The Effects of Cooperative Agent Behavior on Human Cooperativeness

Arlette van Wissen, Jurriaan van Diggelen, and Virginia Dignum

Institute of Information and Computing Sciences,
Utrecht University, the Netherlands
arlette.vanwissen@phil.uu.nl; {jurriaan,virginia}@cs.uu.nl

Abstract. Within negotiation environments, cooperative behavior may emerge from different factors. In this paper, we focus on human-agent interaction and in particular on the question of how the cooperativeness of a software agent affects the cooperativeness of a human player. We implemented three different kinds of agent behavior to determine how they influence helpful human behavior. Our data shows that humans behave more cooperatively towards agents that negotiate with them in a cooperative way and that humans tend to punish egoistic and unfair play by behaving non-cooperatively themselves.

Keywords: human-agent interaction, negotiation, cooperation, Ultimatum Game

1 Introduction

Social relations between software agents and humans is becoming an essential part of our life. Both in everyday life and science humans use, control and cooperate with agents [1]. Many forms of human-machine interaction involve participants that pursue different, sometimes even contradictory, interests. In these scenarios, humans and agents are autonomous entities which interact and cooperate with each other while keeping in mind their own objectives and goals. In this work, we use the word 'agents' to refer solely to software agents and 'actors' to refer to both agents and humans. The interactions between these actors take place in many domains such as games [2], where humans play against agents with contradictory goals, and online auctions [3], where humans and agents compete over the same product. To achieve successful human-machine interaction, we must understand the factors involved in this interaction.

The area of human-agent interaction has long been neglected, ignoring the issue of how the interaction affects the behavior and strategies of the involved actors. Most work concerned with human-agent interaction focuses mainly on how an agent should be designed to model its opponent or on how to design most efficient agent strategies. Still only a very limited number of studies shed light on the social phenomena that occur between humans and agents and on

the influence of this interaction on human behavior [4–6].

This work presents an experimental study of the social phenomena that play a role in human-agent interactions with contradictory interests. In this project we combine and extend existing work to understand the effect of agent behavior on the cooperativeness of people towards their opponents. More specifically, we explore to what extent social principles like reciprocity and fairness influence human bidding behavior in these interactions. This paper is a more elaborate description of the work we presented in [7]. The central question we address is: how does the cooperativeness of an agent affect the cooperativeness of a human player? The study of human bidding behavior in relation to agents with competing goals may provide new insights into how bidding decisions are made and how agents' behaviors are perceived. This knowledge helps to establish successful human-agent interaction.

Since our work concentrates on the interaction between humans and agents, our research method combines methods and results from both behavioral and computer sciences. On the one hand we study human behavior by means of experimental research. We build upon earlier research showing that human reasoning and negotiation strategies are affected by social factors, such as altruism, fairness and reciprocity [4, 8]. On the other hand, we use observations from game theory of cooperation between agents and succesful reciprocal agent strategies. A mixture of punishing defection and rewarding cooperation seems to encourage cooperation [9]. It has also been shown that a significant difference between the reaction of people to agent and to human opponents exists [5, 10]. We combined these observations to test two main hypotheses. First, humans will behave cooperatively towards an altruistic opponent. Second, they will punish egoistic behavior. These hypotheses are discussed in more detail in Section 3.3. Both hypotheses assume similar behavior towards the human and agent opponents and therefore that social tendencies predominate over any prejudices about the nature of the opponent.

To analyze differences in behavior between various kinds of human-agent interactions we use Colored Trails (CT) [11], which is a negotiation environment for designing and evaluating decision-making behavior and dynamics between players. CT is a conceptually simple but strategically complex game that has been developed for testing strategies for automated agents that operate in groups. In our experiment, mixed groups of humans and agents play the game. We implement different agent strategies that vary in their degree of cooperativeness.

We create a CT setup that enables multiple humans to play the game simultaneously against alternate opponents. In some conditions of the experiment subjects were not aware of the identity of their opponent, i.e. they did not know whether they were playing against a human or an agent. This enables us to analyze the relation between the cooperativeness of agents and that of humans.

