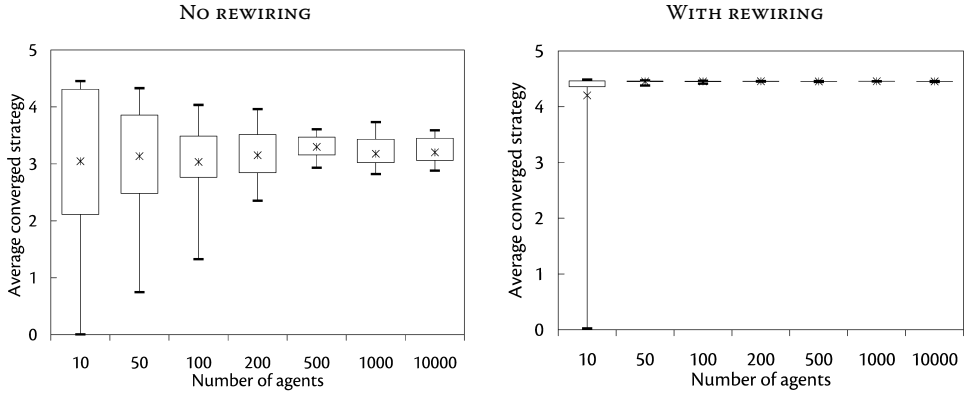


Figure 6.5 Average learned strategy; without rewiring (left) and with rewiring (right)

Learned strategy (Figure 6.5). A setup without rewiring (left) on average converges to a strategy of offering as well as accepting around 3, where 4.5 would be required, as the 33% FS agents present in the population all play with this strategy (i.e., the 66% DS agents on average have a strategy of 2). With increasing population size, this average strategy is not affected; however, it becomes increasingly certainly established. Once again, a setup with rewiring (right) shows convincingly better results. Independent of the population size, the learning agents all converge to the desired strategy, i.e., 4.5.

Performance (Figure 6.6). With a setup without rewiring (left), we already saw that the average learned strategy of the DS agents is not satisfactory. Performance is seriously affected; at around 60%, it indicates that few DS agents ever agree with FS agents. However, average performance is not influenced by the population size. As with the learned strategy, the performance of around 60% only becomes more certainly established. As expected, a setup

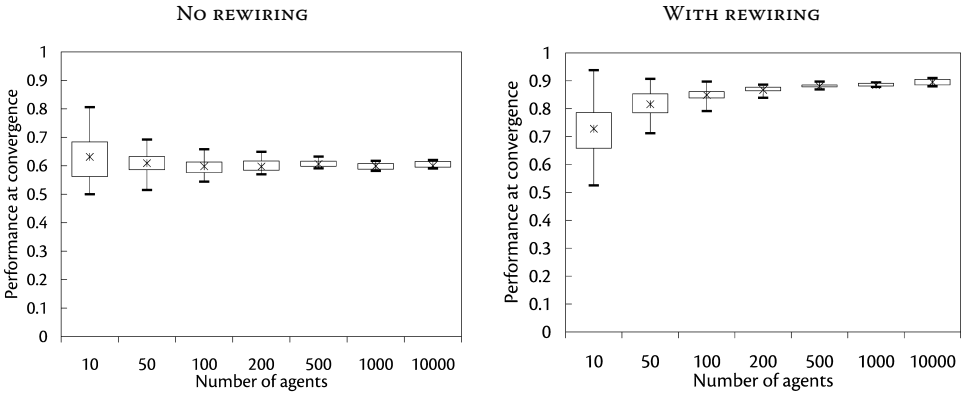


Figure 6.6 Final performance; without rewiring (left) and with rewiring (right)

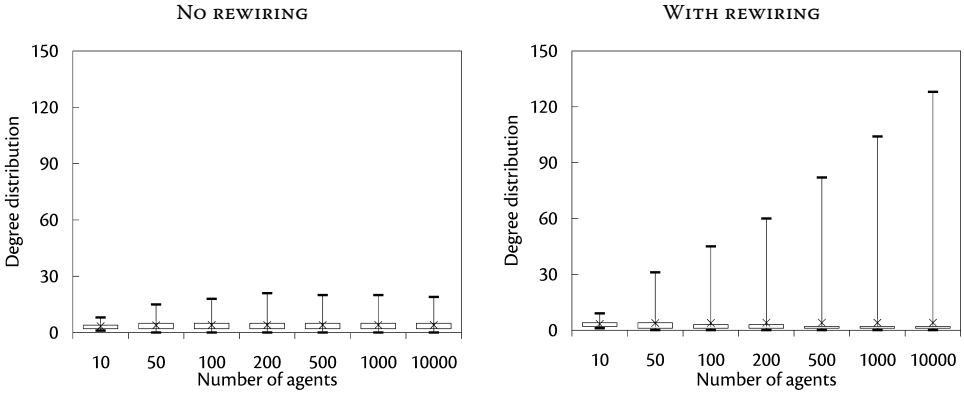


Figure 6.7 Resulting network structure; without rewiring (left) and with rewiring (right)

with rewiring (right) shows much more satisfying results, i.e., generally above 80% agreement. These results are actually positively affected by the population size, as the average performance increases with an increasing population.

Resulting network structure (Figure 6.7). We look at the network structure resulting from learning to reach agreement, and determine whether this structure is influenced by the population size. Obviously, a setup without rewiring (left) does not display any influence here, as the network is static. A setup with rewiring (right) shows an interesting tendency. The average degree of the resulting network remains low, while the maximum degree increases with an increasing population size. Clearly, as the population size increases, the hubs in the scale-free network receive increasingly more preferential attachment, and correspondingly, already less densely connected nodes become even less densely connected. When we examine the number of times agents actually rewire, we find that this number generally lies

below 1,000, i.e., a very low percentage of the total number of games played actually made the agents rewire to a random neighbor of an undesired proposer.

Conclusion. In conclusion to this part of the experiments, we may state that the proposed methodology is not suffering from severe scalability issues. A setup that does not include rewiring is clearly outperformed by one that does include rewiring, but neither a setup without rewiring, nor a setup with rewiring, suffer severely from increasing the number of agents.

Varying the proportion of good examples (FS agents)

Below, we investigate the behavior of the proposed methodology when the proportion of good examples in the population (i.e., FS agents with a strategy of 4.5) is varied. We vary the proportion of FS agents between 0% and 100%, by adapting the probability that a newly generated agent is an FS agent. This implies that the actual number of FS agents in the population varies over individual experiments of 3,000 learning iterations. The remainder of the population consists of DSr and DSh agents in equal proportions. We experiment with a number of population sizes, ranging from 50 to 500.

Since the results for each population size are rather similar, we restrict ourselves to graphically reporting and analysing the results of our experiments with 100 agents in the sequel. A selection of the remaining results is given in Table 6.2. For a setup without rewiring and a setup with rewiring, we report on the population size (Pop), the percentage FS agents used (%FS), the average number of games per agent needed to converge (Games), the average learned strategy (Strat), the average performance (Perf), and the *maximum* number of connections that a single agent has with other agents in the network (Netw). As we will discuss below, the results reported in Table 6.2 for population sizes other than 100 are highly similar to those for a population size of 100 agents.

Point of convergence (Figure 6.8). A setup without rewiring (left) requires increasingly more games per agent to converge, until the proportion of FS agents reaches around 30%. Then, the required number of games decreases again, although there is a great deal of uncertainty. Introducing rewiring (right) yields much better results. The number of games required per agent hardly exceeds 750, and this number decreases steadily with an increasing proportion of the population being an FS agent.

Learned strategy (Figure 6.9). Interestingly, a population consisting of only DS agents tends to converge to offering and accepting the lowest amount possible, both in a setup that does not use rewiring (left), as well as in a setup that does (right). As has been explained in §6.2.6, DS agents tend to adapt their strategies downward more easily than upward. Thus, two DS agents that are having approximately the same strategy, may slowly pull each others' strategy downward. With many DS agents, the probability that this happens increases.

Adding FS agents to the population results in different behavior for the two setups. A setup without rewiring has difficulties moving away from the lowest amount possible; only with a

Table 6.2 Summary of the results of experiments in which the proportion of FS agents is varied. The meaning of the column headers as well as the results for a population of 100 agents are detailed in the text.

No rewiring						With rewiring					
Pop	% FS	Games	Strat	Perf	Netw	Pop	% FS	Games	Strat	Perf	Netw
50	0	663.80	0.01	0.63	15	50	0	639.38	0.01	0.63	22
	30	2,588.50	2.87	0.59	15		30	528.52	4.45	0.81	38
	50	1,800.02	4.12	0.70	16		50	485.60	4.47	0.89	29
	80	259.86	4.47	0.87	15		80	356.34	4.49	0.96	23
200	0	671.30	0.01	0.63	18	200	0	743.00	0.01	0.62	20
	30	2,796.85	2.64	0.57	17		30	540.40	4.45	0.87	52
	50	1,354.80	4.17	0.70	18		50	493.40	4.47	0.91	28
	80	288.35	4.47	0.88	18		80	382.20	4.49	0.97	24
500	0	662.50	0.01	0.64	20	500	0	650.20	0.01	0.65	60
	30	2,793.55	2.85	0.59	20		30	549.95	4.45	0.87	100
	50	1,237.75	4.18	0.69	21		50	498.00	4.47	0.92	55
	80	264.60	4.47	0.89	21		80	380.91	4.49	0.97	35

sufficient number of FS agents (i.e., 30% of the population) does the average learned strategy reflect that the DS agents move towards the strategy dictated by the FS agents. With rewiring, results are different and convincingly better; even with only 10% FS agents, the DS agents on average converge towards offering and accepting the amount dictated by these agents, i.e., 4.5.

Performance (Figure 6.10). The observations concerning the learned strategy, as reported above, are reflected in the performance of the collective of agents. In a setup without rewiring (left), performance decreases initially with an increasing proportion of FS agents, as the DS agents refuse to adapt to the imposed strategy. When the proportion of FS agents becomes sufficiently large, the DS agents start picking up this strategy, resulting in an increasing performance. A setup with rewiring (right) does better, as the performance increases with an increasing number of FS agents. Even though the average learned strategy is close to 4.5 for every proportion of FS agents, low proportions of FS agents still display less performance than higher proportions. This may require an addditional explanation. We note that the box plot of Figure 6.9 shows the distribution of the average strategy over 50 repeated simulations; i.e., it does not show the strategy distribution *within* a single simulation. Thus, even though the average strategy in a single simulation is always very close to 4.5, there is still variance. With a low number of FS agents, this variance is most prominently caused by inertia, i.e., not all DS agents are directly connected to an FS agent, which implies that they need to learn their desired strategy from neighboring agents who are also learning. Especially with rewiring, this may result in two agents playing together that are compatible with most of their neighbors, but not (yet) with each other.

Resulting network structure (Figure 6.11). Clearly, the network structure of a setup without

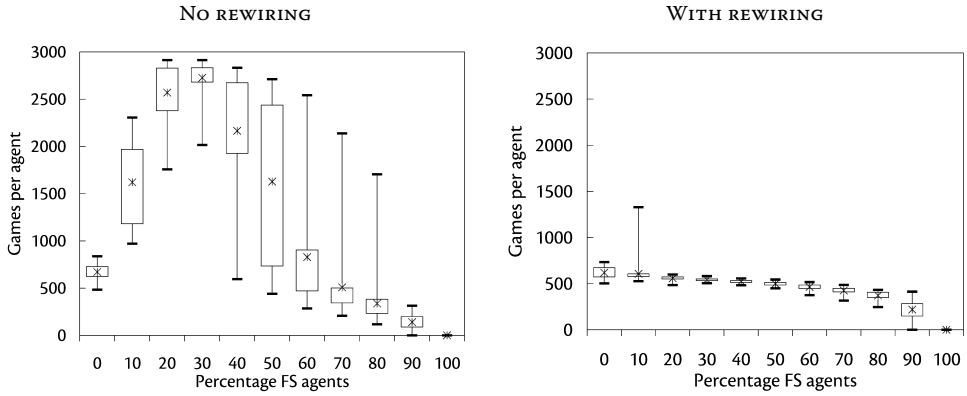


Figure 6.8 Games per agent until convergence; without rewiring (left) and with rewiring (right)

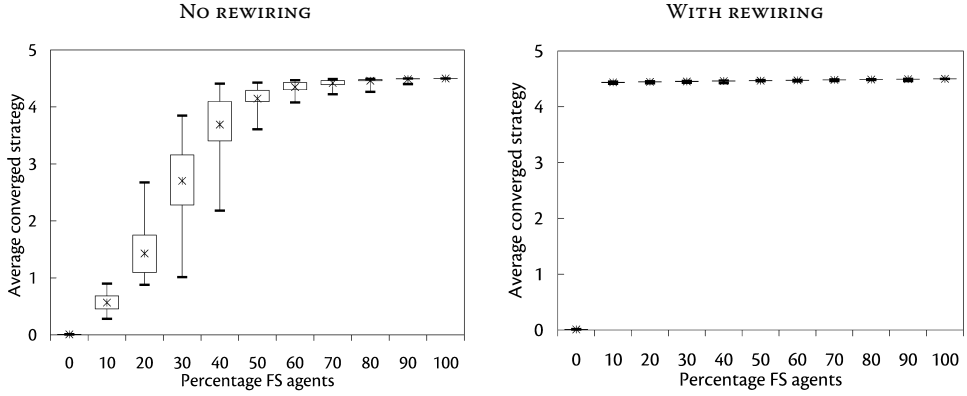


Figure 6.9 Average learned strategy; without rewiring (left) and with rewiring (right)

rewiring (left) is not influenced by varying the proportion of FS agents. When rewiring is used (right), we observe an interesting phenomenon, closely related to our observations in the experiments concerning population size. Once again, the number of times agents actually rewire generally lies below $1,000$. Even though this is a low number, it does affect the network structure in a useful way. With a low proportion of FS agents, there is a large tendency for increased preferential attachment. For instance, with 10% FS agents, there is a single agent that connects to 70 out of 100 other agents. With an increasing proportion of FS agents, the maximum degree of the network decreases, until finally, it closely resembles the original network. Clearly, in the presence of only few examples of the desired strategy, DS agents attempt to connect to other agents that provide such examples. They show interesting and useful emergent behavior.

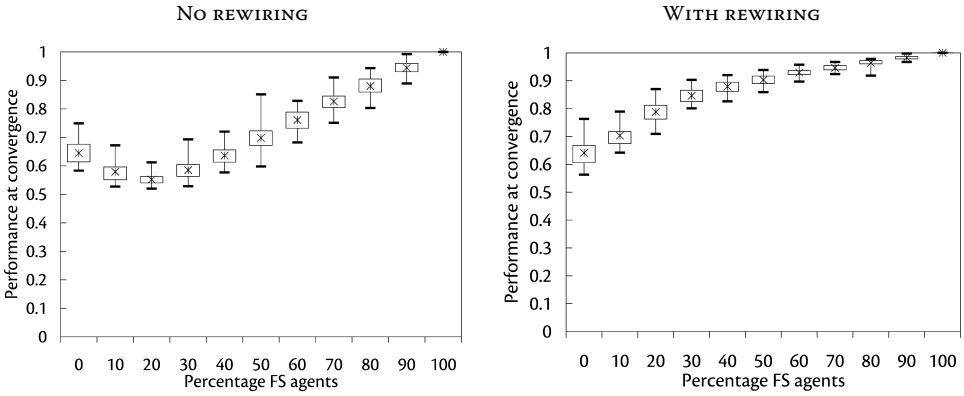


Figure 6.10 Final performance; without rewiring (left) and with rewiring (right)

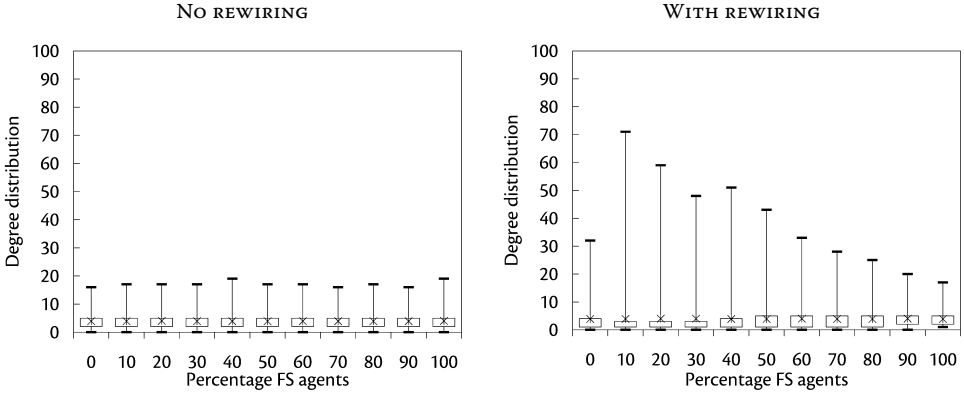


Figure 6.11 Resulting network structure; without rewiring (left) and with rewiring (right)

When we compare the results obtained in a population of 100 agents with the results for other population sizes, as reported in Table 6.2, we see that these are highly similar.

