

# Human-Agent Teamwork in Dynamic Environments

A. van Wissen<sup>a,\*</sup>, Y. Gal<sup>b,c</sup>, B. A. Kamphorst<sup>d</sup>, M. V. Dignum<sup>e</sup>

<sup>a</sup>*VU University Amsterdam, Dept. of Computer Science, De Boelelaan 1081a, the Netherlands*

<sup>b</sup>*Information Systems Engineering Department, Ben-Gurion University of the Negev, Israel*

<sup>c</sup>*School of Engineering and Applied Sciences, Harvard University, USA*

<sup>d</sup>*Utrecht University, Dept. of Philosophy, the Netherlands*

<sup>e</sup>*Delft University of Technology, Faculty of Technology, Policy and Management, Jaffalaan 5 2628 BX Delft*

---

## Abstract

Teamwork between humans and computer agents has become increasingly prevalent. This paper presents a behavioral study of fairness and trust in a heterogeneous setting comprising both computer agents and human participants. It investigates people's choice of teammates and their commitment to their teams in a dynamic environment in which actions occur at a fast pace and decisions are made within tightly constrained time frames, under conditions of uncertainty and partial information. In this setting, participants could form teams by negotiating over the division of a reward for the successful completion of a group task. Participants could also choose to defect from their existing teams in order to join or create other teams. Results show that when people form teams, they offer significantly less reward to agents than they offer to people. The most significant factor affecting people's decisions whether to defect from their existing teams is the extent to which they had successful previous interactions with other team members. Also, there is no significant difference in people's rate of defection from agent-led teams as compared to their defection from human-led teams. These results are significant for agent designers and behavioral researchers who study human-agent interactions.

*Keywords:* human computer interaction, cooperative behavior, teamwork, dynamic environments

---

## 1. Introduction

A significant portion of human activity involves people working in groups. Examples include working on school projects, playing team sports, providing professional health care and executing disaster-relief missions. Increasingly,

---

\*Corresponding author. Tel. +31 (0)20 598 5887.

*Email addresses:* [wissen@few.vu.nl](mailto:wissen@few.vu.nl) (A. van Wissen), [kobig@bgu.ac.il](mailto:kobig@bgu.ac.il) (Y. Gal), [Bart.Kamphorst@phil.uu.nl](mailto:Bart.Kamphorst@phil.uu.nl) (B. A. Kamphorst), [M.V.Dignum@tudelft.nl](mailto:M.V.Dignum@tudelft.nl) (M. V. Dignum)

these groups include computer agents (hereafter referred to as “agents”) as many new technologies are being developed that enable agents to interact with people in a mature and meaningful way. In particular, a lot of advances have been made in assistive technology (Pollack, 2005) and negotiating techniques (Lin, Oshrat & Kraus, 2009b).

Agents can play a variety of roles in their interactions with humans: they can support people in their tasks, for instance in collaborative interfaces and cognitive care systems (Shieber, 1996; Babaian, Grosz & Shieber, 2002; Yorke-Smith, Saadati, Myers & Morley, 2009); serve as proxies for individual people or organizations, for example as bidders in online auctions (Kamar, Horvitz & Meek, 2008; Rajarshi, Hanson, Kephart & Tesauro, 2001); or work autonomously to carry out tasks for which they are responsible, in settings such as computer games and simulation systems for natural disaster relief (Schurr, Patil, Pighin & Tambe, 2006; Murphy, 2004).

Human behavior in group settings has been shown to depend on sociological and psychological aspects, such as trust and fairness (Fehr & Schmidt, 1999; Jones & George, 1998). For agents to be successful in these team interactions, they need to act appropriately, meeting people’s expectations of how other group members will behave. In turn, as agents increasingly become equal partners in their interactions with humans, it is important to understand how the decisions that agents make affect human behavior.

This paper presents an empirical study of the decision-making strategies that people deploy when agents are among the members of the groups in which they participate. It investigates the extent to which people treat agents differently than other people when creating teams to carry out joint tasks. In particular, it studies the effects of trust and fairness concerns in dynamic environments in which people have to manage their (possibly conflicting) commitments. These dynamic environments are characterized by rapidly changing circumstances that participants are expected, and often required, to deal with, and by the amount of information that participants have to consider when they make decisions. Examples of applications for heterogeneous human-agent teams in dynamic environments include autonomous agent pilots that work together with people to complete a joint mission (Tambe, Johnson, Jones, Koss, Laird, Rosenbloom & Schwamb, 1995; Tambe, 1998), or mission execution assistants for astronauts to help them empower their cognitive capacities (van Diggelen, Bradshaw, Grant, Johnson & Neerincx, 2009).

Most research on teamwork focused on interactions solely between people or between agents (cf. Riedl & Vyrastekova (2003); van Beest (2002) and van de Vijzel & Anderson (2004); Sen (1996)). Work that addressed heterogeneous teams of people and agents was often concerned with GUI design or with optimization of agent strategies to outperform people (cf. Shneiderman (1987); Goschnick (2009) and Lin, Kraus, Tykhonov, Hindriks & Jonker (2009a), respectively). Very little literature has been found that addresses the social dynamics that come into play when people interact with agents (e.g., Nass, Fogg & Moon (1996); Blount (1995)). This work contributes to our understanding of these dynamics.

We constructed a dynamic environment for this study, in which, in order to complete group tasks, participants were required to form teams by selecting teammates with appropriate skills and resources. Team leaders negotiated with potential members about the allocation of rewards resulting from the successful completion of a group task. Participants could also choose to carry out individual tasks that did not benefit other team members, to make outside offers to participants in other teams, and to defect from their existing team. Our central hypothesis was that trust and fairness are key factors that affect whether people treat agents differently than humans in this setting.

The results show that when negotiating to create teams, people offer less reward to agents than they do to other people. However, no significant difference was shown in people’s willingness to join human- versus agent-led teams. In addition, people do not defect more often from agent-led teams than they do from human-led teams. Lastly, the main factor contributing to the acceptance of outside offers by team members was a positive relationship with the human or agent making the outside offer, rather than the reward associated with it.

These results highlight the roles played by social factors in people’s decision-making in heterogeneous human-agent groups. They suggest that agent designers should include factors such as trust and fairness in agents’ decision-making models in order to meet people’s expectations of others when they act as team members.

The rest of this paper is organized as follows. Section 2 presents relevant theories and research in the field of human-agent team formation along with a description of related work. In section 3 we present our experimental framework. Section 4 describes our empirical methodology and covers details of the experimental setup. Section 5 presents the experiment results, showing how trust, fairness and the type of actors (human or agent), influence the way that humans interact with agents in our setting. Section 6 provides a discussion of the results in light of our hypotheses. Finally, conclusions of the study and ideas for future work are presented in section 7.