Our results support our hypotheses and show three main findings. First, humans tend to be more cooperative towards a cooperative opponent and do not fully exploit altruistic behavior. Second, humans play less cooperatively towards egoistic opponents. Third, an egoistic opponent is often perceived as a human.

The paper is organized as follows. First, we discuss related work and show how our work differs from it. The conceptual approach to our experiment is presented in Section 3. This section is mainly concerned with introducing the Colored Trails game framework and demonstrating how we use it in this project. In Section 4 the experimental setting will be addressed. We will discuss the results of the experiments in Section 5, followed by a brief conclusion and outline for future work in Section 6.

2 Related Work

This paper may be placed at the intersection of studies on human negotiation behavior and studies on agent negotiation behavior. It therefore builds on three streams of research: human-human interaction, agent-agent interaction and human-agent interaction. These fields are briefly discussed in the following three subsections.

2.1 Human-Human Interaction

The Ultimatum Game [12] is a commonly used game setting to simulate and analyze negotiation behavior that was originally developed by experimental economists. In the classic ultimatum game, two people are given the task to divide a sum of money. One player proposes how to divide the money between them and the other player can accept or reject this proposal. If the second player rejects, neither player receives anything. In the original game, the players interact only once. As it turns out, people usually offer 'fair' (i.e., 50:50) splits, and most offers of less than 20% are rejected. This latter phenomenon is referred to as *inequity aversion*.

Many experiments have been conducted that confirm these properties of human behavior. Social factors, such as altruism, fairness and reciprocity are shown to have an influence on negotiation behavior [8, 13]. When playing against other humans, humans tend to be cooperative towards their opponents but punish them severely in case of unfairness. As soon as human actors are involved, people seem to develop expectations of how other people should behave. People not only care about the outcome of actions; they are also concerned with how they come about [14]. These results show that humans do not always pursue their ultimate utility. Apparently, humans are not completely rational actors as assumed in standard economic game theory [15]. In our study, we use these results to formulate the hypothesis that human expectations of an opponent are also applicable to agents.

2.2 Agent-Agent Interaction

Negotiation between intelligent agents has been widely studied over the past years. Mostly, a Multi-Agent System (MAS) is assumed to consist solely of computer agents. Interacting agents are faced with the challenge of modeling other

agents and their behavior in order to adapt to their environment [16]. Much research has been done on social reasoning models for agents. These models, as opposed to the models which are fully based on game theory, may enable the agent to be more adaptive to the environment and to perform better in more realistic negotiation scenarios [17]. Axelrod [9] has demonstrated that in some MAS's, it can be beneficial for agents to use a cooperative strategy. In these situations cooperative agents do as well as or outperform competitive agents. Moreover, reciprocal behavior allows cooperators to do better than self-interested rational agents [18]. Models that take social principles such as fairness and helpfulness into account were shown to explore new negotiation opportunities [19] and find solutions that correspond to solutions found by humans [17].

2.3 Human-Agent Interaction

Our study is based on two seemingly contradictory findings from empirical research. On the one hand, people seem to display the same social behavior towards agents as they do to humans. The negotiation experiments presented in [4] seem to imply that human responders still care about equality of outcomes while negotiating with agents. On the other hand, experiments indicate that a significant difference exists between the behavior of humans towards agent and human opponents. Humans seem to perceive other humans differently from agents. For example, experiments by Sanfey et al. [5] and Blount [10] show that humans are more likely to accept a proposal from agents than from other humans. More seems to be tolerated in terms of behavior and actions from agent actors than from human actors. Humans and agents seem to have a different set of 'acceptable behavior'. This leaves us with the question how humans actively treat agents in a negotiation setting: Are social factors like fairness and altruism less applicable to agents since humans do not ascribe the same expectations to them?

None of these prior works have investigated the effects of different computational strategies on human behavior in a repeated interaction scenario. Setting up repeated interaction is the key to observing a variety of interesting behavior, since the players are able to react to strategies of their opponent. The work of Gal et al. on reciprocity [20] does evaluate computer models of human behavior in repeated interaction but does not position computer players against people.