Conclusion. In conclusion to this part of our experiments, we may state that a setup that is not using rewiring has severe difficulties converging to a desired example if the proportion of FS agents providing this example is low. Only for, e.g., half of the population consisting of examples, does the other half learn the desired behavior. A setup that is using rewiring has absolutely no problems converging to the desired strategy, even with a low proportion of FS agents. In both cases, completely omitting the examples leads to the agents converging to the individually rational solution. This is caused by an artifact of the learning method used, i.e., as mentioned before, two CALA trying to learn each others' strategy tend to be driven downward to the lowest value allowed.

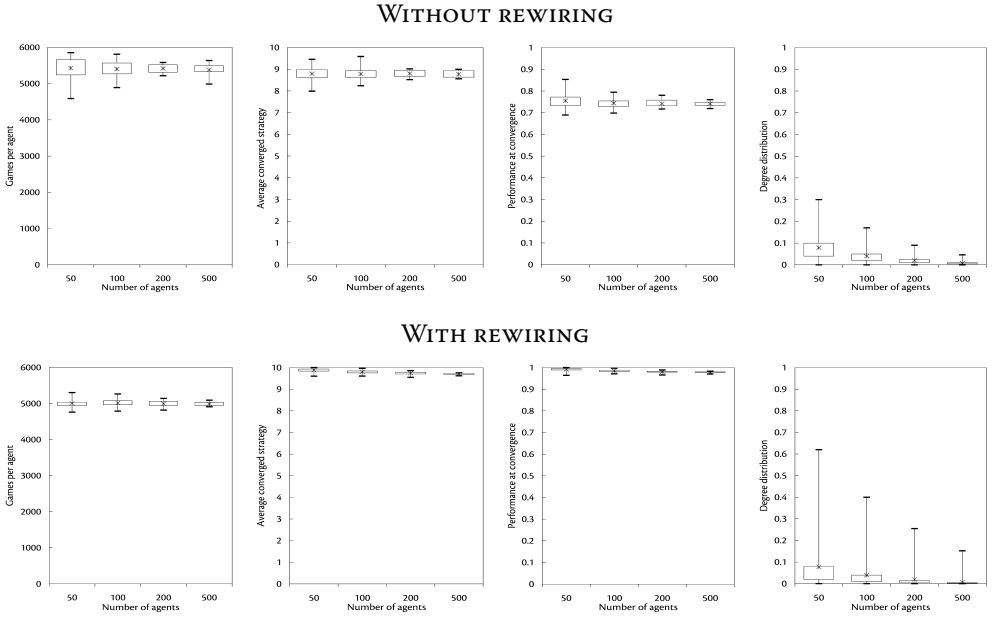


Figure 6.12 Influence of population size in the Public Goods Game

6.4.2 The tragedy of the commons

As with the experiments concerning the UG, our experiments concerning the PGG are presented in two subsections. First, we investigate the influence of varying the population size on the quantities of interest, using a fixed proportion of 33% DSr agents, 33% DSh agents and 33% FS agents. Second, we investigate the influence of varying the proportion of good examples (FS agents), while keeping the other two types in equal proportion.

Varying the population size

Experiments were performed with populations of 50, 100, 200, and 500 agents. Results are shown in Figure 6.12. We compare a setup with a static population structure (without rewiring) to a setup with a dynamic structure (with rewiring).

Point of convergence. Concerning the number of games per agent until convergence, we see that introducing the option to rewire allows agents to reach convergence with approximately 10% fewer games, e.g., 5,000 instead of 5,500. This number is not strongly influenced by the size of the population.

Learned strategy. We observe a good result in a static population structure (i.e., around 9, where 10 would be desired), and an even better result in a dynamic structure.

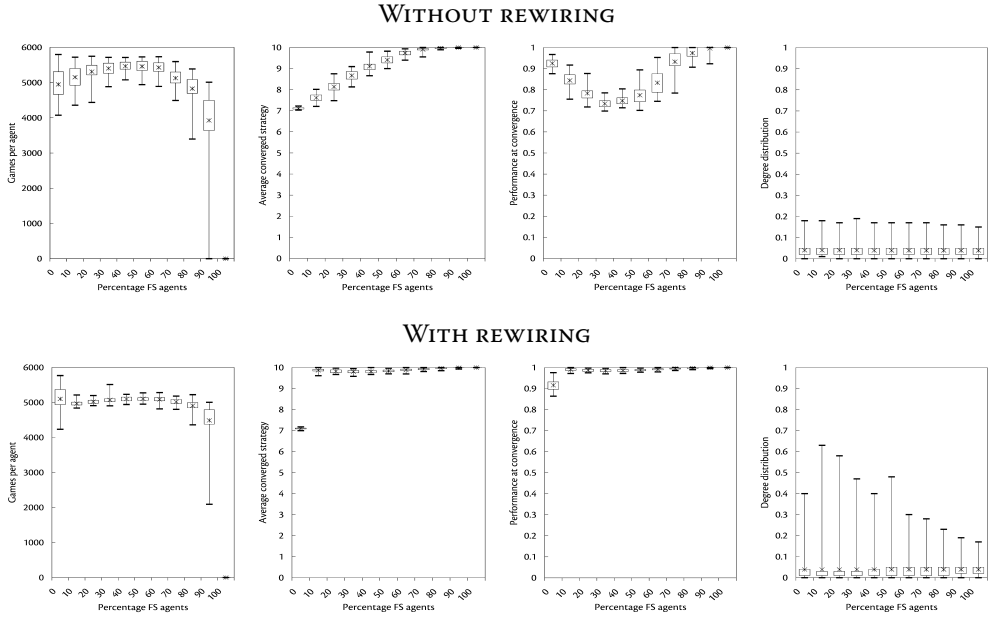


Figure 6.13 Influence of the proportion of fixed-strategy agents in the Public Goods Game

Performance. The performance shows similar characteristics; without rewiring, there is still quite some disagreement on strategies, as only around 75% of the neighbors have a common strategy, whereas with rewiring, there is hardly any disagreement.

Resulting network structure. Looking at the network structure, we may observe that rewiring increases the preferential attachment already present in the scale-free network; very few agents become rather densely connected (e.g., with 50 agents, we observe one agent being connected to more than 30 others), whereas most agents remain sparsely connected or even lose some connections. The performance increase caused by rewiring is even more surprising if we study the number of times agents actually rewired; this turns out to be quite low, i.e., it happens after less than 1% of the games played.

Conclusion. In conclusion to this experiment, we may state that the size of the population is not a strong influence on the quantities of interest. In all cases, a static population reaches a performance of around 75%, whereas a dynamic population achieves nearly 100%.

Varying the proportion of good examples (FS agents)

Once again, we perform experiments in populations of 50, 100, 200, and 500 agents. We vary the proportion of FS agents between 0% and 100%, by adapting the probability that a newly generated agent is an FS agent. This implies that the actual number of FS agents in the

population varies over individual experiments of 6,000 learning iterations. We repeat every experiment 50 times. For the sake of brevity, we only report results of our experiments with a population of 100 agents. The results for other population sizes are highly similar to those reported here. The results for 100 agents are shown in Figure 6.13. We compare a setup with a static population structure (as given in the figure below 'without rewiring') to a setup with a dynamic structure ('with rewiring').

Point of convergence. When we look at the number of games per agent until convergence is reached, we see that populations with a low or high percentage of FS agents converge slightly more quickly than populations with a percentage in between, e.g., 40%. Introducing rewiring reduces the number of games needed by approximately 10%.

Learned strategy. When we look at the average converged strategy, we may observe that (1) a population without any FS agents converges to the most cooperative strategy present, i.e., the 7 of the DSh agents; (2) adding FS agents to the population allows the DS agents to learn an even more cooperative strategy, with much better results for a population that is allowed to rewire.

Performance. Once again, the performance measures reflect the quality of the average learned strategy. Interestingly, with an increasing proportion of FS agents, the performance of a static population initially decreases. If we look at the average learned strategy, we can see why this is; the averages reported reflect that the DS agents learn to contribute 7 at least until there are around 40% of FS agents. With more FS agents, the DS agents learn to contribute more than 7, making them more compatible with the FS agents. When we allow agents to rewire, a low percentage of FS agent is already sufficient to achieve full cooperation.

Resulting network structure. Looking at the network structure resulting from rewiring, as compared to the static network, we see an interesting phenomenon: with a low percentage of FS agents, a few agents show a drastic increase in their number of connections to other agents. For instance, with 10% FS agents, there is a single agent that is connected to 63 of the 100 agents. Clearly, if this single agent is an FS agent, connecting to it is useful, as it allows a DS agent to learn the desired strategy quickly. Once again, the number of times agents actually rewire, is low.

Conclusion. In conclusion to this experiment, we may state that results for the tragedy of the commons are similar to those for the agreement dilemma, except for the notable fact that a population without any FS agents converges to the most cooperative strategy present. This actually implies that FS agents will not be necessary to facilitate cooperation. As long as we have a sufficient number of agents that are playing cooperatively initially, we may reach a fully cooperative outcome.

6.5 Discussion

The results presented in §6.4 suggest that a number of mechanisms that lead to cooperative solutions in social dilemmas with only a discrete set of strategies (e.g., scale-free networks and rewiring), also lead to agreement and cooperation in social dilemmas with continuous strategy spaces.

In this section, however, we show that this is not a trivial issue. More precisely, we discuss a number of mechanisms, as described in Chapter 5, that may enhance the agents' abilities to reach cooperation in social dilemmas with discrete strategy sets, but do not directly enhance the agents' abilities to reach agreement and cooperation in our experiments here, which are using continuous strategy spaces. We empirically analyze why this is the case.[7]

6.5.1 Reputation

Reputation is one of the main concepts used in behavioral economics to explain how fairness emerges (e.g., Bowles et al., 1997; Fehr, 2004). Basically, it is assumed that interactions between people lead to expectations concerning future interactions. These expectations may be positive or negative and may be kept to oneself, or actually shared with peers. If players know about each others' reputation, they may change their strategies, e.g., by offering low amounts to other players that are known to be (relative) defectors; it then becomes less attractive to be a defector.

In work closely related to our work, Nowak et al. (2000) show that reputation deters agents from accepting low offers in the UG. The information that a certain agent accepts low offers will disseminate, and therefore, other agents will provide only low offers to this agent. If, due to this reasoning, all agents refuse to accept low offers, they should provide high offers. Thus, Nowak et al. (2000) argue that the population goes toward providing and accepting high offers. However, we note that any shared strategy (i.e., any agreement) in the UG yields an expected payoff of 50% of the amount at stake for both agents. Thus, reputation may indeed help agents to decide which strategy to play against others, but a preference for playing cooperatively (i.e., providing high offers) does not directly result from reputation.

Spreading reputation

We study the effects of reputation in the UG by optionally adding a second network to our system. As with the interaction network, we consider the reputation network to be scale-free. However, in contrast to the interaction network, the reputation network is assumed to be static, as agents are truthful concerning reputation, making it unnecessary for agents to consider rewiring. Note that two agents sharing reputation information may be connected as well in the interaction network, and as a consequence, two agents playing an UG may

[7] We note that we only discuss the agreement dilemma here. In the tragedy of the commons, the underlying issues with the alternative mechanisms are equally present.

share reputation information with each other. Generally, we can assume that this is not the case.

In effect, after every UG, the responding agent may broadcast reputation information to its neighbors in the reputation network. The information is sent by the responder and concerns the offer just done by the proposer, as this is the only information that is guaranteed to be correct. Agents receive information with a probability:

$$p^{ij} = 1 - \frac{d}{H}. \tag{6.5}$$

Here, d is the distance between the sender and the (potential) receiver j in the reputation network. Thus, reputation information may travel for at most H hops, with a decreasing probability per hop. In our simulations, we set $H = 5$.

We note that reputation information may be helpful only if we allow agents to *do* something with this information. In the aforementioned work by Nowak et al. (2000), reputation is used by agents to determine what to offer to others. Given (1) the observation that reputation, used in this way, should not necessarily promote cooperative strategies (see above), and (2) the fact that we already use CALA to determine what agents offer to each other, we want the reputation to affect something else than agents' strategies.

We will discuss two ways in which agents may use reputation, as taken from literature, i.e., interacting with a preferred neighbor (below), and using reputation to facilitate voluntary participation (§6.5.3).

Using reputation

Without reputation, agents play against a random neighbor in the interaction network. Reputation may be used to make agents prefer interacting with specific neighbors. Chiang (2008) discusses that strategies of fairness could evolve to be dominant if agents are allowed to choose preferred partners to play against. Chiang (2008) allows agents to select partners that have helped the agent previously.

To determine who is a preferred partner, we use the heuristic proposed by Santos et al. (2006c), i.e., an agent prefers playing with (relative) cooperators, as these help it in obtaining a high payoff if it is the responder. Thus, the probability that agent i plays with agent $j \in N^i$, where N^i is the set of agent i's neighbors, is:

$$p^{ij} = \frac{\tilde{\mu}^j - \mu^i}{\sum_{k \in N^i} \tilde{\mu}^k}. \tag{6.6}$$

Here, μ^i, $\tilde{\mu}^j$ and $\tilde{\mu}^k$ are the agents' current strategies (for agents other than i, these are estimates based on reputation and previous experience).

There are two problems with this approach. First, the number of times that an agent i receives information about an agent $j \in N^i$ may be rather low, especially with many agents.

Even with only 50 agents, we observe that only around 25% of the reputation information received by agents actually concerned one of their neighbors. This problem may be addressed by making the reputation network identical to the interaction network (as neighbor relations in both networks are then identical). However, this may be seen as a considerable abstraction. Second, the probability that agent i has information concerning all of his neighbors is low, so we need to specify default values for $\tilde{\mu}^j$. Clearly, any default value is more often wrong than right, unless we use a centralized mechanism to estimate it by, for instance, using the current average population strategy, which is what we do in our simulations.

With this mechanism in place, we perform the same experiments as in §6.4.1, i.e., we vary the population size between 10 and 10,000 agents, and the proportion of FS agents in steps of 10%. A statistical analysis reveals no significant difference between a setup that uses reputation and a setup that does not. When we further analyse the results, we see that, as expected, agents almost always need to resort to default values for their neighbors' strategies. Thus, on average, the reputation system does not often change the probabilities that certain neighbors are selected. Similar results may be expected for the PGG.

6.5.2 Reputation and rewiring

As we have seen in §6.4.1, rewiring works well without reputation (i.e., purely based on an agent's own experience). Adding reputation may be beneficial to agents, as they no longer need to interact with each other to be allowed to unwire. Thus, agents may further increase their preference for certain others.

Reputation information (i.e., the amount offered by a certain agent) propagates through the (static) reputation network, allowing agents receiving such information to unwire from one of their neighbors if they consider this neighbor's behavior to be undesirable. The same rewiring mechanism is used here as detailed in §6.2.7 (i.e., Equation 6.4). We allow the responder in the UG to broadcast reputation information through the reputation network, with a maximum of $H = 5$ hops.

As above, we perform the same experiments as in §6.4.1. Again, there is no significant difference in the main results. We analysed the number of times agents actually rewired, and found that this number on average increases by a factor 2 with respect to a setup in which reputation is not used. However, this increase does not increase performance. On average, agents have only a few neighbors; thus, they generally receive reputation information concerning a neighbor that, in the absence of reputation, they would play against soon anyway. We note that results in the PGG would be similar.