## **2. Related Work**

Our study relates to work that spans the social sciences as well as Artificial Intelligence. Past studies of negotiation have shown that fairness and trust considerations greatly influence people’s behavior (Joseph & Willis, 1963; Pruitt & Carnevale, 1993), and that previous experiences affect trust relationships between group members (Hinckley, 1981). Specifically, people have been shown to reject beneficial offers that are perceived as unfair in two- and three-player team formation games (Okada & Riedl, 2005; van Beest, van Dijk & Wilke, 2004), and that people are likely to choose potential teammates with whom they have developed strong working relationships in the past (Hinds, Carley, Krackhardt & Wholey, 2000). Self-interest and equity concerns have also been shown to affect people’s negotiation behavior when forming teams (van Beest, Andeweg, Koning & van Lange, 2008). These studies were limited to interactions solely comprising people.

One of the pioneering studies to investigate the social factors that come into play in interactions between humans and computers, carried out by Nass et al., shows the conditions under which human behavior towards media is the same as towards other humans (Nass, Fogg & Moon, 1996; Reeves & Nass, 2003). According to Nass' *media equation*, the social dynamics surrounding human-computer interactions mirror those interactions in groups that solely comprise humans. More recent studies have also shown social and psychological dynamics to occur between people and agents. The results in these experiments were not conclusive, however. For example, people were shown to accept lower offers from computer proposers than from human proposers in the ultimatum game (Blount, 1995). Yet in other contexts, where repeated negotiation took place, people exhibited reciprocal behavior towards agents in a manner that was similar to their interaction with other humans (van Wissen, van Diggelen & Dignum, 2009). Our work augments these studies by describing the factors that bring people to negotiate differently with agents than with people in a more realistic setting.

Past studies on team formation were limited to environments in which agreements were binding (van de Vissel & Anderson, 2004; Okada & Riedl, 2005), the division of the rewards was fixed (Breban & Vassileva, 2002; Brooks & Durfee, 2002) and where team membership was pre-defined (Sen, 1996; Breban & Vassileva, 2002). Most studies on team formation did not consider the human role (cf. van de Vissel & Anderson (2004); Breban & Vassileva (2002)) or did not include agents as autonomous team members (cf. van Beest, van Dijk & Wilke (2004); Bolton & Brosig (2007); Clancey, Sachs, Sierhuis & van Hoof (1998); Bradshaw, Acquisti, Allen, Breedy, Bunch, Chambers, Galescu, Goodrich, Jeffers, Johnson, Jung, Kulkarni, Lott, Olsen, Sierhuis, Suri, Taysom, Tonti, Uszok & Hoof (2004)).

Furthermore, work on human-agent interaction has focused primarily on the use of agents as support systems for humans. Some agent systems support people with individual tasks (e.g. personal assistant agents in the realm of task management (Yorke-Smith et al., 2009) or in electronic commerce applications (Kamar et al., 2008; Rajarshi et al., 2001)), others support people who have to work in teams (e.g. multiagent teams that train incident commanders for large scale disasters (Schurr et al., 2006), or intelligent agents that support bidding decisions and negotiations (Pommeranz, Brinkman, Wiggers, Broekens & Jonker, 2009).) Over the past years however, we have witnessed a shift from these types of settings to another: one in which agents act autonomously alongside people as peers. Examples include human-agent teams performing experiments and collecting data in space (Bradshaw et al., 2004) or robots in rescue operations (Murphy, 2004).

Recent work in AI focuses on decision-making models that learn and reason about the role of social factors in agents' decision-making mechanisms rather than use optimal strategies to guide their play. However, computational work that models the way humans make decisions is limited to bilateral negotiation settings that involve repeated interaction in markets (Oshrat, Lin & Kraus, 2009), one-shot negotiation settings games (Gal, Pfeffer, Marzo & Grosz, 2004)

or electronic auctions (Rajarshi et al., 2001). Our work is related to classical work on social identification that show that people are likely to favour those with which they share solidarity or identification (Tajfel & Turner, 1979; Bernstein & Ben-Yossef, 1994), but extends these studies to dynamic environments in which the group structure is not pre-determined, but forms as a process of negotiation.

### 3. Domain Description

The domain used for this study is an extension of the Package Delivery Domain (PDD), proposed to study the effects of reciprocity between agents in team formation (Sen, 1996; van de Vijzel & Anderson, 2004). We have adapted this domain to include multiple participants of varying capabilities who need to accomplish tasks by delivering small and large packages. Small packages may be delivered by individuals without the assistance of others; large packages can only be delivered by a team of participants with appropriate capabilities. The reward awarded to individuals for successfully delivering small packages is constant (3 points). The reward awarded to teams for successfully delivering large packages increases proportionally to the number of members in the team (60 points for two-member teams; 180 points for three-member teams) and is significantly larger than the reward for delivering small packages.<sup>1</sup> This reward is contingent upon the team successfully delivering its package.

Teams are formed by a negotiation process. Participants may initiate a team by inviting others to take part in a collaborative effort to deliver a large package. To do so, they have to individually offer each invitee some of the reward that is associated with the successful delivery of the package. The participant who sends out these invitations and makes the proposals is called the *team initiator*. The *invitee* is required to either accept or reject the offer. This is a forced choice, in that the invitee cannot perform any action until it has responded to the offer. If the invitee accepts the offer, the invitee joins the initiator’s team. The team initiator has to invite each team member separately. When an invitee declines an invitation to join a team, the team initiator will be informed of this decision, and is free to make another offer to the invitee, or to a different participant. Note that invitees may already be team members of another existing team.

Agreements between team members and the team initiator are non-binding. As long as the team has not completed its task, team members can choose to defect from their team and render it unsuccessful in delivering its package. Defection occurs if any of the team members accept an outside offers from another team initiator. The team initiator can also decide to dissolve the team at will at any point, regardless of any outside offers. Participants do not pay a direct cost for defecting from their team.

The delivery of small, individual packages can be performed by participants at any stage of the interaction, whether or not they are part of a team. How-

---

<sup>1</sup>In particular, the relative reward per team member in three-member teams (60) is twice as large as the relative reward per team member in two-member teams.

ever, the delivery of small packages by team members will delay the successful completion of their group task. The interaction terminates after a randomly allocated number of minutes has passed. At this point the performance in the game is determined by the number of points aggregated by each of the participants. Table 1 summarizes the key elements of the package delivery domain.

**Insert table 1 here**

This domain is well suited to study group decision-making in dynamic environments for the following reasons. First, it is highly interactive in that participants have to reason simultaneously about different types of decisions. For example, a team member can receive an outside offer while it is still working with another team. Second, it is characterized by uncertainty at two levels: there is uncertainty about the environment (e.g., the duration of the interaction, the way packages will be distributed), and about other participant’s decisions (e.g., how others decide whether to defect from their commitments, and how much they offer others to join their team). Lastly, it includes a social dilemma, requiring participants to reason about the trade-offs between acting selfishly and incurring a short term benefit (e.g., defecting from a team by accepting a more lucrative offer) and acting cooperatively for the long term gain (e.g., remaining committed to the team and establishing a positive trust relationship).