3 Experimental Approach

Cooperative behavior is a broadly discussed and controversial subject, since there are many theories about the causes and ultimate motives of (non-)cooperative behavior [21]. We do not aim to give any final explanation of underlying motives for cooperation. Instead, we are concerned with the effect of agent behavior on human behavior. In our work, we use the word *cooperative* for interacting together willingly for a common purpose or benefit. This behavior is said to be *altruistic* if it involves a fitness cost to the proposer and confers a fitness benefit on the responder. It is said to be *egoistic* if the proposer acts only to benefit

himself by increasing his fitness. We use the notion of *helpful behavior* to refer to cooperation that does not necessarily imply a fitness cost for the proposer, but in any case yields a fitness benefit for the responder. *Reciprocal* behavior is conduct that corresponds to the behavior of others. In this study, we created a setting that encourages and instigates these different kinds of cooperation.

3.1 Colored Trails

Colored Trails (CT) is a game developed by Grosz et al. [11] which provides a testbed for investigating interactions between players whose task is to reach a goal by exchanging resources. Computational strategies and human decision making can be studied by comparing interactions in both homogeneous and heterogeneous groups of people and computer systems. CT is played on an N×M board of colored squares in which one or more squares can be designated as a goal. Each player has to reach a, possibly different, goal by exchanging chips. In order to move a player piece to an adjacent position, the player has to hand in a chip that has the same color as the square he wants to move to. It is not possible for to do diagonal moves.

3.2 Conceptual Design

Our experiment consists of two main components to examine human cooperativeness: a game and a questionnaire. In this subsection we will elaborate on the conceptual design of the game. We use the CT framework to implement an environment that is similar to an iterated Ultimatum Game (UG). In the stable outcome of an UG, each player has chosen a strategy and no player will benefit by changing his strategy under the assumption that the other players keep their strategy unchanged. This is called the *Nash equilibrium*. A Nash equilibrium would occur when the proposer offers the smallest possible amount of money (in our case: chips) to the responder and the responder accepts. It would be rational for the responder to always accept if the proposal leaves him with more than 0. However, experimental evidence shows that the proposer offers a relatively large share to his opponent and that the responder often rejects smaller positive amounts [12]. This can be interpreted as human willingness to play fair and to punish 'unfair' splits. The results of Gal et al. in [4] seem to be consistent with this scenario.

We use the UG to test whether humans show similar helpful or punishing behavior when they interact with an agent. A game consists of several rounds in which two players have alternating roles: *proposer* or *responder*. During each round, both players try to obtain all the chips they need in order to reach the goal. The proposer creates a proposal to exchange chips with his opponent. The opponent can either accept or reject this proposal. The proposer can also decide not to exchange any chips. In case of rejection, both players receive nothing. This is defined as *one-shot negotiation*. The variety of proposals that can be created show that our scenario is more complex than an UG, where the decision process in narrowed down to a pre-defined list of choices.

The CT environment enables us to create the following protocol:

1. **Orientation phase.** The player is able to get accustomed to the board and its new chips. During this phase, communication is not possible.
2. **Communication phase.** Both players try to obtain all the chips they need to reach the goal. The proposer can offer one proposal and the responder can react to it.
3. **Redistribution phase.** Agreements are enforced by the game controller (which is implemented in CT): after the proposal is accepted or rejected, the game controller will redistribute the chips (if necessary) according to the proposal.
4. **Movement phase.** If the player has the necessary chips to reach the goal, the program will execute the player's movement by moving its piece to the goal via the shortest possible path. If on the other hand the player does not have the required chips, his piece will not move at all. Again, we use an all-or-none approach to stimulate cooperation.

The scoring function is defined as follows: $s = goal + board\_size - taken\_steps$, where *goal* represents the rewarded points for reaching the goal (100), *board_size* represents the size of the board (20) and *taken_steps* is the number of steps of the taken path (variable per round). The scoring function in this experiment does not stipulate a reward dependence, but only a task dependence. Players did not receive penalties for the amount of chips they had left at the end of a round and their score was not dependent on the opponent's score in any way. Since the player scores in our scenario are independent, any act that is beneficial for the opponent does not increase the player's own fitness and can thus be considered as helpful behavior. Cooperation is stimulated by playing an iterated game. Participants have the prospect of encountering their opponent again in the game and are therefore more inclined to cooperate [22].