6.5.3 Volunteering

According to existing research on human fairness (e.g. Sigmund et al., 2001; Boyd and Mathew, 2007; Hauert et al., 2007) the mechanism of *volunteering* may contribute to reaching cooperation in games with only two strategies. The mechanism of volunteering consists

in allowing players not to participate in certain games, enabling them to fall back on a safe 'side income' that does not depend on other players' strategies. Such risk-averse optional participation can prevent exploiters from gaining the upper hand, as they are left empty-handed by more cooperative players preferring not to participate. Clearly, the 'side income' must be carefully selected, such that agents are encouraged to participate if the population is sufficiently cooperative. Experiments show that volunteering indeed allows a collective of players to spend "most of its time in a happy state" (Boyd and Mathew, 2007) in which most players are cooperative. For a broader overview, we refer to Chapter 5.

In this chapter, with the mechanism of rewiring, we basically implemented a way for agents to avoid having to interact with others without explicitly giving up any reward; since (1) we play a fixed number of games between random agents and their neighbors, and (2) every unwiring action of a certain agent is immediately followed by rewiring to another agent, agents do not negatively influence the number of games they play by rewiring. In contrast, the biggest problem when applying volunteering is that we basically do introduce yet another social dilemma; an agent may refrain from participating to make a statement against the other agent, which may convince this other agent to become more social in the future, but to make this statement, the agent must refuse an expected positive payoff. For example, in the UG with randomly assigned roles, the expected payoff is always positive.

Nonetheless, we investigate whether volunteering promotes agreement in games with continuous strategy spaces. We once again use the heuristic proposed by Santos et al. (2006c), which has already been applied in various mechanisms in this chapter: if agent i believes agent j is a (relative) cooperator, then he agrees on playing. If both agents agree, then a game is played. To prevent agents from not playing any game (after all, both agents see each other as a relative cooperator only if they already are playing the same strategy), we introduce a 10% probability that games are played anyway, even if one or both agents does not want to. Reputation may be used here, as it may allow agents to estimate whether one of their neighbors is a relative cooperator or not, without having to play with this neighbor.

However, experimental results point out that agents using volunteering (with and without reputation) have severe difficulties establishing a common strategy (Uyttendaele, 2008). As a result, when measuring performance, we see that only around 50% of the games is played. Of the games played, the performance is similar to a setup with rewiring (e.g., above 80%), which may be expected, as two agents usually only agree to play if their strategies are similar. The reason why agents do not converge properly is quite simple: they avoid playing with other agents that are different from them. Therefore, they do not learn to behave in a way more similar to these others.

6.5.4 General discussion

In general, we may state that with the mechanism of rewiring, we clearly find a good balance between allowing agents to play with more preferred neighbors on the one hand, and forcing agents to learn from those different from them on the other hand. The additions discussed

above allow agents to be too selective, i.e., they have too much influence on the choice of who they play against. While this may be in the interest of individual agents, it generally leads to agents not playing against others that are different from them, instead of learning from these others, as is required to obtain convergence to agreement.

6.6 Chapter conclusion

In conclusion to this chapter, we answer our research questions RQ1 and RQ3–RQ5, as in the previous two chapters.

RQ1 *How are humans using fairness in their decisions?*

In this chapter, we discussed two contributions with regard to this research question. First, existing work has shown that the fact that society is structured in a complex manner, may help humans to find and maintain cooperative strategies (e.g., in social dilemmas). Humans may decide to stop interacting with certain others if they feel these others to be overly defective. This results in a dynamic population structure. Second, we dealt with the observation that altruistic punishment is not equally performed by humans for actions that are 'really bad' as for actions that are only 'quite bad'. We introduce the mechanism of probabilistic punishment to model this.

RQ3 *How can human-inspired fairness be modeled computationally, taking into account the foundations of human fairness?*

In this chapter, we modeled the human tendency to interact in complex networks by using a scale-free network, based on which agents perform pair-wise interactions. Optionally, agents may decide to break the link between them and an undesirable neighbor, which facilitaties a dynamic population structure. Additionally, we introduced probabilistic punishment in the tragedy of the commons.

RQ4 *What are the analytical properties of the computational human-inspired fairness models developed in this research?*

A careful analysis of (dynamic) social networks shows why these networks may be beneficial for agents trying to learn cooperative solutions. Allowing agents to stop interacting with others that are overly defective, clearly is an effective way of isolating defectors. The probability that a cooperative agent is interacting with a defective one decreases, and therefore the probability that the cooperator is tempted to start defecting, also decreases. Concerning probabilistic punishment, we analyze that this mechanism allows agents to reverse the (undesirable) inverse relation between individual contribution and individual reward in the PGG.

RQ5 *What are the (empirical) benefits of incorporating explicitly human-inspired fairness in adaptive multi-agent systems?*

We may give the following four statements here.

First, in the UG, our proposed methodology is able to establish agreement on a common strategy, especially when agents are given the option to rewire in their network of interaction. Humans playing the UG reach an agreement of approximately 80 to 90% (Oosterbeek et al., 2004). Without rewiring, our agents do worse (generally, 65% of the games are successful); with rewiring, they do as well as humans.

In the PGG, combining the proposed methodology with probabilistic punishment allows agents to find and maintain fully cooperative solutions. Without any FS agents, the most cooperative strategy is played by DSh agents, i.e., contributing 7. In Figure 6.13, we see that a population with no FS agents (and 50% DSr agents contributing 0 as well as 50% DSh agents), indeed converges to contributing 7, regardless of whether agents may rewire. With FS agents present, the population converges to 10 (the FS agents' strategy), with distinctly less difficulty for a population that is allowed to rewire in their network of interaction.

Thus, as in games with a discrete strategy set, rewiring greatly enhances the agents' abilities to reach agreement or cooperation, without compromising the scale-free network structure. This indicates that interactions in scale-free networks, as well as rewiring in these networks, are plausible mechanisms for making agents reach agreement and cooperation.

Second, in comparison to methodologies reported on in related work (e.g. Santos et al., 2006b), our methodology facilitates convergence with only a low number of games per agent needed (e.g., 500 instead of 10,000 in the UG). This indicates that CALA are a satisfactory approach when we are aiming at allowing agents to learn from a relatively low number of examples. The benefit of CALA is even larger if we considering the fact that learning to play in a continuous strategy space is obviously markedly more difficult than learning to play one of two possible actions.

Third, the performance of the collective is not seriously influenced by its size. This is clearly influenced by the characteristics of a scale-free, self-similar network.

Fourth, concepts such as reputation or volunteering, which have been reported to facilitate cooperative outcomes in discrete-strategy games, do not seem to have (additional) benefits in continuous strategy spaces.

In general, we may conclude that introducing population structure in a collective of agents greatly enhances their ability to find and maintain desirable, fair solutions, especially if we allow agents to change the structure of the population. A large number of agents (i.e., thousands) can successfully reach agreement in the agreement dilemma, and can successfully find the best (cooperative) strategy in the tragedy of the commons.

7 Conclusions

In this chapter, we present the conclusions of the thesis. We answer our research questions and provide an answer to the problem statement (§7.1). We also discuss our results in a broader perspective by giving ideas for future work (§7.2).

7.1 Answers to the problem statement and research questions

This thesis started with three observations from existing literature. In §1.1, we observed that humans are able to find good solutions to a set of difficult problems called *social dilemmas*, because they do not only decide based on their personal benefit (i.e., a decision based on individual rationality), but also based on the benefit of others (i.e., a decision based on fairness). Moreover, they are willing to punish those who are trying to exploit their care for fairness. In §1.2, we observed that many multi-agent systems have to deal with tasks in which social dilemmas are somehow present, most prominently the task of allocating resources to different agents within the system. In §1.3, we observed that current work concerning fairness in multi-agent systems is not explicitly inspired by human decision-making, which makes current implementations of fairness both insufficiently fair from the human perspective, as well as vulnerable for exploitation by agents who do not care for fairness.

Given these three observations, we argued that multi-agent systems will benefit from the inclusion of explicitly human-inspired fairness. Our problem statement (PS) therefore read as follows.

PS *How can we obtain human-inspired fairness in multi-agent systems?*

In §1.4, we provided five research questions, i.e., RQ1 to RQ5, aimed at addressing this problem statement. Then, in §1.5, we indicated that answering these five questions required us to perform eight steps in our research. RQ1 required four such steps, and RQ2 to RQ5 required one step each. The research questions follow below.

RQ1 *How are humans using fairness in their decisions?*

1. Literature study. We started by performing an extensive literature study in the field of behavioral economics. In this field, experiments with humans are conducted; based on observations, *descriptive models* of human fairness are constructed. We found that the most prominent human fairness mechanisms are (1) altruistic punishment, and (2) withholding action. Ad (1), altruistic punishment implies that we are willing to sacrifice some of our reward in order to reduce the reward of someone that behaved in an unfair manner. Ad (2), withholding action implies that we are willing to refrain from interacting with someone that we believe to behave in an unfair manner, even if the expected reward is positive. The individually rational decision would be *not* to perform altruistic punishment and withhold action, as both mechanisms do not yield immediate personal benefit.

Researchers have conducted numerous experiments, aimed at explaining why humans execute these two mechanisms, even though they are not motivated by rationality. The three most important explanations found in the literature are modeled in three descriptive models, i.e., (i) inequity aversion, (ii) reputation, and (iii) social networks. Ad (i), the inequity aversion model (see §4.1) addresses the observation that humans tend to dislike large differences in rewards, with an emphasis on disadvantageous differences. Ad (ii), the reputation model (see §5.1) focuses on repeated interactions and shows that humans tend to be reciprocal, i.e., they are nice to others that they know to be nice to them, and willing to punish

others that are somehow offensive, even if this punishment is costly. Reputation emerges over time and helps humans to identify who is nice and who should be punished. Ad (iii), the social-network model (see §6.1) addresses the phenomenon that human interactions are not random, but typically take place based on complex (e.g., scale-free) network topologies, which may contribute significantly to the emergence of fair, cooperative behavior.

2. Analysis. In our literature study, we identified two possible opportunities. First, we analysed the inequity-aversion model, which has already been used to explain human behavior in many interactions (modeled in the form of matrix games). It may also explain human behavior (i.e., the tendency to perform altruistic punishment) in the Public Goods Game (see §4.2.2 for an in-depth analysis). Second, we analysed two additional existing models (as discussed above) and found that the three existing models are missing an important element, i.e., *priority*, which indicates that human decision-making is influenced by additional information about others. This information may emerge over time (e.g., reputation), but may also be immediately available (e.g., physical appearance, explicit information concerning the wealth of others). Such additional information allows humans to give priorities to each of the others they interact with. We give a number of examples of interactions in which priorities play a role (see §5.2).

3. Design. In relation to the second opportunity (i.e., priority) described directly above, we design a new model of human fairness, which is an extension of the existing inequity-aversion model, called *priority awareness* (see §5.3).

4. Validation. Using analyses and experiments, we validated to what extent the priority-awareness model adequately predicts human behavior in various settings. We established that, in contrast to the inequity aversion model, priority awareness at least allows us to give a qualitative prediction of human behavior (see §5.4).

After these four steps, we may answer the first research question. In summary, human decisions are not based only on individual rationality, but also on fairness. Their decisions are most notably based on fairness whenever they decide to punish someone (at a cost to themselves) and whenever they decide to withhold action (even though performing an action may lead to a positive reward). We identified three existing descriptive models that clarify what these decisions may be based on, and developed a new model ourselves. Thus, we ended up with four descriptive models, i.e., in the order of being discussed in the thesis, (1) inequity aversion, (2) reputation, (3) priority awareness, and (4) social networks. The third model (priority awareness) is new. We show that it may be used to model both static priorities (i.e., information immediately known) as well as dynamic priorities (e.g., reputation). Thus, we were able to model (2) and (3) in a single descriptive model. The thesis therefore focuses on three descriptive models and devotes one chapter to each model.

The four remaining research questions, RQ2 to RQ5, are addressed by performing one methodological step per question. Each of these questions concerns the issue of transforming the descriptive models, as found by answering RQ1, to computational models of human-inspired fairness. The first of the four research questions reads as follows.

RQ2 *What are the foundations of human-inspired computational fairness?*

5. Formalization. In answer to RQ2, we formalized the foundations of human-inspired computational fairness, which consist of two parts, i.e., we give *requirements* as well as a *template model* for computational models of human-inspired fairness.

The requirements were based on a study of the relevant background knowledge on (general) computational models, i.e., game theory and reinforcement learning (see Chapter 2). In summary, in §3.1, we stated that (R1) any computational model must be based on game-theoretic principles, since game theory is a well-known and established formalism to describe interactions between multiple agents. Moreover, (R2) any computational model must be applicable to a multi-agent system in which agents learn to behave according to reinforcement learning. Finally, (R3) any computational model of human-inspired fairness must, of course, be human-inspired. With respect to this latter requirement, we specified what we mean by 'human-inspired', i.e., agents must be able to determine (R3-Q1) the fairness of an interaction, to determine (R3-Q2) whether altruistic punishment is appropriate, and to determine (R3-Q3) whether withholding action is appropriate.

Our template model is centered around a *utility function*, which satisfies the first two requirements (i.e., R1 and R2), as the concept of a utility function is known in game theory as well as reinforcement learning. In fulfilment of the third requirement (R3), we showed how a utility function may be used to allow agents to decide upon fairness, altruistic punishment, and withholding action (see §3.2). To embed our work properly in the existing literature, we related our work to published work concerning computational fairness (see §3.3).

In the remainder of the thesis, we proposed actual utility functions that may be used. Thus, RQ3 read as follows.

RQ3 *How can human-inspired fairness be modeled computationally, taking into account the foundations of human fairness?*

6. Design. We designed three computational models of human-inspired fairness, based on the three descriptive models found earlier. In the case of inequity aversion (see §4.3), the descriptive model already gave us a utility function, called *Homo Egualis*. For this rather rugged utility function to be applicable computationally, a number of small modifications to our reinforcement learning method (CALA) were presented. For priority awareness (see §5.5), we followed a similar approach.

The social network model is quite different from the other two models, as it describes how fairness emerges from factors *external* to the agent (see §6.2). The agents themselves therefore are not equipped with an 'advanced' utility function, but with a straightforward threshold function. The manner in which agents interact, especially when coupled with the possibility to refrain from interacting with certain others (i.e., withholding action), provides the main incentive to care for fairness.

RQ4 *What are the analytical properties of the computational human-inspired fairness models developed in this research?*

7. Analysis. In order to derive the analytical properties of our three models, we extensively analysed the utility functions and external factors they are based on. Most prominently, we determined expected outcomes obtained by a collective of agents in social dilemma interactions. For inequity aversion (see §4.2) and priority awareness (see §5.4), we calculated the expected outcomes for all social dilemma games under study. For social networks (see §6.1), we looked at similar analytical properties, as reported in the literature, which support the claim that interactions in networks may enhance agents' abilities to care for fairness.

RQ5 *What are the (empirical) benefits of incorporating explicitly human-inspired fairness in adaptive multi-agent systems?*

8. Experiments. Using multi-agent learning algorithms, i.e., learning automata, and our three computational models of fairness, we constructed three adaptive multi-agent systems that use a computational instantiation of human fairness. We performed experiments with these multi-agent systems in order to determine whether they are able to address the problems in which we are interested, i.e., social dilemmas.

For all three models, we established convincing results. First, in the case of inequity aversion (see §4.3), we found that agents learned satisfactory solutions to most of the social dilemmas we investigated, except for those dilemmas in which priorities played a role. Second, priority awareness (see §5.5) was successfully applied to social dilemmas in which priorities were introduced. It was shown that these priorities may change over time (for instance, if they represent reputation information). Third, the social network model (see §6.2), improved upon the results found by the first two models. The fact that agents were interacting in structured populations allowed us to use many more agents than in unstructured populations. Moreover, adding the opportunity for agents to restructure their interaction network enhanced their abilities to find good solutions to all types of social dilemma under study.