#### **4. Empirical Methodology**

As discussed in section 1, we expect that trust and fairness will be the main factors underlying the choice of team members that people make. They will tend to choose those with whom they have successfully interacted in the past and will try to maintain their trust relationships by offering close to fair splits of the rewards. These expectations of people’s behavior are sustained by several studies on human collaboration (Hinckley, 1981; Hinds et al., 2000; Roth, 1995; Rubinstein, 1982). However, we suspect that when interacting with agents, people will develop trust relationships differently (van Wissen et al., 2009; Blount, 1995). In fact, we expect that because people will not attribute any emotional feelings to the agents, they will therefore feel less ‘guilt’ towards agents, resulting in more defection from agent-led teams. For the same reason, we expect people to offer agents less than fair splits. Because we designed the rewards in the game to encourage participants to form teams, we expect them to prefer team formation over individual activity.

On the basis of these assumptions, we generated the following main hypothesis for this research:

Trust and fairness are key factors that affect whether people treat agents differently than humans when interacting in dynamic environments.

In order to perform a more detailed analysis of the main hypothesis, we investigated eight more fine-grained hypotheses:

**Hypothesis 1.** *Participants are more likely to form or join teams than carry out actions individually.*

**Hypothesis 2.** *Larger teams are more likely to fail than smaller teams.*

**Hypothesis 3.** *Human team initiators offer agents less benefit to join their teams than they offer to humans.*

**Hypothesis 4.** *Human team members are more likely to defect from agent-led teams than from human-led teams.*

**Hypothesis 5.** *Team members defect because they receive outside offers that are significantly higher than the allocated reward of their current team.*

**Hypothesis 6.** *Team members are more likely to defect when the received outside offer originates from a participant with whom they have worked with successfully in the past.*

**Hypothesis 7.** *Team initiators prefer to invite those with whom they have established positive relationships in the past.*

**Hypothesis 8.** *Participants who are more likely to defect suffer in performance, because former team members are less likely (i) to invite them to join teams, or (ii) to agree to join teams initiated by defectors.*

In section 6 we will relate the outcomes of the study to these hypotheses. The remainder of this section is dedicated to a description of the domain used in this study.

#### *4.1. Implementation using Colored Trails*

We implemented the Package Delivery Domain using the Colored Trails (CT) framework (Grosz, Kraus, Talman, Stossel & Havlin, 2004; Gal, Grosz, Kraus, Pfeffer & Shieber, 2010). CT is a testbed for empirical studies of decision making that supports the evaluation of negotiation strategies of heterogeneous groups of actors comprised of both humans and agents. The CT PDD implementation is a conceptually simple (but highly versatile) game in which players negotiate and exchange resources to achieve their individual or group goals. CT provides a realistic analogue of the ways in which goals, tasks and resources interact in real-world settings, but abstracts away the complexities of real-world domains. CT has been used to study diverse topics relating to human-agent decision-making, such as the role of gender and social relationships in people’s negotiation behavior (Katz, Amichai-Hamburger, Manisterski & Kraus, 2008), evaluating computational models of human reciprocity (Gal & Pfeffer, 2007), the way people respond to interruptions from computers (Kamar, Gal & Grosz, 2009) and the effect of space-travel conditions on people’s negotiation behavior (Hennes, Tuyls, Neerincx & Rauterberg, 2009). Past studies have shown that CT has a measured effect on human behavior, in that it leads people to exhibit more helpful behavior (and increase their social welfare) than when using traditional payoff matrices in which decision-making is reduced to choosing from a list of possible outcomes. This makes CT the right kind of setting to investigate team formation in dynamic environments that are akin to the real world.

#### 4.1.1. The Board

The game is played by six participants on a  $11 \times 11$  board of colored squares. One central square on the board is designated as the depot. Each participant is represented by an icon on the board and each icon is initially located at a random position on the board (the depot location and package locations excluded). Each icon is assigned a color from the same palette as the squares on the board. A snapshot of one of the boards used in our study is shown in Figure 1a. We used a palette of three possible colors for each square of the board: red, green and blue. Each participant in the game is identified by an icon of one of these three colors.

Figure 1a also shows three of the six participants in the game: a blue ‘me’ icon, representing the location of the participant viewing this board; a red agent icon, located near the bottom-right corner of the board, and a green human icon, located one square to the right and two squares up from the goal depot. The current total reward of the ‘me’ participant is displayed at the bottom of the main board panel.

Participants can move freely on the board, but not diagonally. Also shown in the Figure are the small and large packages dispersed over the board. Small packages appear in white and large packages appear in brown. At any given point there are twelve small packages and six large packages located at random positions on the board. Once a package is delivered — whether by an individual participant or by a team — another package of the same type is generated and positioned at a random location on the board.

#### 4.1.2. Forming Teams and Delivering Packages

Participants collect rewards by delivering packages to the goal depot. The delivery of a small package requires the participant to move across the board and place its icon on the designated package. That package is then picked up automatically. The participant, now visibly carrying the package, must subsequently move towards the goal depot to deliver the package.

In order to deliver large packages, participants have to collaborate as a team. Participants can form teams in one of the following ways:

- **As a team initiator:** by inviting one or two other participants by sending them a proposal.
- **As a team member:** by accepting a proposal that was sent by an initiator.

Teams must meet the following criteria: (i) team members cannot share the same color icon, and (ii) all of the colors on the squares in the path from the position of the large package to the goal depot should be represented by the icons of the team members. For example, in Figure 1a, one of the paths to the depot from the location of the large package and the ‘me’ icon consists of moving five steps to the right and two steps up. This path is outlined in the figure. The squares in this path include three colors, and thus requires a three-member team. However, one path to the depot (six steps to the right, two steps

up, and one step to the left) requires only two colors, green and blue. This path could be realized by a two-member team represented by blue and green icons.

The CT configuration described above was designed to provide an analogy to aspects that characterize team interactions, highlighting the interaction among goals, tasks required to achieve these goals, and resources needed for completing tasks. The colors of players' icons represent capabilities and skills required to fulfill tasks. When participants form teams, they pull together their resources and capabilities towards achieving their joint tasks. This is analogous to the way teams are formed in the real world.

The protocol for forming teams requires that participants reach agreements with one another. To invite others to join, the team initiator must position its icon on a square with a large package. For example, the 'me' icon in Figure 1a is positioned on one of the large packages. Next, the team initiator makes an offer to each of the participants that it wishes to invite to join the team. This offer concerns a split of the reward associated with the successful delivery of the package. An example of such an offer is presented in Figure 1b. Here, a 'me' participant has been invited to form a two-member team by the team initiator (a green icon that represents a person). Should the participant accept this offer, then that participant becomes part of the team. If the team succeed in delivering its package, both the 'me' player and the team initiator will receive 30 points. The Message History Panel, shown in Figure 1d, displays all past offers made or received by the 'me' player and their associated replies. Note that a team initiator can make multiple offers to the same participant.