We use four CT variables to give an indication of the degree of altruistic or egoistic behavior of the players and their willingness to play fair:

- The offered chips. A higher amount and usefulness for the responder are considered more helpful.
- The requested chips. A lesser amount and usefulness for the responder are considered more helpful.
- The response. Accepting is considered more helpful. Rejecting might indicate an 'unfair' or non-beneficial proposal.
- The pursued path. Chips requested or accepted in order to take a suboptimal path to the goal are considered more helpful.

3.3 Agent Models

We developed three agents with different strategies: the altruistic agent (ALT), the egoistic agent (EGO), and the reciprocal agent (REC). The behavior by these agents is extreme and prototypical. Note that it is not our intention to

Egoistic Agent (EGO)	**proposer**	*offers*	randomly: a) no chips c) b) chips that are not beneficial for opponent
		requests	a) chips that it needs to reach the goal b) chips that enable shorter path
	responder	*accepts*	deals with better score than it can obtain with current chipset
		rejects	all other deals
Altruistic Agent (ALT)	**proposer**	*offers*	chips unrequired to reach goal that are useful for opponent
		requests	chips it needs for a path to goal with least costs for responder
	responder	*accepts*	a) deals in favor of itself b) deals in favor of opponent if it can reach its own goal
		rejects	deals that make it unable to reach the goal
Reciprocal Agent (REC)	**proposer**		a) behaves as egoistic proposer if favor balance =<0 b) behaves as altruistic proposer if favor balance >0
	responder		a) behaves as egoistic responder if favor balance =<0 b) behaves as altruistic responder if favor balance >0

Table 1. The different agents and their strategies

develop agents that resemble accurate human behavior or decision making. Our objective is to observe differences in human behavior that are caused by agent behavior. For that purpose, we implemented agents with different and extreme utility functions. ALT is satisfied with a suboptimal path to reach its goal and will accept almost all requests, thereby creating a good reputation for itself. EGO has a very selfish strategy: it will not grant the opponent any favors and always aims for the shortest path. REC has a mildly adaptive strategy that is a combination of the strategies of EGO and ALT. It uses a *favor balance* that keeps track of how cooperative its opponent has been in the game so far. The favor balance is calculated in the following way: *fb = 1 + pos_encounters - neg_encounters*. The default action of REC is to act altruistically, which is ensured by adding 1 to the favor balance. In case a proposal or response is judged as non-cooperative, the number of *neg_encounters* is increased, in case it is considered positive, it will increase *pos_encounters*. A response of the opponent is considered negative if it is a rejection and is judged positive if it is an acceptance. A proposal of the opponent is considered negative if it would leave the agent with a less advantageous chipset than it had beforehand, and is judged positive otherwise. Table 1 gives an overview of the three agent strategies.

We implemented these strategies to explore two hypotheses. Both hypotheses are motivated by the idea that although humans perceive agents differently from other humans, the human tendency to play fair and to encourage others to do the same will dominate.

Hypothesis 1 Cooperative behavior of the agent encourages cooperative behavior of the human opponent.
More specifically, both the altruistic and reciprocal agent, even though the altruistic agent is vulnerable to exploitation, will receive helpful behavior from their opponent.

Hypothesis 2 Egoistic (and therefore non-cooperative behavior) instigates 'punishment-behavior' of humans. We expect human players to offer the egoistic agent as little as possible and to prevent the agent from reaching its goal.