Thus, the answer to RQ5 reads as follows. The (empirical) benefits of incorporating explicitly human-inspired fairness in adaptive multi-agent systems are that (1) this incorporation allows a collective of individually learning agents to find and maintain desired, cooperative solutions to social dilemmas, and (2) they can do this even in the presence of agents that are purely self-interested.

In conclusion, after this last step, we have answered all five research questions. Thus, we are now able to provide an answer to the problem statement, which is repeated here for convenience.

PS *How can we obtain human-inspired fairness in multi-agent systems?*

The thesis proposed an answer to the problem statement, which may essentially be summarized in four steps. First, we investigated human fairness by means of experiments with human subjects. Second, based on experimental observations, we turned to *descriptive models* of human fairness. The main aim of descriptive models is to describe why humans decide to

discipline each other by performing the (successful) mechanisms of altruistic punishment and/or withholding action, even if these mechanisms are both not individually rational. We presented an overview of existing descriptive models, and constructed a new descriptive model based on our own experiments. Third, we discussed how descriptive models may be translated to *computational models* of human-inspired fairness. In order to be useful in a multi-agent system, such models must meet a number of requirements, which we outlined in the thesis. Fourth, after establishing these requirements, we actually translated three descriptive models to computational models of human-inspired fairness, thus including in multi-agent systems a wide range of motivations to perform altruistic punishment and/or withhold action. Multi-agent systems that are equipped with our computational models are shown to be able to find satisfactory outcomes in interactions that are difficult (or even impossible) to address with existing approaches. Moreover, since our fair agents are willing to punish those that behave in an unfair manner, the system as a whole encourages agents (including those not designed by us) to refrain from exploiting others.

7.2 Ideas for future work

In relation to other work concerning fairness, the work reported in the thesis lies somewhere in the middle between theory and practice. On the one hand, we see a great deal of theoretical work, mostly either pursuing experiments with humans, with the aim of constructing descriptive models of human fairness (e.g., Fehr and Schmidt, 1999; Fehr and Gaechter, 2002; Bearden, 2001; Oosterbeek et al., 2004), or pursuing evolutionary-game-theoretic analyses or multi-agent learning, with the aim of validating proposed descriptive models (e.g., Nowak et al., 2000; Santos et al., 2006c). On the other hand, there is a great deal of practical work, mostly pursuing the construction of computational models, aimed at developing multi-agent systems for specific, practical tasks, e.g., resource distribution, load balancing, auctioning, or scheduling (e.g., Chevaleyre et al., 2006; De Jong et al., 2006c; Mao et al., 2006; Verbeeck et al., 2007).

That the work reported on here is indeed in the middle between theory and practice follows from the aim of the work, i.e., to reduce or even bridge the gap between descriptive models of human fairness and computational models, resulting in computational models of human-inspired fairness. Clearly, although this thesis may have achieved a significant reduction of the gap, there still is work to do. Essentially, we placed our work in the middle of the gap, leaving two (smaller) gaps on either side.

In conclusion to this thesis, we will give an overview of research that may be performed in order to reduce these smaller gaps. We will refer to the gaps as the 'gap between theory and current work' and the 'gap between between current work and practice'.

7.2.1 The gap between theory and current work

In the current work, we investigated and developed three computational models of human-inspired fairness. These models have been analysed and implemented in a multi-agent system. We note that a more thorough analysis of the models is possible, and possibly desirable. Below, we provide three ideas.

Using evolutionary game theory. As mentioned in this section's introduction, many researchers have used evolutionary game theory to study attractors produced by certain mechanisms (e.g., by means of replicator dynamics; Nowak et al., 2000; Sigmund et al., 2001). Our approach, i.e., using learning automata, has been shown to yield equivalent results in the case of a small, discrete set of strategies (Hennes, 2008). A similar equivalence between learning automata and replicator dynamics in the continuous case is currently being investigated (Tuyls and Westra, 2009). Learning automata's learning traces, or equivalently, the underlying mechanisms' replicator dynamics, may be plotted in a highly informative figure (a so-called simplex; see Weibull, 1996). In such a figure, we may observe how agents switch strategies, based on their utility, as well as how this utility compares to that of others. Thus, instead of empirically showing the efficacy of our proposed models (i.e., by showing what agents learn as a result of these models), we may obtain even stronger, analytical results (i.e., replicator dynamics) (Tuyls and Parsons, 2007). Since replicator dynamics are usually applied in interactions with at most three discrete strategies, work needs to be done in order to apply them to our models (which use continuous strategy spaces, and therefore an infinite number of strategies).

A more thorough look at priority awareness. In Chapter 5, we use a number of experiments with humans to motivate the development of a new descriptive model of human fairness, called priority awareness. The attentive reader may have noted that the subsequent application of the model to the experiments' main observations was generally qualitative rather than quantitative. The main problem here seems that human priorities are actually also qualitative, or at least definitely not linearly dependent on the available information. For instance, when we play an UG against someone who is ten times richer, we do not simply associate a ten times lower priority with this person than with ourselves. There are many factors that determine the human concept of a fair share. Extremely extensive and well-designed experiments with humans will be necessary in order to establish the effect of one single factor, which may then be captured by means of a priority value.

We emphasize that the conceptual idea behind priority awareness is still valid, as shown in experiments described in the literature (see, e.g., Bearden, 2001; Zollman, 2008), as well as our own experiments. Also, the concept has not yet been sufficiently addressed by descriptive models of fairness. Moreover, priority awareness may be applied in a multi-agent system context successfully, for instance, when we wish to incorporate the important concept of bargaining power, which is not modeled by the other computational models of human-inspired fairness. However, the model may need to be refined for it to be able to capture human behavior adequately.

More analytical results for social networks. In our work on social networks, we followed the general approach of existing work (most notably, Santos and Pacheco, 2005; Santos et al., 2006c), extending this existing work to continuous strategy spaces. This existing work empirically investigates how social networks, and rewiring in these networks, influences results. More precisely, the authors constructed an actual multi-agent system, in which agents learned by means of evolutionary algorithms, and examined how agents' strategies evolve. We also performed an empirical investigation, i.e., using CALA.

In addition to an empirical investigation, we may also follow an analytical approach. Given the stochastic nature of interactions between agents (who plays with who, who punishes whom, who unwires from who, who rewires to whom), we might, for instance, be able to calculate the expected reward for agents that decide to rewire, and compare this to the expected reward for agents that do not. Then, if rewiring yields a higher expected reward, we find analytical support for the claim that rewiring indeed helps.

7.2.2 The gap between current work and practice

A comment we regularly received from the anonymous referees of our submitted publications was that, although we did mention a number of applications for our models, we never actually showed our models being used in an application. Below, we will briefly discuss two interesting applications, one for each of the two social dilemmas under study. There are also possibilities to validate our models in practice, using humans.

The agreement dilemma. As has been indicated various times in the preceding text, the agreement dilemma is prominently present in applications which require interactions with humans. A possible application here is scheduling, e.g., of aircraft departures on a snowy day (Mao et al., 2006). Bad weather will lead to delays, since a lower number of aircraft may depart per hour. By representing airline companies as agents, each of which may have a certain (dynamic) priority, we may allocate the available time slots such that no airline company suffers unfairly more delay than others, while still allowing differences in delays. The possibility for agents to punish each other in case of unfair actions, is also interesting in the domain of auctions (and, more generally speaking, bargaining), where researchers aim to find mechanisms forcing agents to be truthful in their valuation of the goods in question (Preist and van Tol, 1998; Kalagnanam and Parkes, 2004; Chevaleyre et al., 2006; Sandholm, 2006). If, somehow, agents that try to deceive others may suffer punishment, there is a substantial incentive to be indeed truthful.

The tragedy of the commons. The tragedy of the commons is not only present in applications which require interactions with humans, but also, more generally, in any applications where a limited resource needs to be shared by multiple agents. In the preceding text of the thesis, we already mentioned load balancing as a typical application (Verbeeck et al., 2007). In a situation where agents may use either a slow personal client, or a fast common server, everyone would gain by moving their tasks to the server. However, if everyone indeed does

move their tasks to the server, this server probably takes more time to execute each task than the agents' personal client computer would. Essentially, this problem may be directly mapped to a PGG with a (near-)continuum of strategies. We may thus introduce the possibility for agents to punish each other for over-using the server. Moreover, given that most client-server solutions are inherently networked, we may introduce dynamic network structure, i.e., agents that over-use a certain server are (temporarily or permanently) banned from this server, allowing other agents to use its computational resources.

Practical validation. As a validation of our computational models, we might *re-introduce humans*. For instance, we may replace the fixed-strategy agents of Chapter 6, which model human players, by actual human players, and investigate how strategies emerge over time. Human players will punish and rewire if they are insufficiently satisfied. Thus, models may be considered successful if they drive all agents, including humans, to strategies that are decreasingly less frequently leading to punishment and rewiring.

References

P. Arnopoulos. *Sociophysics*. Nova Science Publishers, 1993. Cited on page 96.

R. Aumann. Acceptable points in general cooperative n-person games. In R. D. Luce and A. W. Tucker, editors, *Contributions to the Theory of Games IV,*, volume 40 of *Annals of Mathematics Study*, pages 287–324. Princeton University Press, Princeton NJ, 1959. Cited on page 64.

R. Axelrod. *The Evolution of Cooperation*. New York: Basic Books, 1984. Cited on pages 15, 16, and 64.

R. Axelrod. The Dissemination of Culture: A Model with Local Convergence and Global Polarization. *Journal of Conflict Resolution*, 41:203–226, 1997. Cited on page 96.

A.-L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286: 509–512, 1999. Cited on pages 96, 97, and 101.

N. Barr. *Economics of the welfare state*. New York, Oxford University Press (USA), 2004. Cited on page 36.

A. Barrat, A. Baronchelli, L. Dall'Asta, and V. Loreto. Agreement dynamics on interaction networks with diverse topologies. In *Workshop on Optimization in Complex Networks*, 2007. Cited on page 96.

K. Basu. The Traveler's Dilemma. *Scientific American*, Volume 296, Number 6:68–73, 2007. Cited on page 18.

J. N. Bearden. Ultimatum Bargaining Experiments: The State of the Art. *SSRN eLibrary*, 2001. Cited on pages 25, 48, 61, 73, 102, 128, and 129.

K. Binmore. *Game Theory and The Social Contract Volume 2: Just Playing*. MIT Press, 1998. Cited on page 19.

K. G. Binmore. *Fun and Games: A Text on Game Theory*. D.C. Heath, 1991. Cited on pages 2, 14, 19, and 26.

E. Borel. La theorie du jeu et les equations integrales a noyeau symmetrique. *Comptes Rendus de l'Academie des Sciences*, 173:1304–1308, 1921. [Translation: L. J. Savage. The theory of play and integral equations with skew symmetric kernels. *Econometrica*, 21:97-100, 1953.]. Cited on page 14.

T. Bourke. *Server Load Balancing*. O'Reilly Media, Inc., 2001. Cited on page 7.

S. Bowles, R. Boyd, E. Fehr, and H. Gintis. Homo reciprocans: A Research Initiative on the Origins, Dimensions, and Policy Implications of Reciprocal Fairness. *Advances in Complex Systems*, 4:1–30, 1997. Cited on pages 2, 68, 74, and 117.

M. H. Bowling and M. M. Veloso. Multiagent learning using a variable learning rate. *Artificial Intelligence*, 136:215–250, 2002. Cited on page 57.

R. Boyd and S. Mathew. A Narrow Road to Cooperation. *Science*, 316:1858–1859, 2007. Cited on pages 42, 70, 119, and 120.

R. Boyd, H. Gintis, S. Bowles, and P. J. Richerson. The evolution of altruistic punishment. *Proc Natl Acad Sci USA*, 100:3531–3535, 2003. Cited on pages 3 and 53.

S. J. Brams and A. D. Taylor. An Envy-free Cake Division Protocol. *American Mathematical Monthly*, 102(1):9–18, 1995. Cited on page 39.

S. F. Brosnan and F. B. M. de Waal. Monkeys reject unequal pay. *Nature*, 425(6955):297–299, Sept. 2003. ISSN 0028-0836. Cited on page 46.

L. Cameron. Raising the stakes in the ultimatum game: Evidence from Indonesia. *Journal of Economic Inquiry*, 37:47–59, 1999. Cited on pages 73 and 102.

C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *to appear in Reviews of Modern Physics*, to appear, 2008. Eprint arXiv: 0710.3256. Cited on page 96.

Y. Chevaleyre, P. Dunne, U. Endriss, J. Lang, M. Lemaître, N. Maudet, J. Padget, S. Phelps, J. Rodriguez-Aguilar, and P. Sousa. Issues in Multiagent Resource Allocation. *Informatica*, 30:3–31, 2006. Cited on pages 1, 4, 36, 38, 39, 128, and 130.

Y.-S. Chiang. A Path Toward Fairness: Preferential Association and the Evolution of Strategies in the Ultimatum Game. *Rationality and Society*, 20(2):173–201, 2008. Cited on page 118.

W. Cohen, R. Schapire, and Y. Singer. Learning to order things. *Intelligence Research*, 10: 243–270, 1999. Cited on page 78.

L. Dall'Asta, A. Baronchelli, A. Barrat, and V. Loreto. Agreement dynamics on small-world networks. *Europhysics Letters*, 73(6):pp. 969–975, 2006a. Cited on page 42.

L. Dall'Asta, A. Baronchelli, A. Barrat, and V. Loreto. Non-equilibrium dynamics of language games on complex networks. *Phys. Rev. E*, 74:36–105, 2006b. Cited on page 96.

A. Dannenberg, T. Riechmann, B. Sturm, and C. Vogt. Inequity Aversion and Individual Behavior in Public Good Games: An Experimental Investigation. *SSRN eLibrary*, 2007. Cited on pages 2, 48, 52, 53, and 85.

R. Dawkins. *The Selfish Gene.* Oxford University Press, 1976. Cited on page 71.

R. Dawkins. *The God Delusion.* Boston: Houghton Mifflin, 2006. Cited on page 72.

S. de Jong, K. Tuyls, T. Hashimoto, and H. Iida. Scalable Potential-Field Multi-Agent Coordination in Resource Distribution Tasks. In *Proceedings of the ALAAMAS 2006 Workshop*, 2006a. Cited on page 5.

S. de Jong, K. Tuyls, and I. Sprinkhuizen-Kuyper. Nature-Inspired Multi-Agent Coordination in Task Assignment Problems. In *Proceedings of ALAMAS*, 2006b. Cited on page 5.

S. de Jong, K. Tuyls, and I. Sprinkhuizen-Kuyper. Robust and Scalable Coordination of Potential-Field Driven Agents. In *Proceedings of IAWTIC/CIMCA 2006, Sydney*, 2006c. Cited on pages 4, 5, and 128.

S. de Jong, K. Tuyls, K. Verbeeck, and N. Roos. Priority Awareness: Towards a Computational Model of Human Fairness for Multi-agent Systems. *Adaptive Agents and Multi-Agent Systems III. Adaptation and Multi-Agent Learning*, 4865:117–128, 2008. Cited on pages 2, 48, 61, 76, 78, 80, 82, and 103.

C. N. Dellarocas. The Digitization of Word-of-Mouth: Promise and Challenges of Online Feedback Mechanisms. *SSRN eLibrary*, 2003. Cited on page 69.

D. Denning. *Cryptography and Data Security*. Addison-Wesley, Boston, USA, 1982. Cited on page 83.

P. G. Devine. Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56:5–18, 1989. Cited on page 76.

U. Endriss. Fair Division. Tutorial at the International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), 2008. Cited on pages 1, 4, 7, 18, 29, and 38.

U. Endriss, N. Maudet, F. Sadri, and F. Toni. Negotiating Socially Optimal Allocations of Resources. *Journal of Artificial Intelligence Research*, 25:315–348, 2006. Cited on page 40.

I. Erev and A. E. Roth. Predicting how people play games with unique, mixed strategy equilibria. *American Economic Review*, 88:848–881, 1998. Cited on page 4.

A. Falk and U. Fischbacher. A theory of reciprocity. *Games and Economic Behavior*, 54: 293–315, 2006. Cited on pages 68, 69, and 71.