### **Insert Figure 1 here**

To actually deliver the package, all team members have to position their icons on the large package. This is the same square in which the team initiator is located. Once all team members have united on the location of the designated large package, the team automatically delivers the package. This completes their joint task and the associated rewards are distributed to the team members according to the agreed upon splits. We note that team initiators must reach separate agreements with each of the members in the team, and team members cannot observe agreements of other team members. However, team members may continue to make offers or to receive offers from non-team members.

The Coalition Panel, shown in Figure 1c, shows the structure and performance of past teams in which the 'me' player participated. It shows all previous interactions with team members within one round. For example, in Figure 1c the team comprising the 'me' player and the green human was initiated by the human, and did not succeed, while the team comprising the 'me' player and the green agent was initiated by the 'me' player, and did succeed.

The setting was designed to bring out social dilemmas that characterize groups: team initiators are required to choose between a human and an agent sharing the same color when joining teams (e.g. for a participant who wants to work with someone who has the capabilities that are represented by the color blue, there is always a choice between working with a blue human or a blue

agent); team members may need to decide whether to defect from an existing team in light of an outside offer from another team; participants must choose between delivering small packages and incurring (low) rewards with certainty, and forming teams that are potentially more lucrative, but carry a higher risk of failure. This configuration promotes the development of trust relationships between participants and enables us to study whether humans prefer to work in teams with agents or humans.

#### *4.1.3. Defection*

Recall that any member of a team can choose to dissolve the team by defecting. All members can defect by accepting outside offers by non-team members, the initiator can also defect at will by moving off the square containing the large package associated with the team. There is no score penalty associated with the decision to defect. When a team is dissolved, it fails to deliver its package and none of its members receive any reward.

#### *4.1.4. Scoring*

The scoring function for the game was set as follows: Each small package that is delivered to the goal depot yields a reward of 3 points to the participant delivering it. The delivery of a large package by a two or three-member team yields a reward of 60 or 180 points, respectively. This reward is divided among all team members, depending on the agreements reached between the team initiator and each of the team members. These values were determined after several pilot studies in such a way to incentivize participants to form teams, despite the increase in waiting time and risk of defection that comes with creating teams. The score for each participant is aggregated over time.

#### *4.1.5. Availability of information*

To prevent overloading participants with information we imposed the following constraints. First, the exact duration of each game was not revealed to participants, so not to incentivize strategic reasoning about future decisions. Second, participants could not see the location of the icons of other participants on the board, except when they were on the same team. Third, participants could not see each other's performance during the interaction, to avoid their decision strategies being influenced by performance comparison with other participants.

The inherent uncertainty in the game makes it difficult for players to engage in strategic (game theoretic) reasoning about the best action to take. Therefore, we can attribute the decisions made by participants to the type of social factors that we wish to study.

### *4.2. Agent Reasoning*

The design of agents can strongly influence the outcomes of a game. For instance, a badly designed agent could be excluded from team participation for exhibiting random or inconsistent behavior. On the other hand, an agent

that behaves predictably according to a strategy, could be singled out because of that strategy. In order to decrease these effects, we replaced agents with people throughout the study. That is, we used a ‘Wizard of Oz’ (WOZ) type of experimental setting, which is common in studies of human-agent interaction (Kelley, 1984). WOZ testing is an experimental user interface evaluation method in which the user of the system is made to believe that he or she is interacting with a fully implemented system, though in fact the whole system or a part of it is controlled by one or several humans (“wizards”). WOZ tests have proven useful in evaluating interfaces and supporting the design process of agents (Wooffitt, 1997; N. Dahlbäck, 1993).

In our setup, all participants in a game were in fact human, but appeared to be agents to some of the other participants. Importantly, this fact was not revealed to the subjects during the study.<sup>2</sup> We varied the identity of participants that were presented as agents at each round of the study. This was done in a semi-random way with the constraints that all participants appeared as human to themselves and that there were as many agents as humans in the game. In any given round it was possible for a participant to appear as a human to one participant, and as an agent to another participant. Yet the true identity of the players was never revealed. Humans and agents are represented by icons, examples of which can be found in Figure 1.

There were two major benefits to this deception. First, it isolated the effects concerning the type of subject (whether human or agent) on people’s behavior. Had we used actual agents it would have been impossible to distinguish those effects relating to the type of participant from those relating to the strategies used by the agents. By having humans play the role of agents, we could attribute differences in behavior solely to the way people relate to different participant types. Moreover, since the players appeared differently to all players it was ensured that the choice to work with a human or an agent would not be guided by the strategy employed by that player.

The second benefit is that this deception allows us to use the data we collected in this experiment as a baseline for evaluating decision-making strategies of yet to be developed agents in future studies.

### *4.3. Experimental Setup*

Our experiment generated a total of 90 games, played by 18 subjects from diverse socio-economic backgrounds. Although the number of subjects is small, the number of games played provide a statistically significant amount of data. 44% of the subjects were male. 50% of the subjects were younger than 25, 44% were between 25-29 years old and 6% were between 30-34 years old. The majority (72%) of the subjects were students. Subjects went through an identical 45-minute tutorial of the game that consisted of a handout, a video, and a trial

---

<sup>2</sup>The deception was approved by the Institutional Review Board of Harvard University, the academic institution sponsoring the study.

round.<sup>3</sup>

Participation in the study was contingent on passing a comprehension quiz about the game. During the tutorial we used neutral terminology to avoid introducing competitive or cooperative factors to players' reasoning other than their own. For example, we used 'participant' instead of 'opponent' and 'interaction' instead of 'game'. Subjects were not allowed to communicate and could not see each others' console.

Each subject participated in a series of rounds in which the board configuration as well the identity of the other participants were varied. The duration of each round was set at random between 4 and 10 minutes. After each round, players were randomly designated a color. Subjects were monetarily compensated according to their total performance in the rounds they played. At the end of the session subjects were asked to fill out a questionnaire, asking them to explain about their choices in certain situations in the interactions.<sup>4</sup>

## 5. Results

Analysis showed that participants varied significantly in their total rewards. The mean total reward for participants was 565 points, with a standard deviation of 300 points.

**Result 1.** *The successful delivery of large packages was a stronger predictor of performance than the delivery of small packages.*

Although both group-type actions (delivering large packages) and individual-type actions (delivering small packages) were predictors of performance in the game, the delivery of large packages was a more significant predictor of performance than was the delivery of small packages (linear regression  $F(2, 69) = 105.71, r^2 = 0.7539, p < 0.0001$ ). In addition, participants attempted to form three-member teams more often than they attempted to form two-member teams (295 times versus 202 times, respectively). This shows that participants were willing to take the extra risks associated with creating larger teams.

