The Effects of Cooperative Agent Behavior on Human Cooperativeness

(Extended Abstract)

Arlette van Wissen
Institute of Information and Computing Sciences
Utrecht University
P.O. Box 80.089
arlette.vanwissen@phil.uu.nl

Jurriaan van Diggelen
Institute of Information and Computing Sciences
Utrecht University
P.O. Box 80.089
jurriaan@cs.uu.nl

Virginia Dignum
Institute of Information and Computing Sciences
Utrecht University
P.O. Box 80.089
virginia@cs.uu.nl

ABSTRACT

In this paper we examine the question of how cooperativeness of a software agent affects cooperativeness of a human player. Our data shows that humans behave more cooperatively towards agents that negotiate with them in a cooperative way.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems—*Human factors*; I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent agents*

General Terms

Human factors

Keywords

human-agent interaction, social factors, coopeative behavior

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. Many forms of human-machine interaction involve participants that pursue different, or even contradictory, interests. To achieve successful human-machine interactions, we must understand the factors involved in this interaction. 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.

This work presents an experimental study of the social phenomena that play a role in human-agent interactions with contradictory interests. More specifically, we explore to what extent social principles like reciprocity and fairness influence human bidding behavior in these interactions. We develop agent behavior models that exhibit different cooperativeness strategies, differing from pure egoism to pure altruism. The central question we address is: how does the cooperativeness of an agent affect the cooperativeness of a human player? This study may provide new insights into how bidding decisions are made and how agents' behaviors are perceived. A complete description of this study can be found in [2].

Cite as: The Effects of Cooperative Agent Behavior on Human Cooperativeness, (Extended Abstract), Arlette van Wissen, Jurriaan van Diggelen, Virginia Dignum, *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009),* Decker, Sichman, Sierra and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. 1179–1180

2. EXPERIMENTAL APPROACH

To analyze differences in behavior between various kinds of human-agent interactions we use Colored Trails (CT) [1], which is a negotiation environment for designing and evaluating decision-making behavior and dynamics between players. We create a CT setup that enables multiple humans to play the game simultaneously against alternate opponents.

In an Ultimatum Game, it would be rational for a responder to always accept if the proposal leaves him with a payoff of 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. This can be interpreted as human willingness to play fair and to punish 'unfair' splits.

Our 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. We use *one-shot negotiation*: if the responder rejects the proposal, both players receive nothing. If the player has the necessary chips to reach the goal, the program will move its piece to the goal via the shortest possible path. In case the player does not have the required chips, its piece will not move at all.

The scoring function is defined as follows: $s = goal + board - steps$, where *goal* represents the rewarded points for reaching the goal (100), *board* represents the size of the board (20) and *steps* is the number of steps of the taken path (variable per round). 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, the requested chips, the response and the pursued path.

We developed three agents with different strategies (cf., Table 1). Our first hypothesis states that 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. Our second hypothesis is that non-cooperative behavior instigates 'punishment-behavior' of humans. We expect human players to prevent the egoistic agent from reaching its goal. 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.

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

3. EXPERIMENTAL DESIGN

The CT games were played with *full board visibility* and *full chip visibility*. Participants played on a 4×5 board with one square designated as a goal. A total of 30 subjects participated in the experiment. All participants were students between 20-30 years of age, almost 50% of them were female. Participants were instructed to perform a non-competitive task: they were to try to maximize their own scores, not to minimize other players' scores. Participants played three games against the agent opponents and one game against another human participant. The group that played the human condition 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. Therefore we created three different conditions in which we provided them with true, false or no information about the nature of their opponent.

After playing the game, participants filled out a questionnaire.The questionnaire provided us with insights of how different aspects of cooperation come about. The logs of each CT game contained information of player behavior.

4. RESULTS

Our results support both our hypotheses. However, more extensive research with a larger number of participants has to be done to significantly demonstrate these results.

Human participants regard themselves more cooperative towards opponents who they perceive to be cooperative (Table 2). The questionnaires reveal that the collaboration of human players increases as their opponent shows more cooperative 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 0 to 1 scale. Furthermore, players consistently identified themselves as more cooperative than their human and egoistic opponents.

Table 3 shows the average score of subjects per round, playing against different opponents. There are two interesting main results. First, the score of the agent increases if it has a more cooperative strategy. Second, human cooperation with the egoistic agent does pay off, since the average score of both the human player and the agent transcend the minimum score average of 70. 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. The questionnaires reveal that players were prepared to give the altruist chips that would help it reach its goal.

Nature opponent	coop. player	coop. opponent
HUMAN	0.57	0.33
EGO	0.38	0.22
ALT	0.48	0.92
REC	0.4	0.98

Table 2: Perception of cooperation. Scale 0-1: 0 = non cooperative, 1 = utmost cooperative

Nature opponent	human player score	opponent score
HUMAN	84	84
EGO	76	80
ALT	108	86
REC	111	86

Table 3: Average score per round

At the end of the game, players were asked to identify their opponent as human or agent. Remarkably, EGO was correctly identified only at chance level, i.e., in 50% of the cases the egoistic agent was mistaken for human. ALT and REC on the other hand, were identified correctly as an agent 83% of the time.

5. CONCLUSION & DISCUSSION

Our experiment provides preliminary insights in negotiation behavior between humans and agents with contradictory interests. 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 restricted to negotiation settings within a quite controlled environment. In future work, we intend 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.

Acknowledgements.

We thank Ya'akov Gal and Bart Kamphorst for helpful comments and assistance with the CT system. This research is funded by NWO, Veni-grant 639.021.509.

6. REFERENCES

[1] B. Grosz, S. Kraus, S. Talman, B. Stossel, and M. Havlin. The influence of social dependencies on decision-making *AAMAS*, 2004.

[2] A. van Wissen, J. van Diggelen, and V. Dignum. The effects of cooperative agent behavior on human cooperativeness. *BNAIC*, 2008.