E. Fehr. Don't lose your reputation. *Nature*, 432:499–500, 2004. Cited on pages 68, 69, 74, and 117.

E. Fehr and S. Gaechter. Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives*, 14:159–181, 2000. Cited on pages 51 and 68.

E. Fehr and S. Gaechter. Altruistic punishment in humans. *Nature*, 415:137–140, 2002. Cited on pages 2, 3, 27, 51, 53, 68, and 128.

E. Fehr and K. Schmidt. A Theory of Fairness, Competition and Cooperation. *Quart. J. of Economics*, 114:817–868, 1999. Cited on pages 1, 2, 3, 5, 32, 42, 46, 47, 48, 53, 59, 61, 72, 74, 82, 106, 128, 150, and 154.

J. Ferber. *Multi-Agent Systems. An Introduction to Distributed Artificial Intelligence*. Addison-Wesley, Boston, USA, 1999. Cited on page 3.

U. Fischbacher, C. M. Fong, and E. Fehr. Fairness and the Power of Competition. Working Paper 133, Institute for Empirical Research in Economics, University of Zurich, January 2003. Cited on page 25.

P. C. Fishburn. *Utility Theory for Decision Making*. Robert E. Krieger Publishing Co., 1970. Cited on page 32.

S. Fortunato. Damage spreading and opinion dynamics on scale-free networks. *Physica A*, 348:683, 2005. Cited on page 96.

S. Fortunato and C. Castellano. Scaling and universality in proportional elections. *Physical Review Letters*, 99:138701, 2007. Cited on page 96.

S. Galam. Sociophysics: a personal testimony. *Physica A: Statistical and Theoretical Physics*, 336(1-2):49–55, May 2004. ISSN 03784371. Cited on page 96.

E. Gerding, D. van Bragt, and J. L. Poutré. Multi-Issue Negotiation Processes by Evolutionary Simulation: Validation and Social Extensions. *Computational Economics*, 22:39–63, 2003. Cited on page 47.

H. Gintis. *Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Interaction*. Princeton University Press, Princeton, USA, 2001. Cited on pages 1, 5, 14, 23, 46, 69, and 74.

B. Grosskopf. Reinforcement and Directional Learning in the Ultimatum Game with Responder Competition. *Experimental Economics*, 6(2):141–158, October 2003. Cited on page 56.

W. Gueth, R. Schmittberger, and B. Schwarze. An Experimental Analysis of Ultimatum Bargaining. *Journal of Economic Behavior and Organization*, 3 (4):367–388, 1982. Cited on pages 2 and 25.

M.-L. Halko and T. Seppala. Ultimatum Game Experiments. Discussion paper 140, Helsinki Center of Economic Research, University of Helsinki, Finland, 2006. Cited on page 25.

W. D. Hamilton. The genetical evolution of social behaviour I and II. *Journal of Theoretical Biology*, 7:1–16 and 17–52, 1964. Cited on page 71.

G. Hardin. The Tragedy of the Commons. *Science*, 162:1243–1248, 1968. Cited on pages 2 and 26.

J. C. Harsanyi. Approaches to the Bargaining Problem Before and After the Theory of Games: A Critical Discussion of Zeuthen's, Hicks', and Nash's Theories. *Econometrica*, 24: 144–157, 1956. Cited on page 18.

C. Hauert, S. D. Monte, J. Hofbauer, and K. Sigmund. Volunteering as red queen mechanism for cooperation in public goods games. *Science*, 296:1129–1132, 2002. Cited on pages 3, 27, 42, 51, 68, 70, and 98.

C. Hauert, A. Traulsen, H. Brandt, M. A. Nowak, and K. Sigmund. Via freedom to coercion: the emergence of costly punishment. *Science*, 316:1905–1907, 2007. Cited on pages 42, 51, 68, 70, 98, and 119.

R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence: models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002. Cited on page 96.

D. Hennes. Multi-agent Learning in Stochastic Games - Piecewise and State-coupled Replicator Dynamics. Master's thesis, Universiteit Maastricht, 2008. Cited on page 129.

J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, and H. Gintis. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford University Press, Oxford, UK, 2004. Cited on pages 2, 25, and 43.

J. Henrich, R. McElreath, A. Barr, J. Ensimger, C. Barrett, A. Bolyanatz, J. C. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe, D. Tracer, and J. Ziker. Costly Punishment Across Human Societies. *Science*, 312:1767–1770, 2006. Cited on page 51.

T. Huynh, N. R. Jennings, and N. Shadbolt. Certified Reputation - How an Agent Can Trust a Stranger. In *The Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1217–1224, May 2006a. Cited on page 69.

T. Huynh, N. R. Jennings, and N. Shadbolt. An integrated trust and reputation model for open multi-agent systems. *Journal of Autonomous Agents and Multi-Agent Systems*, 13(2): 119–154, 2006b. Cited on page 41.

E. Ising. Beitrag zur Theorie des Ferromagnetismus. *Z. Phys.*, 31:253–258, 1925. Cited on page 96.

J. H. M. Janssens. Collaborative image ranking. Master's thesis, Faculty of Humanities and Sciences of the Universiteit Maastricht, 2008. Cited on page 76.

N. R. Jennings, K. Sycara, and M. Wooldridge. A Roadmap of Agent Research and Development. *Autonomous agents and Multi-Agent Systems*, 1:275 – 306, 1998. Cited on pages 3, 149, and 153.

J. Kalagnanam and D. C. Parkes. Auctions, Bidding and Exchange Design. In D. Simchi-Levi, S. D. Wu, and M. Shen, editors, *Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era*, Int. Series in Operations Research and Management Science, chapter 5, pages 1–84. Kluwer Academic Publishers, Dordrecht, the Netherlands, 2004. Cited on pages 4 and 130.

G. Kalisch, J. W. Milnor, J. Nash, and E. D. Nering. Some experimental n-person games. Technical report, The Rand Corporation, U.S. Air Force, 1952. Cited on page 19.

L. Keller and K. G. Ross. Selfish genes: a green beard in the red fire ant. *Nature*, 394(6693): 573–575, 1998. ISSN 0028-0836. Cited on page 72.

M. Kendall. A new measure of rank correlation. *Biometrika*, 30:81–89, 1938. Cited on page 80.

D. Knoch, A. Pascual-Leone, K. Meyer, V. Treyer, and E. Fehr. Diminishing Reciprocal Fairness by Disrupting the Right Prefrontal Cortex. *Science*, 314(5800):829–832, 2006. ISSN 1095-9203. Cited on page 71.

O. Leimar and P. Hammerstein. Evolution of Cooperation through Indirect Reciprocity. *Proceedings: Biological Sciences*, Vol. 268, No. 1468:745–753, 2001. Cited on pages 69 and 70.

N. Lemmens, S. de Jong, K. Tuyls, and A. Nowé. Bee System with inhibition Pheromones. In *European Conference on Complex Systems (ECCS)*, 2007. Cited on page 5.

N. Lemmens, S. de Jong, K. Tuyls, and A. Nowé. Bee behaviour in multi-agent systems: A bee foraging algorithm. *Adaptive Agents and Multi-Agent Systems III - Lecture Notes in Artificial Intelligence*, 4865, 2008. Cited on page 5.

A. Lotem, M. A. Fishman, and L. Stone. Evolution of Cooperation between individuals. *Nature*, 400:226–227, 1999. Cited on page 69.

D. MacKay. *Information theory, inference and learning algorithms*. Cambridge University Press, 2003. Cited on page 77.

X. Mao, A. ter Mors, N. Roos, and C. Witteveen. Agent-Based Scheduling for Aircraft Deicing. In P.-Y. Schobbens, W. Vanhoof, and G. Schwanen, editors, *Proceedings of the 18th Belgium - Netherlands Conference on Artificial Intelligence*, pages 229–236. BNVKI, October 2006. ISBN 1568-7805. Cited on pages 4, 128, and 130.

J. Maynard-Smith. *Evolution and the Theory of Games*. Cambridge University Press, 1982. Cited on page 19.

J. Maynard-Smith and G. R. Price. The logic of animal conflict. *Nature*, 246:15–18, 1973. Cited on page 19.

D. M. Messick and M. B. Brewer. Solving social dilemmas: A review. *Review of personality and social psychology*, 4:11–44, 1983. Cited on page 23.

M. Milinski, D. Semmann, and H. J. Krambeck. Reputation helps solve the tragedy of the commons. *Nature*, 415:424–426, 2002. Cited on pages 24, 42, 64, 68, and 74.

H. Moulin. *Axioms of cooperative decision making*. Econometric society monographs, Cambridge University Press, 1988. Cited on pages 36 and 37.

K. Narendra and M. Thathachar. *Learning Automata: An introduction*. Prentice-Hall International, Boston, USA, 1989. Cited on pages 9, 21, 22, and 65.

J. Nash. Equilibrium Points in N-person Games. *Proceedings of the National Academy of Sciences*, 36:48–49, 1950a. Cited on pages 14 and 17.

J. Nash. The Bargaining Problem. *Econometrica*, 18:155–162, 1950b. Cited on pages 18 and 25.

J. Nash. *The Essential John Nash*. Princeton University Press, 2001. Cited on page 25.

M. A. Nowak and K. Sigmund. Evolution of indirect reciprocity by image scoring. *Nature*, 393(6685):573–577, 1998. ISSN 0028-0836. Cited on page 69.

M. A. Nowak and K. Sigmund. Evolution of indirect reciprocity. *Nature*, 437(7063): 1291–1298, 2005. ISSN 0028-0836. Cited on page 69.

M. A. Nowak, K. M. Page, and K. Sigmund. Fairness versus reason in the Ultimatum Game. *Science*, 289:1773–1775, 2000. Cited on pages 24, 42, 69, 117, 118, 128, and 129.

M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg. Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428:646–650, 2004. Cited on page 70.

R. Nydegger and H. Owen. Two-person bargaining, an experimental test of the Nash axioms. *International Journal of Game Theory*, 3:239–250, 1974. Cited on pages 26 and 50.

H. Oosterbeek, R. Sloof, and G. van de Kuilen. Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis. *Experimental Economics*, 7:171–188, 2004. Cited on pages 2, 25, 33, 43, 48, 61, 72, 73, 76, 78, 80, 102, 106, 122, and 128.

J. O. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, Boston, USA, Cambridge, MA, 1994. Cited on page 14.

K. Panchanathan and R. Boyd. Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, 432:499–502, 2004. Cited on pages 3 and 69.

V. Prasnikar and A. E. Roth. Considerations of Fairness and Strategy: Experimental Data from Sequential Games. *The Quarterly Journal of Economics*, 107(3):865–88, 1992. Cited on page 25.

C. Preist and M. van Tol. Adaptive agents in a persistent shout double auction. In *ICE '98: Proceedings of the first international conference on Information and computation economies*, pages 11–18. ACM Press, 1998. ISBN 1-58113-076-7. Cited on pages 4 and 130.

M. Rabin. Incorporating Fairness into Game Theory and Economics. *American Economic Review*, 83:1281–1302, 1993. Cited on page 46.

J. Rawls. *A Theory of Justice*. Oxford University Press, 1971. Cited on page 38.

S. Reece, A. Rogers, S. Roberts, and N. Jennings. Rumours and reputation: evaluating multi-dimensional trust within a decentralized reputation system. In *Proc. 6th Int. J. Conference on Autonomous Agents and Multi-agent Systems (AAMAS)*, pages 1063–1070, 2007. Cited on page 69.

L. E. Reichl. *A Modern Course in Statistical Physics*. Wiley-Interscience, 2nd edition, 1980. Cited on page 96.

B. Rockenbach and M. Milinski. The efficient interaction of indirect reciprocity and costly punishment. *Nature*, 444(7120):718–723, Dec. 2006. ISSN 0028-0836. Cited on pages 23 and 42.

J. A. Rodriguez-Aguilar. *On the design and construction of agent-mediated electronic institutions*. PhD thesis, Monografies de l'Institut d'Investigació en Intelligència Artificial, 2003. Cited on page 4.

A. Roth and M. Malouf. Game Theoretic Models and the Role of Information in Bargaining. *Psychological Review*, 86:574–594, 1979. Cited on pages 26 and 50.

A. E. Roth, V. Prasnikar, M. Okuno-Fujiwara, and S. Zamir. Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study. *American Economic Review*, 81(5):1068–95, December 1991. Cited on page 73.

T. Sandholm. *Combinatorial Auctions*, chapter Optimal Winner Determination Algorithms. MIT Press, 2006. Cited on pages 39 and 130.

A. G. Sanfey, J. K. Rilling, J. A. Aronson, L. E. Nystrom, and J. D. Cohen. The Neural Basis of Economic Decision-Making in the Ultimatum Game. *Science*, 300:1755–1758, 2003. Cited on page 71.

F. C. Santos and J. M. Pacheco. Scale-Free networks provide a unifying framework for the Emergence of Cooperation. *Physical Review Letters*, 95:98–104, 2005. Cited on pages 42, 64, and 130.

F. C. Santos, J. M. Pacheco, and T. Lenaerts. Emergence of Cooperation in Heterogeneous Structured Populations. *Proceedings of the 10th International Conference on the Simulation and Synthesis of Living Systems (Alife X)*, 1:432–437, 2006a. Cited on pages 42 and 97.

F. C. Santos, J. M. Pacheco, and T. Lenaerts. Evolutionary Dynamics of Social Dilemmas in Structured Heterogeneous Populations. *Proc. Natl. Acad. Sci. USA*, 103:3490–3494, 2006b. Cited on pages 97, 101, and 122.

F. C. Santos, J. M. Pacheco, and T. Lenaerts. Cooperation Prevails When Individuals Adjust Their Social Ties. *PLoS Comput. Biol.*, 2(10):1284–1291, 2006c. Cited on pages 42, 97, 104, 107, 118, 120, 128, and 130.

U. Schwalbe and P. Walker. Zermelo and the Early History of Game Theory. *Games and Economic Behavior*, 34:123–137, 2001. Cited on page 14.

R. Selten and R. Stoecker. End behavior in sequences of finite Prisoner's Dilemma supergames : A learning theory approach. *Journal of Economic Behavior & Organization*, 7 (1):47–70, March 1986. Cited on page 56.

A. Sen. Markets and freedom: Achievements and limitations of the market mechanism in promoting individual freedoms. *Oxford Economic Papers*, 45(4):519–541, 1993. Cited on page 36.

A. K. Sen. *Collective choice and social welfare.* Holden Day, 1970. Cited on page 4.

Y. Shoham, R. Powers, and T. Grenager. If multi-agent learning is the answer, what is the question? *Artificial Intelligence*, 171(7):365–377, 2007. Cited on pages 3, 30, 149, and 153.

K. Sigmund. Keynote talk. European Conference on Complex Systems, 2007. Cited on page 26.

K. Sigmund, C. Hauert, and M. A. Nowak. Reward and punishment. *Proceedings of the National Academy of Sciences*, 98(19):10757–10762, 2001. Cited on pages 2, 53, 64, 68, 69, 119, and 129.

H. Simon. *Models of Man.* John Wiley, London, UK, 1957. Cited on page 19.

H. Simon and A. Newell. *Human Problem Solving.* Prentice-Hall, Englewood Cliffs, USA, 1972. Cited on page 19.

R. Slonim and A. Roth. Learning in high stakes ulitmatum games: An experiment in the Slovak republic. *Econometrica*, 66:569–596, 1998. Cited on pages 73 and 102.

J. Sonnegard. Determination of first movers in sequential bargaining games: An experimental study. *Journal of Economic Psychology*, 17:359–386, 1996. Cited on page 73.

H. Steinhaus. The Problem of Fair Division. *Econometrica*, 16:101–104, 1948. Cited on page 39.

P. G. Straub and J. K. Murninghan. An experimental investigation of ultimatum games: information, fairness expectations and lowest acceptable offers. *Journal of Economic Behavior and Organization*, 27:345–364, 1995. Cited on page 73.

W. Stromquist. How to Cut a Cake Fairly. *American Mathematical Montly*, 87(8):640–644, 1980. Cited on page 39.

R. Sugden. *The economics of rights, co-operation and welfare*. Basil Blackwell, Oxford, UK, 1986. Cited on page 70.

R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998. A Bradford Book. Cited on pages 9, 20, and 21.