### 5.1. Team Formation

We distinguish between several events relating to the way teams were created, maintained and dissolved in the game. We consider a team to be *formed* once all of its intended members have successfully accepted proposals from the team initiator; a team is considered *successful* if it delivers its large package to the depot, and a team has *failed* if at least one of the team members or the

---

<sup>3</sup>The tutorial can be found at <http://www.phil.uu.nl/~wissen/downloads/cihb/tutorial.pdf>. The video is available at <http://www.phil.uu.nl/~wissen/downloads/cihb/pddmovie.zip> and provides an impression of the flow of the game.

<sup>4</sup>The questionnaire is available at <http://www.phil.uu.nl/~wissen/downloads/cihb/questionnaire.pdf>.

initiator defected before the team was able to deliver its package.

**Insert Table 2 here**

Table 2 lists the number of two- and three-member teams formed by participants, and the number of times that those teams were successful.

**Result 2.** *Significantly more three-member teams were formed than two-member teams (goodness of fit test  $p < 0.0001$ ).*

Table 2 shows that there were 205 formed three-member teams and only 122 formed two-member teams. This result is striking given that there were more risks involved in the creation of three-member teams versus two-member teams. Larger teams consisted of more team members that could choose to defect on their respective agreements. Another source of risk was time. The lifespan of a successful team is measured as the time that elapsed from the first offer made by the team initiator to the first team member, until the team has successfully delivered the package. For teams to be successful, the team initiator and each of the team members had to agree on a split of the reward, and the team members must not accept outside offers as long as the package has not been delivered. The average lifespan of three-member teams (three minutes) was significantly longer than the life-span of two-member teams (one minute). The additional time can be linked to the fact that the team initiator had to negotiate with each potential team member separately, providing more opportunities for outside offers and possible defections by team members.

*5.2. The Effect of Team Size and Fairness*

According to the data, two-member teams were more robust than three-member teams. Table 2 shows that 68% of formed two-member teams were successful, while only 56% of formed three-member teams were successful ( $\chi^2(N = 1, 304) = 11.8, p = 0.0001$ ).

**Result 3.** *The defection rate in three-member teams was higher than the defection rate of two-member teams.*

However, the average likelihood of defection for individual participants was similar for two- and three-member teams (the average likelihood of defection in three-member teams was 16%, whereas it was 15% for two-member teams).

To compare between offers made by team initiators when forming two- and three-member teams, we defined a measure of fairness in our scenario that depended on the relative size of the split offered by team initiators to their respective team members. For example, a 100% fair offer to a participant in a two-member team (with a total reward of 60 points) was 30 points, and a 50% fair offer was 15 points. Similarly, a 100% fair offer to a participants in a three-member team (with a total reward of 180 points) was 60 points, and a 50% fair offer was 30 points.

**Result 4.** *Offers to form two-player teams were significantly fairer than offers to form three-player teams.*

As shown in Table 3, offers made by team initiators to potential members of two-member teams were fairer (99%) than offers made by team initiators to potential members of three-member teams (83%) (combined t-test  $t(690) = 8.90, p < 0.001$ ).

### 5.3. The Effect of Participant Type

In this section we study whether being presented with humans or agents affected the way participants made decisions. We focused the analysis on the extent to which participants engaged in fair behavior, as defined above, given the way their negotiation partner was presented to them.

**Insert Table 3 here**

The left-hand column of Table 3 presents the average fairness of offers made by team initiators to potential members that they perceived to be an agent or a human.

**Result 5.** *Offers made by team initiators to people were fairer than offers made by team initiators to agents.*

As shown by Table 3, the average fairness of offers made to humans (94%) was significantly higher than the average fairness of offers made to agents (83%) (combined t-test  $t(692) = 1.45, p < 0.0001$ ).

In the questionnaire subjects described their strategy for making offers. 56% mentioned that they often or always propose an even split to their members and 22% talked about creating splits that are ‘fair’, ‘decent’ or ‘reasonable’. However, 22% of the subjects also mentioned that they proposed lower offers to agents than to humans, which corresponds to the data in Table 3.

As shown by Figure 2, the type of (potential) members influenced the decisions and strategies of participants. 39% of the participants reported in the questionnaire to find the type of the player important for the decision to be part of a team. 44% stated that the type of the member affected their decision as team initiators about how to make offers. They made statements like “I felt I could ‘pay’ computers less without guilt” and “I tried to lure computers with a very low [offer]”.

Although team initiators made more offers to agents (369 proposals in sum) than they did to humans (328 proposals in sum), this difference was not statistically significant. There was also no significant difference between the percentage of accepted offers from human and agent initiators. This was also confirmed by the questionnaire: Figure 2 shows that for accepting a proposal, subjects did not give much consequence to the type of the initiator.

Furthermore, for both initiator types, offers averaging below 83% fair were mostly rejected, and offers averaging above 91% were mostly accepted. Lastly,

the type of the team initiator was of no significance to the defection behavior of team members.

**Result 6.** *There is no difference in the rate of defection from agent- versus human-led teams.*

**Insert Figure 2 here**

#### 5.4. Analysis of Defection Behavior

In section 4 we proposed that two factors would affect participants' decisions whether to defect from their current team: the size of the outside offer (i.e., participants would defect when an outside offer was higher than the one from their current agreement); and trust relationships (i.e., participants would be more likely to defect from their team when an offer originates from participants with whom they had worked with successfully in the past).

Results showed that both of these factors significantly affected participants' decisions whether to defect from their team, but that they varied in magnitude.

**Result 7.** *Although on average the outside offers were higher than the participants' current agreements, the average accepted outside offer had no significant increase over the current agreement.*

Outside offers were on average 6 points more beneficial for team members than the reward allocated to them by their current team ( $t(52) = 1.04, p < 0.05$ ). Although this gain in benefit is small, we note that participants could not see the rewards associated with non-team members and that this number also includes outside offers that were not accepted. Moreover, outside offers that were accepted did not offer a significant increase in reward as compared to the agreement with the current team.

To analyze trust relationships, we examined whether participants preferred to interact with others they had interacted with successfully in the past. Our results show that 56% of all successful three-member teams and 42% of all successful two-member teams consisted of team configurations that occurred in the past and were successful.

To provide a finer grained measure of this relationship, we defined the *trust value* between any two participants as the number of times they were both members of the same two- or three-member team. We will use this notion as a measure for the extent to which participants that have interacted successfully in the past will interact in future.<sup>5</sup>

**Insert Figure 3 here**

---

<sup>5</sup>This measure has also been used in recent work in AI that models trust among computer agents (Castelfranchi & Falcone, 1998; Jones, Fullam & Barber, 2007; Griffiths & Luck, 2003; Huynh, Jennings & Shadbolt, 2004).

**Result 8.** *Participants showed a preference to interact with those they successfully interacted with before.*

Figure 3 shows a graphical representation of the trust relationships that were formed in two of the games, played by groups of 6 participants. Nodes in the graph represent participants, and edges connect participants who were members of the same team in interactions.<sup>6</sup> Edges are labeled with trust values for each relationship. Thicker edges represent stronger relationships, meaning that participants were members of the same team multiple times. As shown by the figure, diverse relationships were established. For example, Figure 3a shows that in this particular game players with ID’s 5 and 1 preferred to work together, and Figure 3b shows that player 0 was not a popular team member in this game. Furthermore, analysis reveals that there is a positive, significant correlation of 0.31 between the trust value that exists between the team initiator and the member, and the number of times a team member accepted an offer of a team initiator.