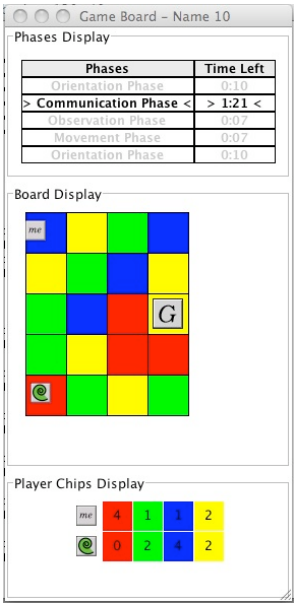

Fig. 1. The configuration of the CT game showing the board, the phases of a round and the players' chipsets

4 Experimental Design

4.1 CT Configuration

CT makes use of a wide variety of parameters that allow for increasing complexity of the game along different dimensions. The games were played with *full board visibility* and *full chip visibility*: both players had complete knowledge of the colors of the squares, both their positions on the board and the chip distribution. The full board and chip visibility allowed players to make deliberate decisions about helping their opponent or not. The PathFinder is an additional tool for players that shows the chips needed to take the shortest path to the goal. In this experiment, players were only allowed to see their own chips in the PathFinder and were not shown any paths longer than the shortest one.

Figure 1 shows our basic CT configuration. The game is played on a 4×5 board with one square designated as a goal. The scoring function, board configuration and positions of the players and the goal on the board remain the same throughout the game. The player to start and the role he[1] performs (proposer or responder) are determined randomly. In each round, the game manager randomly allocates one of ten different sets of chips to the players. The chipsets

[1] Both male and female subjects as well as gender-neutral agents were involved in our experiment. For purposes of convenience and clarity we will refer here to all subjects as male. This is however completely arbitrary.

are composed to stimulate 'interesting' negotiation behavior. There are always enough chips in the game for both players to reach the goal, but the players never both start with a chipset that allows reaching the goal via the shortest path. The strategies of the agents are reflected in the chips they are willing to transfer. For example, it is considered altruistic to propose a chip distribution that allows the responder to reach the goal but leaves the proposer with a longer path to the goal. 'Helpful behavior' entails a broader range of conduct in which a proposer can perform actions that increase the fitness of the responder without necessarily decreasing the proposer's own fitness. In this scenario, a proposer displays helpful behavior if he transfers chips to a responder to help him reach the goal while the proposer can still reach the goal via the shortest path.

4.2 Experimental Setup

A total of 30 subjects participated in the experiment. All participants were graduate and undergraduate students between 20-30 years of age. 80% of them studied Artificial Intelligence, 20% were enrolled in other studies. Almost 50% of the participants were female. Each subject participated in 4 games, adding up to a total of 120 games played. Participants were instructed to perform a non-competitive task: they were to try to maximize their own scores, not to minimize other players' scores. This is also reflected in the scoring function, since the majority of the points can be received by reaching the goal, not by choosing the shortest path. It follows that there is no strictly competitive or zero-sum iteration. A small financial bonus payment would be awarded to the player who obtained the highest score. Players were aware of this, but since the bonus would be awarded to all player with the highest score, we expect the bonus did not stimulate players to minimize their opponent's score, but only to maximize their own score.

Given that our main goal was to examine the differences in human behavior when confronted with (non-)cooperative behavior of an agent, we had to compare this behavior in some way to the behavior of humans towards a human opponent. For this reason, participants played three games against agent opponents and one game against another human participant. The group that played against human opponents was used as a control group to demonstrate the degree of cooperativeness of the players. We expected that telling the participants they would be playing against computer agents could alter their behavior. In order to investigate the effect of knowledge and beliefs of their opponent on the acceptance level of 'unfair' or non-cooperative behavior, we deliberately manipulated their knowledge and beliefs.

Hence the general setup included four independent variables:

1. **Nature** of the player: human or agent.
2. **Strategy** of the agent: altruist, egoist or reciprocal.
3. **Knowledge** of the nature of opponent: True Belief (TB), False Belief (FB), No Belief (NB).
4. **Order** of played opponents: egoist, altruist, reciprocal, human.

We created three groups of ten participants each and combined them with different belief conditions. The belief condition provided the participants only with information about the nature of the opponent; the strategy of the agents was never disclosed. The game controller randomly determined the order of the opponents for each of the three groups. Each game consisted of ten rounds. In the TB-condition we told the participants the truth about whether they were playing against an agent or against another human. In the FB-condition we misled them by announcing their opponent would be a human when in fact it was an agent, and the other way around. Unlike in the other conditions, we did not give the players any information at all about their opponent in the NB-condition.