P. 't Hoen, K. Tuyls, L. Panait, S. Luke, and J. L. Poutré. *Learning and Adaptation in Multiagent Systems (LAMAS)*, volume 3898 of *Springer Lecture Notes in Artificial Intelligence (LNAI)*, chapter An Overview of Cooperative and Competitive Multiagent Learning, pages 1–49. Springer Verlag, 2006. Cited on page 20.

M. A. L. Thathachar and P. S. Sastry. *Networks of Learning Automata: Techniques for Online Stochastic Optimization*. Kluwer Academic Publishers, Dordrecht, the Netherlands, 2004. Cited on pages 9, 22, 23, 56, and 100.

K. Tuyls and S. Parsons. What evolutionary game theory tells us about multiagent learning. *Artificial Intelligence*, 171(7):406–416, 2007. Cited on pages 20 and 129.

K. Tuyls and R. Westra. *Accepted in Multi-Agent Systems: Simulation and Applications*, chapter Replicator Dynamics in Discrete and Continuous Strategy Spaces. 2009. Cited on page 129.

A. Tversky. *Preference, Belief, and Similarity: Selected Writings*. MIT Press, 2004. Cited on page 17.

S. Uyttendaele. Fairness and agreement in complex networks. Master's thesis, MICC, Maastricht University, 2008. Cited on page 120.

J. van Huyck, R. Battalio, S. Mathur, and A. Ortmann. On the origin of convention: evidence from symmetric bargaining games. *International Journal of Game Theory*, 34:187–212, 1995. Cited on page 50.

K. L. Vaughn. *The New Palgrave : A dictionary of Economics*, volume 2, chapter Invisible Hand. Macmillan, London, UK, 1987. Cited on page 26.

K. Verbeeck, J. Parent, and A. Nowe. Homo Egualis Reinforcement Learning Agents for Load Balancing. In *First NASA Workshop on Radical Agents Concepts*, number LNAI2564 in Lecture Notes on Artificial Intelligence, pages 81–91. Springer Verlag, 2002. Cited on page 26.

K. Verbeeck, A. Nowé, J. Parent, and K. Tuyls. Exploring Selfish Reinforcement Learning in Repeated Games with Stochastic Rewards. *Journal of Autonomous Agents and Multi-Agent Systems*, 14:239–269, 2007. Cited on pages 5, 7, 47, 128, and 130.

L. von Ahn. Games with a Purpose. *IEEE Computer*, 39(6):92–94, 2006. Cited on pages 74 and 76.

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behaviour*. Princeton University Press, 1944. Cited on page 14.

P. Walker. A Chronology of Game Theory. http://www.econ.canterbury.ac.nz/personal_pages/paul_walker/gt/hist.htm, retrieved March 6, 2008, 2005. Cited on page 14.

E. Walster, W. G.W., and E. Bershcheid. *Equity: Theory and Research*. Allyn and Bacon, Inc., 1978. Cited on page 46.

C. Watkins. *Learning from delayed rewards*. PhD thesis, University of Cambridge, Psychology Department, 1989. Cited on page 21.

C. Wedekind and M. Milinski. Cooperation Through Image Scoring in Humans. *Science*, 288(5467):850–852, 2000. Cited on page 69.

J. W. Weibull. *Evolutionary Game Theory.* MIT Press, 1996. Cited on page 129.

E. R. Weintraub, editor. *Toward a History of Game Theory*. Duke University Press, Durham, NC, 1992. Cited on page 14.

G. Weiss. *Multiagent Systems. A Modern Approach to Distributed Artificial Intelligence*. MIT Press, Boston, USA, 1999. Cited on page 3.

D. Weyns, N. Boucke, T. Holvoet, and W. Schols. Gradient Field-Based Task Assignment in an AGV Transportation System. In *Proceedings of EUMAS*, pages 447–458, 2005. Cited on pages 4 and 5.

S. C. Wheeler and R. E. Petty. The effects of stereotype activation on behavior: a review of possible mechanisms. *Psychol Bull.*, 127(6):797–826, 2001. Cited on page 76.

M. Yaari and M. Bar-Hillel. On dividing justly. *Social Choice and Welfare*, 1:1–24, 1981. Cited on pages 26 and 50.

T. Yamagishi. The provision of a sanctioning system as a public good. *J. Pers. and Soc. Psych.*, 51(1):110–116, 1986. Cited on pages 27, 51, and 68.

E. Zermelo. Neuer Beweis für die Möglichkeit einer Wohlordnung. *Mathematische Annalen*, 65(1):107–128, Mar. 1907. Cited on page 14.

F. Zeuthen. *Problems of Monopoly and Economic Warfare.* London: George Routledge and Sons, 1930. Cited on pages 14 and 18.

K. J. S. Zollman. Explaining Fairness in Complex Environments. *Politics, Philosophy, and Economics*, 7(1):81–97, 2008. Cited on pages 73 and 129.

List of figures

List of tables

List of definitions

Summary

Within the field of artificial intelligence, the research area of multi-agent systems investigates societies of autonomous entities, called *agents*, that need to cooperate or compete in order to achieve a certain goal (Jennings et al., 1998; Shoham et al., 2007). Humans may be part of these societies. Example applications include resource distribution, auctions, and load balancing. In many of these applications, we observe elements of so-called *social dilemmas*, in which taking into account fairness and social welfare is necessary. In some dilemmas, humans are known to care strongly for fairness and social welfare; in others, caring for fairness and social welfare is essential for agents to achieve a satisfactory solution.

In this thesis, we show how agents may be stimulated to care for the fairness of their actions. The human-inspired mechanisms of altruistic punishment and withholding action are central to our approach. Chapter 1 therefore presents the following problem statement.

PS *How can we obtain human-inspired fairness in multi-agent systems?*

The remainder of Chapter 1 provides an overview of five research questions resulting from this problem statement. These questions are given below.

RQ1 *How are humans using fairness in their decisions?*

RQ2 *What are the foundations of human-inspired computational fairness?*

RQ3 *How can human-inspired fairness be modeled computationally, taking into account the foundations of human fairness?*

RQ4 *What are the analytical properties of the computational human-inspired fairness models developed in this research?*

RQ5 *What are the (empirical) benefits of incorporating explicitly human-inspired fairness in adaptive multi-agent systems?*

Thereafter, Chapter 1 discusses the research methodology followed. Next, in Chapter 2, we review the fundamental background knowledge required for research in multi-agent systems in general, i.e., game theory and multi-agent reinforcement learning. Chapter 2 also elaborately explains the social dilemmas we investigate throughout the thesis, i.e., the *Ultimatum Game*, the *Nash Bargaining Game*, and the *Public Goods Game*.

The five research questions are addressed in Chapters 3 to 6. RQ2 is addressed before the other four research questions, i.e., in Chapter 3. Essentially, we there present two foundations, i.e., (1) a set of three *requirements* that need to be met by human-inspired computational fairness models, and (2) a *template model* based on these requirements. We require that any computational model should be (R1) rooted in a game-theoretic background (as game theory is a well-established manner of describing interactions between multiple

agents), (R2) computationally applicable in an adaptive multi-agent system (i.e., we require tractable solution concepts), as well as (R3) inspired by humans. With respect to the last requirement, we state that agents must be able to answer three questions, i.e., (R3-Q1) to what extent an interaction is fair, (R3-Q2) whether one or more of their peers need(s) to be punished, and (R3-Q3) whether it is desirable to withhold action, i.e., not to participate in a certain interaction. We present a template model, based on the well-known concept of a utility function, and show how this model may be instantiated in such a way that our requirements are met.

The following three chapters, i.e., Chapters 4 to 6, form the core of the thesis, and follow a similar structure. In each chapter, we discuss a specific computational model of human-inspired fairness, based on the foundations presented in Chapter 3. For each model, we address RQ1 by discussing a specific descriptive model of human fairness. We then create a computational model of fairness, incorporating this specific descriptive model (RQ3), analyze the computational model (RQ4), and use the model in an adaptive multi-agent system that is learning to find good solutions to social dilemmas (RQ5).

Chapter 4 presents a computational model based on a descriptive model of *inequity aversion*, as developed by Fehr and Schmidt (1999). Inequity aversion entails that human decisions are influenced by differences in observed rewards. The descriptive model of Fehr and Schmidt (1999) is able to explain a great deal of (irrational) human decisions in interactions where limited resources need to be shared. Even though this is the case, the model has not yet convincingly found its way into multi-agent systems. We address this issue by developing a computational model of fairness, based on inequity aversion. We show that our computational model allows (a small number of) agents to reach satisfactory, human-inspired solutions to the social dilemmas under study.

In Chapter 5, we discuss that human behavior is not only influenced by (differences in) observed rewards, but also by additional information humans may have or gather about the others participating in an interaction. Existing research proposes reputation-based approaches to model this phenomenon. We argue that an important element is missing from reputation-based models, i.e., that there may be additional information that is immediately present, e.g., bargaining powers, stereotypes, or priorities. We present a descriptive model named *priority awareness* to address both additional information that is immediately present, as well as reputation. In an approach similar to that of Chapter 4, we show how a computational model of fairness may be based on priority awareness, and how this influences outcomes to interactions between agents.

In Chapter 6, we increase the scale of our work from at most a few dozen to a few thousand agents. In the last ten years, a great deal of research has been devoted to *social networks*, which have been shown to have a decisive impact on the manner in which humans (as well as artificial agents) change their behavior as a result of interactions with others, based on neighbor relations in a certain network structure. Existing work examining how networked agents change their behavior in social-dilemma-like interactions has thus far been limited to social dilemmas with only a discrete, small number of possible actions (e.g., two). Since

we are interested in addressing more realistic social dilemmas, we investigate how network structure influences agents' behavior in social dilemmas with a continuum of actions. We show that a number of mechanisms promoting desirable behavior in discrete dilemmas also work in continuous dilemmas (i.e., most prominently the possibility of agents to withhold action by breaking the link between them and an undesirable neighbor), while a number of other mechanisms do not provide additional benefits (e.g., reputation).

We conclude in Chapter 7 by answering our research questions, summarizing our findings, answering the problem statement, and looking at opportunities for future work. We show that human-inspired fairness in multi-agent systems may be obtained by means of a four-step process, i.e., (1) experiments with human subjects, (2) modeling human fairness in descriptive models, (3) establishing the foundations of computational models of human-inspired fairness, and (4) translating descriptive models to computational models respecting the foundations. Using the three computational models presented in this thesis, we may conclude that agents are able to find desirable solutions to social dilemmas, even in the presence of agents that do not care about fairness and try to undermine a desirable, fair solution.

Samenvatting

Binnen de kunstmatige intelligentie bevindt zich een onderzoeksgebied genaamd 'multi-agent systemen'. In dit gebied wordt onderzoek gedaan naar collectieven van autonome entiteiten, die men *agenten* noemt. Deze agenten moeten samenwerken of met elkaar wedijveren om een bepaald doel te bereiken (Jennings et al., 1998; Shoham et al., 2007). Het is mogelijk dat mensen deelnemen aan zulke collectieven. Voorbeeld-toepassingen zijn onder meer gelegen in het verdelen van hulpbronnen en beloningen. In veel toepassingen vinden we elementen terug van de zogenoemde *sociale dilemma's*. In dit soort dilemma's moet overwogen worden om eerlijkheid en sociaal welbevinden in acht te nemen. In sommige gevallen is dit wenselijk omdat mensen hier zeer veel belang aan hechten; in andere gevallen is het strikt noodzakelijk, omdat er anders geen bevredigende oplossing wordt gevonden.

In dit proefschrift tonen we hoe agenten gestimuleerd kunnen worden om te geven om de eerlijkheid van hun acties. De mechanismen van altruïstisch bestraffen en afzien van actie, die geïnspireerd zijn door mensen, nemen een centrale plaats in. Hoofdstuk 1 presenteert een probleemstelling als volgt.

PS *Hoe kan op mensen geïnspireerde eerlijkheid verkregen worden in multi-agent systemen?*

De rest van Hoofdstuk 1 geeft een overzicht van vijf onderzoeksvragen die voortvloeien uit deze probleemstelling. Deze vragen worden hieronder weergegeven.[1]

RQ1 *Hoe gebruiken mensen eerlijkheid in hun beslissingen?*

RQ2 *Wat zijn de fundamenten van door mensen geïnspireerde, computationele eerlijkheid?*

RQ3 *Hoe kan door mensen geïnspireerde eerlijkheid computationeel gemodelleerd worden, met inachtname van de fundamenten van menselijke eerlijkheid?*

RQ4 *Wat zijn de theoretische eigenschappen van modellen van door mensen geïnspireerde, computationele eerlijkheid?*

RQ5 *Wat zijn de (empirische) voordelen van het toevoegen van expliciet door mensen geïnspireerde eerlijkheid aan multi-agent systemen?*

Daarna vervolgt Hoofdstuk 1 met een bespreking van onze onderzoeksmethodologie. Vervolgens bespreken we in Hoofdstuk 2 de fundamentele achtergrondkennis die vereist is voor vrijwel elk onderzoek gerelateerd aan multi-agent systemen, te weten speltheorie en 'multi-agent reinforcement learning'. Hoofdstuk 2 legt bovendien uitgebreid de sociale dilemma's uit die we in dit proefschrift steeds gebruiken, namelijk de zogenoemde *Ultimatum Game*, de *Nash Bargaining Game*, en de *Public Goods Game*.

De vijf onderzoeksvragen worden beantwoord in de Hoofdstukken 3 tot en met 6. RQ2 wordt beantwoord in Hoofdstuk 3, voor de andere vier onderzoeksvragen aan bod komen.

[1] We merken op dat de afkorting 'RQ' is afgeleid van de Engelstalige term 'research question'.

In feite bekijken we twee fundamenten, namelijk (1) drie *vereisten* waaraan moet worden voldaan door modellen van computationele, door mensen geïnspireerde eerlijkheid, en (2) een *sjabloonmodel* dat wordt gebaseerd op die drie vereisten. Om precies te zijn vereisen we dat ieder computationeel model (R1) geworteld moet zijn in een speltheoretische achtergrond (omdat speltheorie een goed gefundeerde manier is om interacties tussen meerdere agenten te beschrijven), (R2) computationeel toepasbaar moet zijn in een adaptief multi-agent systeem (dat wil zeggen, de modellen zijn berekenbaar), en tenslotte (R3) geïnspireerd moet zijn door mensen. De laatste vereiste werken we nog verder uit door te stellen dat agenten in staat moeten zijn om drie vragen te beantwoorden, te weten (R3-Q1) in hoeverre een interactie eerlijk is, (R3-Q2) of één of meer van de anderen bestraft dient te worden, en (R3-Q3) of het wenselijk is om af te zien van actie. We presenteren een sjabloonmodel, gebaseerd op het bekende concept van een nutsfunctie, en tonen hoe dit model geïnstantieerd kan worden op een manier die onze vereisten respecteert.

De volgende drie hoofdstukken (Hoofdstuk 4 tot en met 6) vormen de kern van dit proefschrift, en hebben allemaal een gelijkaardige structuur. In ieder van deze hoofdstukken behandelen we een specifiek computationeel model van door mensen geïnspireerde eerlijkheid, gebaseerd op de fundamenten uit Hoofdstuk 3. Voor elk model beantwoorden we RQ1 door aandacht te besteden aan een specifiek descriptief model van menselijke eerlijkheid. We maken vervolgends een computationeel model van door mensen geïnspireerde eerlijkheid dat is gebaseerd op het specifieke descriptieve model (RQ3). Dan analyseren we het computationele model (RQ4) en gebruiken we het in een adaptief multi-agent systeem dat leert om goede oplossingen te vinden voor sociale dilemma's (RQ5).