We did not find that participants with a high defection rate were more likely to suffer in their performance than participants with a low defection rate.

**Result 9.** *Defections were not a significant predictor of performance. The average score of participants with high defection rates was not significantly different than the score of participants with low defection rates.*

## 6. Discussion

Result 1 confirms that participants understood the rules of interaction, at least to the extent that performing team activities was potentially more lucrative for participants than acting individually. Moreover, result 2 suggests that the subjects understood that collaborating in three-member teams could result in greater gains than in two-member teams. Together, results 1 and 2 confirm hypothesis 1. What is striking is that participants were more likely to form larger teams, despite the additional risks of defection and loss of time.<sup>7</sup> Result 2 was echoed in the post-study questionnaire, in which most subjects reported that the size of the reward was the most important factor that determined whether they would join teams.

Result 3 concerns the influence of team size on people’s decision to defect. In accordance with hypothesis 2, result 3 demonstrates that more defections occurred in three-member teams, leading to more unsuccessful teams. This effect was to be expected, because an increase of the number of team members leads to an increase in the likelihood of defection by one of the team members.

---

<sup>6</sup>Note that a complete graph (with all nodes connected) is not possible because participants of the same color could not be teammates in our setting.

<sup>7</sup>Note that we do not expect larger teams to be less successful than smaller teams in all settings, as the effect of team size can vary with the task at hand. However, we do wish to stress that result 2 is noteworthy given the current domain.

However, the relative rate of defection (per participant) was similar for two- and three-member teams. Therefore participants were not more likely to defect in larger teams.

The procedure of splitting rewards in our game can be considered a repeated Ultimatum Game (Guth, Schmittberger & Schwarz, 1982). The results of this study concerning the fairness of offers correspond to results found in several Ultimatum Game studies between people: (i) there are virtually no offers that are more than 100% fair, (ii) there are almost no offers with a fairness lower than 20%, and (iii) low offers are frequently rejected (Slonim & Roth, 1997; Hoffman, McCabe & Smith, 1996). Result 4 shows that the size of the team influences the fairness of the offers that initiators make to others. A possible explanation for this is that the relative payoff for members of large teams could be significantly higher than a fair offer in small teams, even though the offer in itself would not be considered fair (e.g. accepting an unfair offer of 45 points in a three-member team is potentially more beneficial than a fair offer of 30 points in a two-member team). We suggest that people preferred the higher payoff in disregard of their fairness concerns. An interesting note is that the fairness of the offers did not influence the loyalty that participants had to their teams.

Result 5 confirms hypothesis 3, namely that participants were more likely to be generous to humans than to agents when they were forming teams. However, we found no other differences in team creation and defection behavior relating to agents and people. For instance, hypothesis 4 was refuted by result 6. These results differ from past studies in which people were willing to accept lower offers from agents than they would from people in Ultimatum Games (Sanfey et al., 2003; Blount, 1995). We suggest that the reason for this discrepancy is the focus of past works on one-shot settings which are significantly different from the repeated interactions that characterize dynamic environments. We conclude that the only aspect in which agents were treated differently than people in our domain was the reward they were offered to join teams.

Contrary to hypothesis 5, result 7 shows that outside offers with significantly higher rewards were regularly declined, whereas participants regularly accepted offers with no increase, or a very small increase of the reward. Result 8 shows that people were likely to choose team members with whom they have had successful interactions in the past, confirming hypothesis 7. When the issue of trust is not taken into account, there was no *a priori* reason for participants to prefer one team configuration over another, given that the high uncertainty and partial information that characterizes our setting did not allow participants to strategize about the consequence of a potential defection in the future. This shows that trust was a more significant factor than potential reward in people's reasoning about whether to defect. Therefore, taking results 7 and 8 into account, we consider hypothesis 6 to be confirmed.

Having established that positive interactions build trust relationships in our domain, we examined whether negative interactions would lead to punishment. We did not find that defection led to lower performance, as shown by result 9. Hypothesis 8 is thereby refuted.

One possible explanation for this is that participants could not observe de-

fections in other teams, and since our game was not designed as a reputation game, participants could not share information about the reliability of other participants.

Table 4 gives an overview of the hypotheses we set out to investigate (see section 4) and the results that confirmed or refuted them.

#### **Insert Table 4 here**

We conclude this section by stating the significance of the results to behavioral and computational researchers. First, the fact that people are as loyal and committed to agent-led teams suggests that people are amenable towards working alongside autonomous systems in dynamic environments. What is more, agent-led activities may succeed in these contexts at least as well as human-led activities since people will carry through their commitments.

Second, the results confirm our main hypothesis, namely that when people are allowed to choose between humans and agents, they treat agents differently from humans with respect to social factors such as trust and fairness. This study has shown that people tend to discriminate agents by offering them less than humans. This implies that for the purpose of designing agents that are able to interact in an intelligent way in dynamic environments, the agents should be able to reason and behave in a ‘human-like’ fashion, which means taking into account factors such as trust and fairness.

Third, our data reveals that participants were likely to interact with those they successfully interacted with before. This confirms that the design of agents that interact with people over long periods of time, such as companions and game characters, must support the preservation of past experiences as a basis to generate trust between team members (Castelfranchi & Falcone, 1998; Bosse, Jonker, Treur & Tykhonov, 2007).

## **7. Conclusions & Future Work**

This paper presented an empirical study of team formation of heterogeneous human-agent teams in settings that are characterized by a high degree of uncertainty and changing information. It studied the effects of trust and fairness on people’s behavior towards humans and agents in a collaborative environment, in which agreements were non-binding and teams were not pre-defined. The study was conducted using a specially created testbed that provided an analogy to real-world task settings in which interactions are fast-paced and contain a lot of relevant information that participants have to consider when they make decisions. It showed that when negotiating to create teams, people offer less reward to agents than they do to people, but that people are as loyal to agent-led teams as they are to human-led teams. In addition, people preferred to create teams with others with whom they have had positive interactions (humans and agents alike), rather than those who offered them large rewards.

The CT configuration in this work was designed to provide an analogy to aspects that characterize team interactions, such as choosing eligible team mem-

bers and making commitments to fulfill joint tasks. It can therefore be used to serve as a generic framework for multiple domain representations. Future work will have to show whether the dynamics between humans and agents that were found in this domain extend to other interactions.

We are extending this work on several levels. First, we are creating a computational model of people's play in this game, for the purpose of building agents that use computational strategies to negotiate with people in this game. Second, we are extending the communication protocol in the game to support a reputation mechanism that will keep track of participants' defection rates in the game.