4.3 Evaluation

At the end of each game, participants were asked to fill out a questionnaire. This questionnaire provided us with insights of how different aspects of cooperation come about. They answered questions about the nature and strategy of their opponent and about their own strategy and cooperativeness. We used this information to find out how the participants perceived their opponents and how this influenced their cooperative behavior.

The logs of each CT game contained all the vital information of the game, such as score, the proposals and responses made, whether both players reached their goal and if so, how many steps it took. These logs provided us with information of how often players made fair trades or helped their opponents.

5 Results

We hypothesized humans to behave more cooperatively towards opponents that behave cooperatively or altruistically themselves (H1). We also expected that egoistic behavior would be punished (H2). Our results support both hypotheses. However, more extensive research with a larger number of participants has to be done to significantly demonstrate these results.

The questionnaire shows that our agents' strategies correctly represent the different types of behavior. 100% of the participants found ALT moderately to very cooperative: 83% found it very cooperative and 17% found it partially cooperative. Interestingly, REC was also considered very cooperative: 97% of the subjects perceived it as cooperative. This can be explained by the fact that the first action REC takes is an altruistic one; that is, it only adopts an egoistic strategy when the opponent treats it in an egoistic way. The majority of the players found EGO to be not cooperative at all but surprisingly a considerable number of subjects believed the agent to be partially (30%), or occasionally even fully (7%), cooperative. In comparison, 53% of the subjects considered their human opponents moderately to very cooperative. As it turned out, participants judged the egoistic agent as being cooperative because it sometimes offered chips when it did not need anything in return. It did not seem to matter that these chips were of no use to the responder.

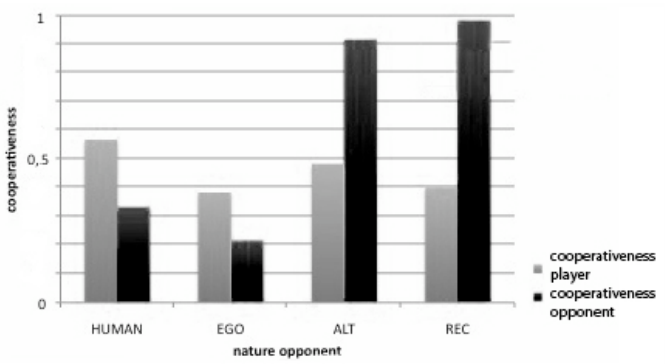

Fig. 2. Results of how cooperative humans perceived themselves and their opponents to be. Scale 0-1: 0 = non cooperative, 0.5 = partially cooperative, 1 = very cooperative

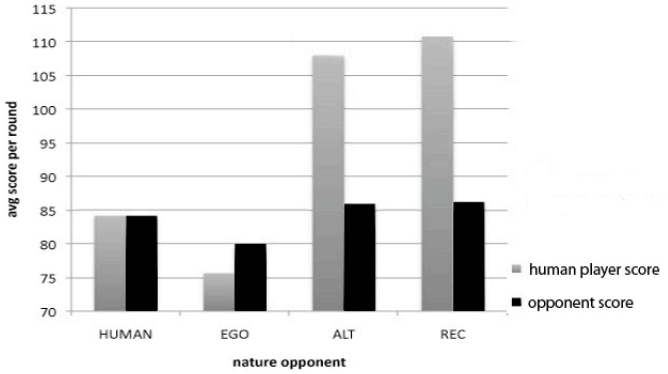

Fig. 3. Average score per round

The data from our survey shows that human participants regard themselves more cooperative towards opponents who they perceive to be cooperative. These results can be found in Fig. 2. The collaboration of human players increases as their opponent shows more cooperative or altruistic behavior. This is most clear when we compare the degree of cooperativeness towards altruistic and egoistic agents, respectively 0.49 and 0.39 on our scale from 0 to 1. Furthermore, players consistently identified themselves as more cooperative than their human and egoistic opponents.