Hoofdstuk 4 behandelt een computationeel model dat is gebaseerd op een descriptief model van aversie tegen onrecht, zoals ontwikkeld door Fehr and Schmidt (1999). Het model beschrijft dat mensen beïnvloed worden door (in hun ogen) onrechtvaardige verschillen in beloning. Het descriptieve model van Fehr and Schmidt (1999) is in staat om een grote verscheidenheid aan (irrationeel) menselijk gedrag te verklaren in interacties waarbij beperkte hulpbronnen dienen te worden gedeeld. Ondanks dit feit is het model tot nog toe niet zeer overtuigend van nut geweest in multi-agent systemen. We laten zien dat het model wel degelijk nut heeft, door een computationeel model van eerlijkheid te ontwikkelen dat is gebaseerd op aversie tegen onrecht, en dit model succesvol toe te passen. Een (relatief kleine) groep agenten kan, gebruikmakend van het computationeel model, bevredigende en op mensen geïnspireerde oplossingen vinden voor de sociale dilemma's die we bestuderen.

In Hoofdstuk 5 bespreken we dat menselijk gedrag niet alleen wordt beïnvloed door (verschillen tussen) waargenomen beloning, maar ook door additionele informatie die mensen kunnen hebben of verkrijgen over de anderen die aan een interactie deelnemen. Bestaand onderzoek stelt voor om dit fenomeen te modelleren door middel van reputatie. We observeren dat er een belangrijk element mist in de bestaande modellen van reputatie, namelijk dat de relevante additionele informatie geregeld onmiddellijk beschikbaar is; we denken daarbij aan bijvoorbeeld onderhandelingspositie, stereotypes, of prioriteiten. We introduceren een descriptief model gebaseerd op prioriteiten, dat zowel onmiddellijk beschikbare

als ook door reputatie verkregen additionele informatie kan modelleren. Door middel van een aanpak die verder sterk lijkt op die van Hoofdstuk 4 laten we zien hoe we een computationeel model kunnen baseren op het idee van prioriteiten, en hoe deze prioriteiten invloed hebben op de uitkomst van interacties tussen agenten.

In Hoofdstuk 6 stappen we over van systemen met maximaal enkele tientallen agenten naar systemen met enkele duizenden agenten. De laatste tien jaar is een opmerkelijk grote hoeveelheid onderzoek gericht geweest op *sociale netwerken*, waarvan is aangetoond dat ze een grote invloed hebben op de manier waarop mensen (en ook kunstmatige agenten) hun gedrag veranderen als gevolg van interacties met anderen, en op basis van buur-relaties zoals geordend in een bepaalde netwerkstructuur. Bestaand werk dat heeft onderzocht hoe 'genetwerkte' agenten hun gedrag veranderen in sociale dilemmas, is vooralsnog beperkt geweest tot sociale dilemma's met een discreet, klein aantal mogelijke acties (bijvoorbeeld twee). Wij zijn in ons werk geïnteresseerd in meer realistische sociale dilemma's. Daarom onderzoeken we hoe netwerkstructuur het gedrag van agenten beïnvloedt als agenten interacteren in sociale dilemma's met een continuum van acties. We laten zien dat een aantal mechanismen die aantoonbaar werken om gewenst gedrag te bevorderen in discrete dilemma's, ook werken in continue dilemma's (het beste voorbeeld hier is de mogelijkheid voor agenten om af te zien van actie door de verbinding te verbreken tussen henzelf en een ongewenste buur in het netwerk). Er zijn echter ook een aantal mechanismen die geen merkbaar positief effect hebben in continue dilemma's (een voorbeeld is reputatie).

Het proefschrift wordt afgerond in Hoofdstuk 7, waarin we onze onderzoeksvragen beantwoorden, onze bevindingen samenvatten, de probleemstelling beantwoorden, en kijken naar mogelijkheden voor toekomstig werk. We laten zien dat door mensen geïnspireerde eerlijkheid in multi-agent systemen verkregen kan worden door middel van vier stappen. Ten eerste worden er experimenten met mensen uitgevoerd. Ten tweede modelleren we menselijke eerlijkheid in descriptieve modellen. Ten derde bepalen we de fundamenten van computationele modellen van door mensen geïnspireerde eerlijkheid. Ten vierde vertalen we descriptieve modellen naar computationele modellen die worden gebaseerd op de fundamenten. Gebruik makend van de drie computationele modellen in dit proefschrift, mogen we concluderen dat agenten in staat zijn om goede, gewenste oplossingen te vinden in sociale dilemma's, zelfs als er agenten aanwezig zijn die niet geïnteresseerd zijn in eerlijkheid, en die daarom proberen een gewenste, eerlijke oplossing te ondergraven.

Curriculum vitae

Steven de Jong was born in Heerlen, The Netherlands, on October 6, 1981. In the same city, he attended primary school and high school (Gymnasium B, graduated 1999, *cum laude*). He then started to study Architecture (Eindhoven University of Technology), but exchanged this study for Knowledge Engineering (Maastricht University) after his propedeutics.

In Maastricht, he quickly gained the attention of the teachers. He was offered a position as a student assistent by the end of his first year. As a student assistent, he worked on the LOK project with Ben Torben-Nielsen and Rembrandt Puijker, developing educational software assisting students in learning epistemic logic as well as the subsumption archictecture. The latter resulted in the notorious program SB-MASE. He also worked on the ANITA project with Femke de Jonge, resulting in a prototype of multi-agent software for safe information sharing between different departments of the Dutch police.

In 2004, he graduated (*cum laude*) with a Master thesis entitled "Hybrid AI approaches for playing resource management games", supervised by Prof. H.J. van den Herik, Prof. E.O. Postma, Dr. N. Roos, and Dr. P.H.M. Spronck. He immediately started to work as a Ph.D. student at the Institute of Knowledge and Agent Technology (IKAT), which would later change its name to Maastricht ICT Competence Centre (MICC), and then again to the Department of Knowledge Engineering (DKE).

During 4 years of being a Ph.D. student, he investigated various topics in collaboration with many colleagues, eventually settling on his own topic, as reported on in this thesis. He also was actively involved in education, i.e., by guiding skills classes, giving lectures, and supervising four Bachelor students, as well as four Master students. After obtaining his Ph.D., Steven continues his research as a postdoctoral researcher in the group led by Prof. Katia Sycara at Carnegie Mellon University, Pittsburgh, PA, United States.

In his spare time, Steven enjoys words, music notes, and pictures, and tries to be actively involved in each of these art forms. Moreover, he enjoys his friends, traveling, swimming, hiking, as well as Mediterranean food. Finally, he thinks LEGO bricks may be the best invention of the past century (but why are they not perfectly cubic...).

4 June 2009

Publications

The scientific work performed during the author's four-year Ph.D. resulted in the following publications, which are sorted by date.

S. de Jong, P. Spronck, and N. Roos. Requirements for resource management game AI. *Reasoning, Representation, and Learning in Computer Games: Proceedings of the IJCAI Workshop*, Technical Report AIC-05-127, pp. 43-48, 2005.

S. de Jong, N. Roos, and I. Sprinkhuizen-Kuyper. Evolutionary planning heuristics in production management. *Proceedings of the Belgisch-Nederlandse AI Conferentie (BNAIC)*, pp. 96–103, 2005.

S. de Jong and B. Torben-Nielsen. Biologically inspired robust robotics (extended abstract). *Proceedings of Lerende Oplossingen*, 2005.

B. Torben-Nielsen and S. de Jong. Biologically realistic self-repair applied to robotics (extended abstract). *Proceedings of Lerende Oplossingen*, 2005.

G. Chaslot, S. de Jong, J. Saito, and J. Uiterwijk. Monte-Carlo Tree Search in production management problems. *Proceedings of the Belgisch-Nederlandse AI Conferentie (BNAIC)*, pp. 91–98, 2006.

S. de Jong, K. Tuyls, T. Hashimoto, and H. Iida. Scalable potential-field multi-agent coordination in resource distribution tasks. *Proceedings of ALAAMAS, an AAMAS workshop*, 2006.

S. de Jong, K. Tuyls, and I. Sprinkhuizen-Kuyper. Nature-inspired multi-agent coordination in task assignment problems. *Proceedings of the ALAMAS workshop*, 2006.

S. de Jong, K. Tuyls, and I. Sprinkhuizen-Kuyper. Robust and scalable coordination of potential-field driven agents. *Proceedings of the International Conference on Intelligent Agents, Web Technologies and Internet Commerce (IAWTIC)*, pp. 230–242, 2006.

N. Lemmens, S. de Jong, K. Tuyls, and A. Nowé. A bee algorithm for multi-agent systems: recruitment and navigation combined. *Proceedings of ALAg, an AAMAS workshop*, pp. 66–70, 2007.

N. Lemmens, S. de Jong, K. Tuyls, and A. Nowé. Bee system with inhibition pheromones. *Proceedings of the European Conference on Complex Systems (ECCS)*, (electronic proceedings), 2007. Short version also published in: *Proceedings of the 19th Belgian-Dutch Conference on Artificial Intelligence (BNAIC)*, pp. 373–375, 2007.

N. Lemmens, S. de Jong, K. Tuyls, and A. Nowé. Bee behaviour in multi-agent systems: a bee foraging algorithm. *Adaptive Agents and Multi-Agent Systems III - Lecture Notes in Artificial Intelligence*, Vol. 4865:145–156, 2008.

S. de Jong, K. Tuyls, K. Verbeeck, and N. Roos. Priority awareness: towards a computational model of human fairness for multi-agent systems. *Adaptive Agents and Multi-Agent Systems III - Lecture Notes in Artificial Intelligence*, Vol. 4865:117–128, 2008.

S. de Jong, K. Tuyls, and K. Verbeeck. Artificial agents learning human fairness. *Proceedings of the 7th International Conference on Adaptive Agents and Multi-Agent Systems (AAMAS)*, pp. 863–870, 2008.

S. de Jong, K. Tuyls, and K. Verbeeck. Fairness in multi-agent systems. *Knowledge Engineering Review*, Vol. 23(2):153-180, 2008.

S. de Jong, R. van de Ven, and K. Tuyls. The influence of physical appearance on a fair share, *Proceedings of the Belgisch-Nederlandse AI Conferentie (BNAIC)*, pp. 105-112, 2008.

M. Ponsen, K. Tuyls, S. de Jong, J. Ramon, T. Croonenborghs, and K. Driessens. The dynamics of human behaviour in Poker. *Proceedings of the 20th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC)*, pp. 225–232, 2008. Also presented at: *The European Conference on Complex Systems (ECCS)*, 2008.

S. de Jong, and K. Tuyls. Learning to cooperate in public-goods interactions. *Presented at the European Workshop on Multi-Agent Systems (EUMAS)*, 2008.

S. de Jong, and K. Tuyls. Learning to cooperate in a continuous tragedy of the commons. *Proceedings of the 8th International Conference on Adaptive Agents and Multi-Agent Systems (AAMAS)*, (in press), 2009.

S. de Jong, S. Uyttendaele, and K. Tuyls. Learning to reach agreement in continuous strategy spaces. *Journal of Artificial Intellligence Research*, Vol. 33:551-574, 2008.

SIKS dissertations series

1998

1 Johan van den Akker (CWI) *DEGAS - An Active, Temporal Database of Autonomous Objects*

2 Floris Wiesman (UM) *Information Retrieval by Graphically Browsing Meta-Information*

3 Ans Steuten (TUD) *A Contribution to the Linguistic Analysis of Business Conversations within the Language/Action Perspective*

4 Dennis Breuker (UM) *Memory versus Search in Games*

5 Eduard W. Oskamp (RUL) *Computerondersteuning bij Straftoemeting*

1999

1 Mark Sloof (VU) *Physiology of Quality Change Modelling; Automated Modelling of Quality Change of Agricultural Products*

2 Rob Potharst (EUR) *Classification using Decision Trees and Neural Nets*

3 Don Beal (UM) *The Nature of Minimax Search*

4 Jacques Penders (UM) *The Practical Art of Moving Physical Objects*

5 Aldo de Moor (KUB) *Empowering Communities: A Method for the Legitimate User-Driven Specification of Network Information Systems*

6 Niek J.E. Wijngaards (VU) *Re-Design of Compositional Systems*

7 David Spelt (UT) *Verification Support for Object Database Design*

8 Jacques H.J. Lenting (UM) *Informed Gambling: Conception and Analysis of a Multi-Agent Mechanism for Discrete Reallocation*

2000

1 Frank Niessink (VU) *Perspectives on Improving Software Maintenance*

2 Koen Holtman (TU/e) *Prototyping of CMS Storage Management*

3 Carolien M.T. Metselaar (UvA) *Sociaal-organisatorische Gevolgen van Kennistechnologie; een Procesbenadering en Actorperspectief*

4 Geert de Haan (VU) *ETAG, A Formal Model of Competence Knowledge for User Interface Design*

5 Ruud van der Pol (UM) *Knowledge-Based Query Formulation in Information Retrieval*

6 Rogier van Eijk (UU) *Programming Languages for Agent Communication*

7 Niels Peek (UU) *Decision-Theoretic Planning of Clinical Patient Management*

8 Veerle Coupé (EUR) *Sensitivity Analyis of Decision-Theoretic Networks*

9 Florian Waas (CWI) *Principles of Probabilistic Query Optimization*

10 Niels Nes (CWI) *Image Database Management System Design Considerations, Algorithms and Architecture*

11 Jonas Karlsson (CWI) *Scalable Distributed Data Structures for Database Management*

2001

1 Silja Renooij (UU) *Qualitative Approaches to Quantifying Probabilistic Networks*

2 Koen Hindriks (UU) *Agent Programming Languages: Programming with Mental Models*

3 Maarten van Someren (UvA) *Learning as Problem Solving*

4 Evgueni Smirnov (UM) *Conjunctive and Disjunctive Version Spaces with Instance-Based Boundary Sets*

5 Jacco van Ossenbruggen (VU) *Processing Structured Hypermedia: A Matter of Style*

Abbreviations. SIKS – Dutch Research School for Information and Knowledge Systems; CWI – Centrum voor Wiskunde en Informatica, Amsterdam; EUR – Erasmus Universiteit, Rotterdam; KUB – Katholieke Universiteit Brabant, Tilburg; KUN – Katholieke Universiteit Nijmegen; RUG – Rijksuniversiteit Groningen; RUL – Rijksuniversiteit Leiden; RUN – Radboud Universiteit Nijmegen; TUD – Technische Universiteit Delft; TU/e – Technische Universiteit Eindhoven; UL – Universiteit Leiden; UM – Universiteit Maastricht; UT – Universiteit Twente; UU – Universiteit Utrecht; UvA – Universiteit van Amsterdam; UvT – Universiteit van Tilburg; VU – Vrije Universiteit, Amsterdam.