### Acknowledgements

We are grateful to the members of the Artificial Intelligence Research Group at Harvard University for their indispensable feedback on the setup of the experiment. We are especially thankful for the contributions of Prof. Barbara Grosz and Prof. Stuart Shieber of Harvard's School of Engineering and Applied Sciences. Furthermore, we would like to thank the members of the Harvard Decision Science Laboratory for enabling us to perform the experiments. Finally, we thank Maarten Engelen for his assistance with the initial implementation of the domain.

### References

- Babaian, T., Grosz, B., & Shieber, S. (2002). A writer's collaborative assistant. In *Intelligent User Interfaces Conference* (pp. 7–14).
- van Beest, I. (2002). The social psychology of coalition formation. In *Proceedings of ECPR Conference*. Turin, Italy.
- van Beest, I., Andeweg, R. B., Koning, L., & van Lange, P. A. M. (2008). Do groups exclude others more readily than individuals in coalition formation? *Group Processes Intergroup Relations*, *11*, 55–67.
- van Beest, I., van Dijk, E., & Wilke, H. (2004). The interplay of self-interest and equity in coalition formation. *European Journal of Social Psychology*, *34*, 547–565.
- Bernstein, G., & Ben-Yossef, M. (1994). Cooperation in intergroup and single-group social dilemmas. *Journal of Experimental Social Psychology*, *30*, 52–67.
- Blount, S. (1995). When social outcomes aren't fair. *Organizational Behavior and Human Decision Processes*, *63*, 131–144.
- Bolton, G., & Brosig, J. (2007). *How Do Coalitions Get Built: Evidence From an Extensive Form Coalition Game with Renegotiation and Externalities*. Working Paper Series in Economics, University of Cologne.

- Bosse, T., Jonker, C., Treur, J., & Tykhonov, D. (2007). Formal analysis of trust dynamics in human and software agent experiments. In *Proceedings of the Eleventh International Workshop on Cooperative Information Agents, CIA'07. Lecture Notes in Artificial Intelligence* (pp. 343–359). Springer Verlag.
- Bradshaw, J., Acquisti, A., Allen, J., Breedy, M., Bunch, L., Chambers, N., Galescu, L., Goodrich, M., Jeffers, R., Johnson, M., Jung, H., Kulkarni, S., Lott, J., Olsen, D., Sierhuis, M., Suri, N., Taysom, W., Tonti, G., Uszok, A., & Hoof, R. V. (2004). Teamwork-centered autonomy for extended human-agent interaction in space applications. In *In Proceedings of the AAAI Spring Symposium* (pp. 22–24). AAAI Press.
- Breban, S., & Vassileva, J. (2002). Using inter-agent trust relationships for efficient coalition formation. In *Proceedings of the 13th Canadian Conference on AI*.
- Brooks, C. H., & Durfee, E. H. (2002). Congregating and market formation. In *Proceedings of the First International Joint Conference on Autonomous Agents in Multi-Agent Systems* (pp. 96–103). ACM Press.
- Castelfranchi, C., & Falcone, R. (1998). Principles of trust for mas: Cognitive anatomy, social importance, and quantification. In *Proceedings of the International Conference on Multi-Agent Systems* (pp. 72–79). IEEE Press.
- Clancey, J., Sachs, P., Sierhuis, M., & van Hoof, R. (1998). Brahms: Simulating practice for work systems design. *International Journal of Human-Computer Studies*, (pp. 831–865).
- van Diggelen, J., Bradshaw, J., Grant, T., Johnson, M., & Neerincx, M. (2009). Policy-based design of human-machine collaboration in manned space missions. In *Proceedings of the Third IEEE International Conference on Space Mission Challenges for Information Technology*.
- Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*, 817–868.
- Gal, Y., Grosz, B., Kraus, S., Pfeffer, A., & Shieber, S. (2010). Agent decision-making in open-mixed networks. *Artificial Intelligence*, *174*, 1460–1480.
- Gal, Y., & Pfeffer, A. (2007). Modeling reciprocity in human bilateral negotiation. In *Proceedings of Association for the Advancement of Artificial Intelligence*.
- Gal, Y., Pfeffer, A., Marzo, F., & Grosz, B. (2004). Learning social preferences in games. In *Proceedings of Association for the Advancement of Artificial Intelligence*.

- Goschnick, S. (2009). People-oriented programming: From agent-oriented analysis to the design of interactive systems. In J. Jacko (Ed.), *LNCS 5610: Human Computer Interaction, Part I* (pp. 836–845). Springer.
- Griffiths, N., & Luck, M. (2003). Coalition formation through motivation and trust. In *Proceedings of International Conference on Autonomous Agents and Multiagent Systems* (pp. 17–24). New York, NY, USA: ACM.
- Grosz, B., Kraus, S., Talman, S., Stossel, B., & Havlin, M. (2004). The influence of social dependencies on decision-making: Initial investigations with a new game. In *Proceedings of International Conference on Autonomous Agents and Multiagent Systems*.
- Guth, W., Schmittberger, R., & Schwarz, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, (pp. 367–388).
- Hennes, D., Tuyls, K., Neerincx, M., & Rauterberg, G. (2009). Micro-scale social network analysis for ultra-long space flights. In *The IJCAI-09 Workshop on Artificial Intelligence in Space*.
- Hinckley, B. (1981). *Coalitions and Politics*. New York: Harcourt Brace Jovanovich, Inc.
- Hinds, P., Carley, K., Krackhardt, D., & Wholey, D. (2000). Choosing work group members: Balancing similarity, competence, and familiarity. *Organizational Behavior and Human Decision Processes*, 81, 226–251.
- Hoffman, E., McCabe, K., & Smith, V. (1996). On expectations and monetary stakes in ultimatum games. *International Journal of Game Theory*, .
- Huynh, D., Jennings, N., & Shadbolt, N. (2004). Developing an integrated trust and reputation model for open multi-agent systems. In *Proceedings of the 7th International Workshop on Trust in Agent Societies*.
- Jones, C., Fullam, K., & Barber, S. (2007). Exploiting untrustworthy agents in team formation. In *Proceedings of IEEE* (pp. 299–302). Washington, DC, USA: IEEE Computer Society.
- Jones, G., & George, J. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *The Academy of Management Review*, 23, 531–546.
- Joseph, M., & Willis, R. (1963). An experiment analog to two party bargaining. *Behavioral Science*, 8, 17–1127.
- Kamar, E., Gal, Y., & Grosz, B. (2009). Modeling user perception of interaction opportunities for effective teamwork. In *IEEE Conference on Social Computing* (pp. 271–277). Vancouver, British Columbia.