Figure 3 shows the average score of subjects per round, playing against different opponents. The categories on the horizontal axis represent the experimental condition, i.e., the category 'ALT' refers to the experimental condition of the participants playing against an altruistic opponent. The categories show the av-

Table 2. Average score of the human players and their opponents in the different belief conditions

	Player Score	Opponent Score			
		HUM	EGO	ALT	REC
TB	90.9	76.1	74.7	85.4	84.5
FB	95.3	84.4	79.8	88.7	84.3
NB	97.7	92.0	85.4	83.6	89.8

Table 3. The percentage of human players that correctly identified the nature of their opponent.

Nature opponent	% players that identified opponent's nature
HUMAN	63 %
EGO	50 %
ALT	83 %
REC	83 %

erage of all belief conditions (True Belief, False Belief and No Belief). The two main results are the following:

1. The score of the agent increases if it has a more cooperative strategy.
2. Human cooperation with the egoistic agent does pay off, because the average score of both the human player and the agent exceed the minimum score average of 70.

Chipsets were thus distributed that in some rounds, one of the players was able to reach the goal with the initial distribution. Without any cooperation, players were able to obtain an average of 70 points per round. In our experiment, the altruistic and reciprocal agent have average scores of 85,9 and 86,2 respectively. The egoistic agent stays behind with an average score of 80. This also demonstrates that the perceived increase in human cooperativeness is actually reflected in the game.

Interestingly, humans do not fully exploit altruistic players. The average score per round of both ALT and REC are higher than the ones of EGO and the human opponent in the human-human condition. Altruistic agents reach their goal slightly more often (75% of the time) than both human (73% of the time) and egoistic players (69% of the time). The questionnaires reveal that players were prepared to give the altruist chips that would help it reach its goal. The logs confirm this: players transferred on average 4.0 chips per game to EGO and 5.6 chips to ALT.

In the different game conditions players received no, false or true information about the nature of their opponents. Table 2 shows the average score of the players in these conditions. The 'Player Score' denotes the average score of the human players on all rounds. The 'Opponent Score' column shows the average scores of the different types of opponents. Note that the HUM and EGO scores increase as the amount of (true) information about the nature of their opponent decreases. Since human opponents were perceived as rather uncooperative (cf. Fig. 2), this might indicate that egoistic strategies perform better as they have to deal with more uncertainty.

At the end of the game, players were asked to identify their opponent as human or agent. The results can be found in Table 3. Remarkably, EGO was correctly identified only at chance level, i.e., in 50% of the cases the egoistic

agent was mistaken for human. A possible explanation might be that people expect other people to behave selfishly and egoistically in negotiations and as a consequence this behavior is perceived as more human-like [23]. The results of the questionnaire clearly show that the behavior of the altruistic agent is seen as 'stupid', 'dumb' and 'too cooperative to be human'. This corresponds to findings of Kraus and Grosz in [11], who suggest that system designers should build cooperative agents because people expect agents to be cooperative.

6 Discussion

Our experiment provides preliminary insights in negotiation behavior between humans and agents with contradictory interests. Our results suggest that cooperative agents stimulate cooperative human behavior and that humans have certain expectations of the behavior of human and agent opponents. These insights can be applied in more complex and realistic domains, such as e-commerce and auctions. However, the applicability of these finding is limited and is restricted to negotiation settings within a controlled environment. Less regulated interactions, e.g., in social networks, may rely on alternative expectations that require more dynamic, adaptive and complex behavior. We intend to take a step in this direction in future work and use small social networks (teams) to examine cooperative behavior between team players with partly contradictory goals. We also plan to extend our agent strategies to more complex and adaptive ones so as to make them more realistic.

Taking all this into account, our results provide insights into social aspects of human-agent negotiation in specific domains, which can be used to further explore social relations between humans and agents.

7 Conclusion

In this paper we have explored the question of how the behavior of a software agent affects the cooperativeness of its human opponent in a negotiation setting. Initial results show that humans behave more cooperatively towards agents that negotiate with them in a cooperative way. We also find that humans tend to punish egoistic and unfair play by behaving non-cooperatively themselves. Furthermore, egoistic agents are more likely to be identified as humans than altruistic agents.