6 Martijn van Welie (VU) *Task-Based User Interface Design*

7 Bastiaan Schonhage (VU) *Diva: Architectural Perspectives on Information Visualization*

8 Pascal van Eck (VU) *A Compositional Semantic Structure for Multi-Agent Systems Dynamics*

9 Pieter Jan 't Hoen (RUL) *Towards Distributed Development of Large Object-Oriented Models, Views of Packages as Classes*

10 Maarten Sierhuis (UvA) *Modeling and Simulating Work Practice BRAHMS: a Multiagent Modeling and Simulation Language for Work Practice Analysis and Design*

11 Tom M. van Engers (VU) *Knowledge Management: The Role of Mental Models in Business Systems Design*

2002

1 Nico Lassing (VU) *Architecture-Level Modifiability Analysis*

2 Roelof van Zwol (UT) *Modelling and Searching Web-based Document Collections*

3 Henk Ernst Blok (UT) *Database Optimization Aspects for Information Retrieval*

4 Juan Roberto Castelo Valdueza (UU) *The Discrete Acyclic Digraph Markov Model in Data Mining*

5 Radu Serban (VU) *The Private Cyberspace Modeling Electronic Environments Inhabited by Privacy-Concerned Agents*

6 Laurens Mommers (UL) *Applied Legal Epistemology; Building a Knowledge-based Ontology of the Legal Domain*

7 Peter Boncz (CWI) *Monet: A Next-Generation DBMS Kernel For Query-Intensive Applications*

8 Jaap Gordijn (VU) *Value Based Requirements Engineering: Exploring Innovative E-Commerce Ideas*

9 Willem-Jan van den Heuvel (KUB) *Integrating Modern Business Applications with Objectified Legacy Systems*

10 Brian Sheppard (UM) *Towards Perfect Play of Scrabble*

11 Wouter C.A. Wijngaards (VU) *Agent Based Modelling of Dynamics: Biological and Organisational Applications*

12 Albrecht Schmidt (UvA) *Processing XML in Database Systems*

13 Hongjing Wu (TU/e) *A Reference Architecture for Adaptive Hypermedia Applications*

14 Wieke de Vries (UU) *Agent Interaction: Abstract Approaches to Modelling, Programming and Verifying Multi-Agent Systems*

15 Rik Eshuis (UT) *Semantics and Verification of UML Activity Diagrams for Workflow Modelling*

16 Pieter van Langen (VU) *The Anatomy of Design: Foundations, Models and Applications*

17 Stefan Manegold (UvA) *Understanding, Modeling, and Improving Main-Memory Database Performance*

2003

1 Heiner Stuckenschmidt (VU) *Ontology-Based Information Sharing in Weakly Structured Environments*

2 Jan Broersen (VU) *Modal Action Logics for Reasoning About Reactive Systems*

3 Martijn Schuemie (TUD) *Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy*

4 Milan Petkovic (UT) *Content-Based Video Retrieval Supported by Database Technology*

5 Jos Lehmann (UvA) *Causation in Artificial Intelligence and Law – A Modelling Approach*

6 Boris van Schooten (UT) *Development and Specification of Virtual Environments*

7 Machiel Jansen (UvA) *Formal Explorations of Knowledge Intensive Tasks*

8 Yong-Ping Ran (UM) *Repair-Based Scheduling*

9 Rens Kortmann (UM) *The Resolution of Visually Guided Behaviour*

10 Andreas Lincke (UT) *Electronic Business Negotiation: Some Experimental Studies on the Interaction between Medium, Innovation Context and Cult*

11 Simon Keizer (UT) *Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks*

12 Roeland Ordelman (UT) *Dutch Speech Recognition in Multimedia Information Retrieval*

13 Jeroen Donkers (UM) *Nosce Hostem – Searching with Opponent Models*

14 Stijn Hoppenbrouwers (KUN) *Freezing Language: Conceptualisation Processes across ICT-Supported Organisations*

15 Mathijs de Weerdt (TUD) *Plan Merging in Multi-Agent Systems*

16 Menzo Windhouwer (CWI) *Feature Grammar Systems - Incremental Maintenance of Indexes to Digital Media Warehouse*

17 David Jansen (UT) *Extensions of Statecharts with Probability, Time, and Stochastic Timing*

18 Levente Kocsis (UM) *Learning Search Decisions*

2004

1 Virginia Dignum (UU) *A Model for Organizational Interaction: Based on Agents, Founded in Logic*

2 Lai Xu (UvT) *Monitoring Multi-party Contracts for E-business*

3 Perry Groot (VU) *A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving*

4 Chris van Aart (UvA) *Organizational Principles for Multi-Agent Architectures*

5 Viara Popova (EUR) *Knowledge Discovery and Monotonicity*

6 Bart-Jan Hommes (TUD) *The Evaluation of Business Process Modeling Techniques*

7 Elise Boltjes (UM) *Voorbeeld$_{IG}$ Onderwijs; Voorbeeldgestuurd Onderwijs, een Opstap naar Abstract Denken, vooral voor Meisjes*

8 Joop Verbeek (UM) *Politie en de Nieuwe Internationale Informatiemarkt, Grensregionale Politiële Gegevensuitwisseling en Digitale Expertise*

9 Martin Caminada (VU) *For the Sake of the Argument; Explorations into Argument-based Reasoning*

10 Suzanne Kabel (UvA) *Knowledge-rich Indexing of Learning-objects*

11 Michel Klein (VU) *Change Management for Distributed Ontologies*

12 The Duy Bui (UT) *Creating Emotions and Facial Expressions for Embodied Agents*

13 Wojciech Jamroga (UT) *Using Multiple Models of Reality: On Agents who Know how to Play*

14 Paul Harrenstein (UU) *Logic in Conflict. Logical Explorations in Strategic Equilibrium*

15 Arno Knobbe (UU) *Multi-Relational Data Mining*

16 Federico Divina (VU) *Hybrid Genetic Relational Search for Inductive Learning*

17 Mark Winands (UM) *Informed Search in Complex Games*

18 Vania Bessa Machado (UvA) *Supporting the Construction of Qualitative Knowledge Models*

19 Thijs Westerveld (UT) *Using generative probabilistic models for multimedia retrieval*

20 Madelon Evers (Nyenrode) *Learning from Design: facilitating multidisciplinary design teams*

2005

1 Floor Verdenius (UvA) *Methodological Aspects of Designing Induction-Based Applications*

2 Erik van der Werf (UM) *AI techniques for the game of Go*

3 Franc Grootjen (RUN) *A Pragmatic Approach to the Conceptualisation of Language*

4 Nirvana Meratnia (UT) *Towards Database Support for Moving Object data*

5 Gabriel Infante-Lopez (UvA) *Two-Level Probabilistic Grammars for Natural Language Parsing*

6 Pieter Spronck (UM) *Adaptive Game AI*

7 Flavius Frasincar (TU/e) *Hypermedia Presentation Generation for Semantic Web Information Systems*

8 Richard Vdovjak (TU/e) *A Model-driven Approach for Building Distributed Ontology-based Web Applications*

9 Jeen Broekstra (VU) *Storage, Querying and Inferencing for Semantic Web Languages*

10 Anders Bouwer (UvA) *Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments*

11 Elth Ogston (VU) *Agent Based Matchmaking and Clustering - A Decentralized Approach to Search*

12 Csaba Boer (EUR) *Distributed Simulation in Industry*

13 Fred Hamburg (UL) *Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen*

14 Borys Omelayenko (VU) *Web-Service configuration on the Semantic Web; Exploring how semantics meets pragmatics*

15 Tibor Bosse (VU) *Analysis of the Dynamics of Cognitive Processes*

16 Joris Graaumans (UU) *Usability of XML Query Languages*

17 Boris Shishkov (TUD) *Software Specification Based on Re-usable Business Components*

18 Danielle Sent (UU) *Test-selection strategies for probabilistic networks*

19 Michel van Dartel (UM) *Situated Representation*

20 Cristina Coteanu (UL) *Cyber Consumer Law, State of the Art and Perspectives*

21 Wijnand Derks (UT) *Improving Concurrency and Recovery in Database Systems by Exploiting Application Semantics*

2006

1 Samuil Angelov (TU/e) *Foundations of B2B Electronic Contracting*

2 Cristina Chisalita (VU) *Contextual issues in the design and use of information technology in organizations*

3 Noor Christoph (UvA) *The role of metacognitive skills in learning to solve problems*

4 Marta Sabou (VU) *Building Web Service Ontologies*

5 Cees Pierik (UU) *Validation Techniques for Object-Oriented Proof Outlines*

6 Ziv Baida (VU) *Software-aided Service Bundling - Intelligent Methods & Tools for Graphical Service Modeling*

7 Marko Smiljanic (UT) *XML schema matching – balancing efficiency and effectiveness by means of clustering*

8 Eelco Herder (UT) *Forward, Back and Home Again - Analyzing User Behavior on the Web*

9 Mohamed Wahdan (UM) *Automatic Formulation of the Auditor's Opinion*

10 Ronny Siebes (VU) *Semantic Routing in Peer-to-Peer Systems*

11 Joeri van Ruth (UT) *Flattening Queries over Nested Data Types*

12 Bert Bongers (VU) *Interactivation - Towards an e-cology of people, our technological environment, and the arts*

13 Henk-Jan Lebbink (UU) *Dialogue and Decision Games for Information Exchanging Agents*

14 Johan Hoorn (VU) *Software Requirements: Update, Upgrade, Redesign - towards a Theory of Requirements Change*

15 Rainer Malik (UU) *CONAN: Text Mining in the Biomedical Domain*

16 Carsten Riggelsen (UU) *Approximation Methods for Efficient Learning of Bayesian Networks*

17 Stacey Nagata (UU) *User Assistance for Multitasking with Interruptions on a Mobile Device*

18 Valentin Zhizhkun (UvA) *Graph transformation for Natural Language Processing*

19 Birna van Riemsdijk (UU) *Cognitive Agent Programming: A Semantic Approach*

20 Marina Velikova (UvT) *Monotone models for prediction in data mining*

21 Bas van Gils (RUN) *Aptness on the Web*

22 Paul de Vrieze (RUN) *Fundaments of Adaptive Personalisation*

23 Ion Juvina (UU) *Development of Cognitive Model for Navigating on the Web*

24 Laura Hollink (VU) *Semantic Annotation for Retrieval of Visual Resources*

25 Madalina Drugan (UU) *Conditional log-likelihood MDL and Evolutionary MCMC*

26 Vojkan Mihajlovic (UT) *Score Region Algebra: A Flexible Framework for Structured Information Retrieval*

27 Stefano Bocconi (CWI) *Vox Populi: generating video documentaries from semantically annotated media repositories*

28 Borkur Sigurbjornsson (UvA) *Focused Information Access using XML Element Retrieval*

2007

1 Kees Leune (UvT) *Access Control and Service-Oriented Architectures*

2 Wouter Teepe (RUG) *Reconciling Information Exchange and Confidentiality: A Formal Approach*

3 Peter Mika (VU) *Social Networks and the Semantic Web*

4 Jurriaan van Diggelen (UU) *Achieving Semantic Interoperability in Multi-agent Systems: a dialogue-based approach*

5 Bart Schermer (UL) *Software Agents, Surveillance, and the Right to Privacy: a Legislative Framework for Agent-enabled Surveillance*

6 Gilad Mishne (UvA) *Applied Text Analytics for Blogs*

7 Natasa Jovanovic' (UT) *To Whom It May Concern - Addressee Identification in Face-to-Face Meetings*

8 Mark Hoogendoorn (VU) *Modeling of Change in Multi-Agent Organizations*

9 David Mobach (VU) *Agent-Based Mediated Service Negotiation*

10 Huib Aldewereld (UU) *Autonomy vs. Conformity: an Institutional Perspective on Norms and Protocols*

11 Natalia Stash (TU/e) *Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System*

12 Marcel van Gerven (RUN) *Bayesian Networks for Clinical Decision Support: A Rational Approach to Dynamic Decision-Making under Uncertainty*

13 Rutger Rienks (UT) *Meetings in Smart Environments; Implications of Progressing Technology*

14 Niek Bergboer (UM) *Context-Based Image Analysis*

15 Joyca Lacroix (UM) *NIM: a Situated Computational Memory Model*

16 Davide Grossi (UU) *Designing Invisible Handcuffs. Formal investigations in Institutions and Organizations for Multi-agent Systems*

17 Theodore Charitos (UU) *Reasoning with Dynamic Networks in Practice*

18 Bart Orriens (UvT) *On the development and management of adaptive business collaborations*

19 David Levy (UM) *Intimate relationships with artificial partners*

20 Slinger Jansen (UU) *Customer Configuration Updating in a Software Supply Network*

21 Karianne Vermaas (UU) *Fast diffusion and broadening use: A research on residential adoption and usage of broadband internet in the Netherlands between 2001 and 2005*

22 Zlatko Zlatev (UT) *Goal-oriented design of value and process models from patterns*

23 Peter Barna (TU/e) *Specification of Application Logic in Web Information Systems*

24 Georgina Ramírez Camps (CWI) *Structural Features in XML Retrieval*

25 Joost Schalken (VU) *Empirical Investigations in Software Process Improvement*

2008

1 Katalin Boer-Sorbán (EUR) *Agent-Based Simulation of Financial Markets: A modular, continuous-time approach*

2 Alexei Sharpanskykh (VU) *On Computer-Aided Methods for Modeling and Analysis of Organizations*

3 Vera Hollink (UvA) *Optimizing hierarchical menus: a usage-based approach*

4 Ander de Keijzer (UT) *Management of Uncertain Data - towards unattended integration*

5 Bela Mutschler (UT) *Modeling and simulating causal dependencies on process-aware information systems from a cost perspective*

6 Arjen Hommersom (RUN) *On the Application of Formal Methods to Clinical Guidelines, an Artificial Intelligence Perspective*

7 Peter van Rosmalen (OU) *Supporting the tutor in the design and support of adaptive e-learning*

8 Janneke Bolt (UU) *Bayesian Networks: Aspects of Approximate Inference*

9 Christof van Nimwegen (UU) *The paradox of the guided user: assistance can be counter-effective*

10 Wauter Bosma (UT) *Discourse oriented Summarization*

11 Vera Kartseva (VU) *Designing Controls for Network Organizations: a Value-Based Approach*

12 Jozsef Farkas (RUN) *A Semiotically oriented Cognitive Model of Knowlegde Representation*

13 Caterina Carraciolo (UvA) *Topic Driven Access to Scientific Handbooks*

14 Arthur van Bunningen (UT) *Context-Aware Querying; Better Answers with Less Effort*

15 Martijn van Otterlo (UT) *The Logic of Adaptive Behavior: Knowledge Representation and Algorithms for the Markov Decision Process Framework in First-Order Domains*

16 Henriette van Vugt (VU) *Embodied Agents from a User's Perspective*

17 Martin Op't Land (TUD) *Applying Architecture and Ontology to the Splitting and Allying of Enterprises*

18 Guido de Croon (UM) *Adaptive Active Vision*

19 Henning Rode (UT) *From document to entity retrieval: improving precision and performance of focused text search*

20 Rex Arendsen (UvA) *Geen bericht, goed bericht. Een onderzoek naar de effecten van de introductie van elektronisch berichtenverkeer met een overheid op de administratieve lasten van bedrijven*

21 Krisztian Balog (UvA) *People search in the enterprise*

22 Henk Koning (UU) *Communication of IT-architecture*

23 Stefan Visscher (UU) *Bayesian network models for the management of ventilator-associated pneumonia*

24 Zharko Aleksovski (VU) *Using background knowledge in ontology matching*

25 Geert Jonker (UU) *Efficient and Equitable exchange in air traffic management plan repair using spender-signed currency*

26 Marijn Huijbregts (UT) *Segmentation, diarization and speech transcription: surprise data unraveled*

27 Hubert Vogten (OU) *Design and implementation strategies for IMS learning design*

28 Ildiko Flesh (RUN) *On the use of independence relations in Bayesian networks*

29 Dennis Reidsma (UT) *Annotations and subjective machines- Of annotators, embodied agents, users, and other humans*

30 Wouter van Atteveldt (VU) *Semantic network analysis: techniques for extracting, representing and querying media content*

31 Loes Braun (UM) *Pro-active medical information retrieval*

32 Trung B. Hui (UT) *Toward affective dialogue management using partially observable markov decision processes*

33 Frank Terpstra (UvA) *Scientific workflow design; theoretical and practical issues*

34 Jeroen de Knijf (UU) *Studies in Frequent Tree Mining*

35 Benjamin Torben-Nielsen (UvT) *Dendritic morphology: function shapes structure*

2009

1 Rasa Jurgelenaite (RUN) *Symmetric Causal Independence Models*

2 Willem Robert van Hage (VU) *Evaluating Ontology-Alignment Techniques*

3 Hans Stol (UvT) *A Framework for Evidence-based Policy Making Using IT*

4 Josephine Nabukenya (RUN) *Improving the Quality of Organisational Policy Making using Collaboration Engineering*

5 Sietse Overbeek (RUN) *Bridging Supply and Demand for Knowledge Intensive Tasks - Based on Knowledge, Cognition, and Quality*

6 Muhammad Subianto (UU) *Understanding Classification*

7 Ronald Poppe (UT) *Discriminative Vision-Based Recovery and Recognition of Human Motion*

8 Volker Nannen (VU) *Evolutionary Agent-Based Policy Analysis in Dynamic Environments*

9 Benjamin Kanagwa (RUN) *Design, Discovery and Construction of Service-oriented Systems*

10 Jan Wielemaker (UVA) *Logic programming for knowledge-intensive interactive applications*

11 Alexander Boer (UVA) *Legal Theory, Sources of Law & the Semantic Web*

12 Peter Massuthe (TUE, Humboldt-Universtät zu Berlin) *Operating Guidelines for Services*

13 Steven de Jong (UM) *Fairness in Multi-Agent Systems*