Our results are not surprising in the wider context of recent work on human behavior analysis and game theory [10, 12]. They confirm the expected results that it is beneficial to act cooperatively in order to induce cooperativeness. However, they are important since previous work has yielded contradictory results which makes it difficult to make assumptions about social conducts between humans and agents. Although our experiment is small in scale, our results provide a foundation for further research in this area.

In order to achieve statistical significance of the results, in the future, we will extend our experiment to a larger number of participants. We plan to extend the agents' behaviors to more complex and adaptive ones.

Acknowledgements. We thank Ya'akov Gal and Bart Kamphorst for helpful comments and assistance with the Colored Trails system. This research is funded by the Netherlands Organization for Scientific Research (NWO), through Veni-grant 639.021.509.

References

1. Jennings, N., Sycara, K., Wooldridge, M.: A roadmap of agent research and development. Autonomous agents and Multi-Agent Systems (1) (1998) 257–306
2. Riedl, M., Saretto, C., Young, R.: Managing interaction between users and agents in a multi-agent storytelling environment. AAMAS (2003)
3. Rajarshi, D., Hanson, J., Kephart, J., Tesauro, G.: Agent-human interactions in the continuous double auction. IJCAI (2001)
4. Gal, Y., Pfeffer, A., Marzo, F., Grosz, B.: Learning social preferences in games. AAAI (2004)
5. Sanfey, A., Rilling, J., Aronson, J., Nystrom, L., Cohen, J.: The neural basis of economic decision-making in the ultimatum game. Science (300) (2003) 1755–1758
6. Sierhuis, M., Bradshaw, J., Acquisti, A., van Hoof, R., Jeffers, R., Uszok, A.: Human-agent teamwork and adjustable autonomy in practice. i-SAIRAS (2003)
7. van Wissen, A., van Diggelen, J., Dignum, V.: The effects of cooperative agent behavior on human cooperativeness (short paper). AAMAS (2009 (to appear))
8. Camerer, C.: Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press (2003)
9. Axelrod, R.: The Evolution of Cooperation. Basic Books (1984)
10. Blount, S.: When social outcomes aren't fair. Organizational Behavior and Human Decision Processes **63**(2) (1995) 131–144
11. Grosz, B., Kraus, S., Talman, S., Stossel, B., Havlin, M.: The influence of social dependencies on decision-making: Initial investigations with a new game. AAMAS (2004)
12. Guth, W., Schmittberger, R., Schwarz, B.: An experimental analysis of ultimatum bargaining. Journal of Economic Behavior and Organization (3) (1982) 367–388
13. Loewenstein, G., Thompson, L., Bazerman, M.: Social utility and decision making in interpersonal contexts. Journal of Personality and Social Psychology (1989)
14. Ross, M., Fletcher, G.: Attribution and social perception. In Lindzey, G., Aronson, E., eds.: Handbook of Social Psychology. Random House (1985)
15. Kagel, J., Roth, A.: The Handbook of Experimental Economics. Princeton University Press (1995)
16. Weiss, G., Sen, S.: Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence. MIT Press (1999)
17. de Jong, S., Tuyls, K., Verbeeck, K.: Fairness in multi-agent systems. The Knowledge Engineering Review (2008)
18. Danielson, P.: Competition among cooperators: Altruism and reciprocity. In: PNAS. (May 2002)
19. Hogg, L., Jennings, N.: Socially intelligent reasoning for autonomous agents. IEEE Trans on Systems, Man and Cybernatics - Part A (2001) 381–399
20. Gal, Y., Pfeffer, A.: Modeling reciprocity in human bilateral negotiation. AAAI-07 (2007)
21. Katz, L., ed.: Evolutionary Origins of Morality: Cross-Disciplinary Perspectives. Imprint Academic. (2000)
22. Binmore, K.: Fun and Games: a Text on Game Theory. D.C. Heath and Company (1992)
23. Epley, N., Caruso, E.M., Bazerman, M.H.: When Perspective Taking Increases Taking: Reactive Egoism in Social Interaction. SSRN eLibrary (2005)
24. Nassiri-Mofakham, F., Ghasem-Aghaee, N., Nematbakhsh, M., Baraani-Dastjerdi, A.: A personality-based simulation of bargaining in e-commerce. Simulation Gaming **39**(1) (2008)