- Kamar, E., Horvitz, E., & Meek, C. (2008). Mobile opportunistic commerce: Mechanisms, architecture, and application. In *Proceedings of International Conference on Autonomous Agents and Multiagent Systems* (p. 1087). ACM New York, NY, USA.
- Katz, R., Amichai-Hamburger, Y., Manisterski, E., & Kraus, S. (2008). Different orientations of males and females in computer-mediated negotiations. *Computers in Human Behavior*, *24*, 516–534.
- Kelley, J. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Office Information Systems*, *2*, 26–41.
- Lin, R., Kraus, S., Tykhonov, D., Hindriks, K., & Jonker, C. (2009a). Supporting the design of general automated negotiators. In *International Workshop on Agent-based Complex Automated Negotiations*.
- Lin, R., Oshrat, Y., & Kraus, S. (2009b). Investigating the benefits of automated negotiations in enhancing negotiation skills of people. In *Proceedings of International Conference on Autonomous Agents and Multiagent Systems*.
- Murphy, R. R. (2004). Human-robot interaction in rescue robotics. *IEEE Transactions On Systems, Man, and Cybernetics. Part C: Applications and Reviews*, *34*.
- N. Dahlbäck, L. A., A. Jönsson (1993). Wizard of oz studies - why and how. *Knowledge-Based Systems*, *6*, 258–266.
- Nass, C., Fogg, B., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, *65*.
- Okada, A., & Riedl, A. (2005). Inefficiency and social exclusion in a coalition formation game: Experimental evidence. *Games and Economic Behavior*, *50*, 278–311.
- Oshrat, Y., Lin, R., & Kraus, S. (2009). Facing the challenge of human-agent negotiations via effective general opponent modeling. In *Proceedings of International Conference on Autonomous Agents and Multiagent Systems* (pp. 377–384).
- Pollack, M. (2005). Intelligent technology for an aging population: The use of AI to assist elders with cognitive impairment. *AI Magazine*, *26*, 9–24.
- Pommeranz, A., Brinkman, W., Wiggers, P., Broekens, J., & Jonker, C. (2009). Towards design guidelines for negotiation support systems: An expert perspective using scenarios. In *Proceedings of European Conference on Cognitive Ergonomics (ECCE)*.
- Pruitt, D., & Carnevale, P. (1993). *Negotiation in Social Conflict*. Buckingham: Open University Press.

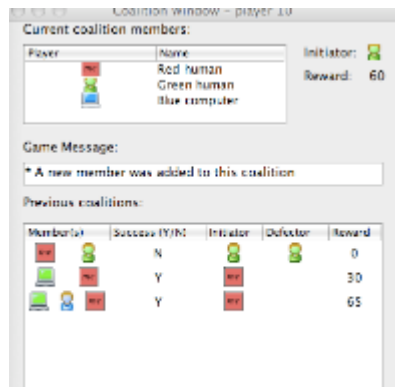
- Rajarshi, D., Hanson, J., Kephart, J., & Tesauro, G. (2001). Agent-human interactions in the continuous double auction. In *Proceedings of The International Joint Conference on Artificial Intelligence*.
- Reeves, B., & Nass, C. (2003). *The Media Equation : How People Treat Computers, Television, and New Media Like Real People and Places (CSLI Lecture Notes S.)*. Center for the Study of Language and Informatics.
- Riedl, A., & Vyrastekova, J. (2003). *Responder Behavior in Three-Person Ultimatum Game Experiments*. Technical Report Tilburg University, Center for Economic Research.
- Roth, A. (1995). *The handbook of experimental economics*. Princeton University Press.
- Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, (pp. 97–109).
- Sanfey, A., Rilling, J., Aronson, J., Nystrom, L., & Cohen, J. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*, 1755–1758.
- Schurr, N., Patil, P., Pighin, F., & Tambe, M. (2006). Using multiagent teams to improve the training of incident commanders. In *Proceedings of International Conference on Autonomous Agents and Multiagent Systems*.
- Sen, S. (1996). Reciprocity: a foundational principle for promoting cooperative behavior among self-interested agents. In *Proceedings of the Second International Conference on Multiagent Systems* (pp. 315–321). AAAI Press.
- Shieber, S. M. (1996). A call for collaborative interfaces. *Computing Surveys*, *28A (electronic)*.
- Shneiderman, B. (1987). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley Publishing Company.
- Slonim, R., & Roth, A. (1997). Financial incentives and learning in ultimatum and market games: An experiment in the slovak republic. *Econometrica*, .
- Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. *The social psychology of intergroup relations*, *33*, 33–47.
- Tambe, M. (1998). Implementing agent teams in dynamic multi-agent environments. *Applied Artificial Intelligence*, *12*, 189–210.
- Tambe, M., Johnson, W. L., Jones, R. M., Koss, F., Laird, J. E., Rosenbloom, P. S., & Schwamb, K. (1995). Intelligent agents for interactive simulation environments. *AI Magazine*, *16*, 15–39.

- van de Vijssel, M., & Anderson, J. (2004). Coalition formation in multi-agent systems under real-world conditions. In *Proceedings of Association for the Advancement of Artificial Intelligence*.
- van Wissen, A., van Diggelen, J., & Dignum, V. (2009). The effects of cooperative agent behavior on human cooperativeness. In *Proceedings of International Conference on Autonomous Agents and Multiagent Systems* (pp. 1179–1180).
- Wooffitt, R. (1997). *Humans, Computers and Wizards. Analysing Human (Simulated) Computer Interaction..* Routledge: London.
- Yorke-Smith, N., Saadati, S., Myers, K., & Morley, D. (2009). Like an intuitive and courteous butler: A proactive personal agent for task management. In *Proceedings of International Conference on Autonomous Agents and Multiagent Systems* (pp. 337–344).

## Images



(a) Main Game Panel (shown at the onset of the game) (b) Proposal Panel (used by players to make offers)



(c) Coalition Panel (displays structure and performance of past and present teams)

Action History - Name 0

Proposer	Responder	Proposed Chips	Response
me	[Avatar]	25	accept
[Avatar]	me	30	reject
[Avatar]	me	30	reject
[Avatar]	me	30	accept
me	[Avatar]	50	accept
me	[Avatar]	60	accept
me	[Avatar]	25	accept

(d) Message History Panel (shows past offers and responses)

Figure 1: Snapshots of CT GUI

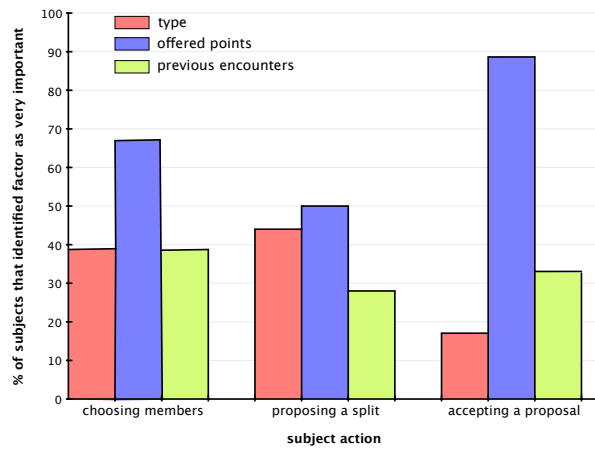


Figure 2: Factors of importance for subjects' decisions according to post-study questionnaire

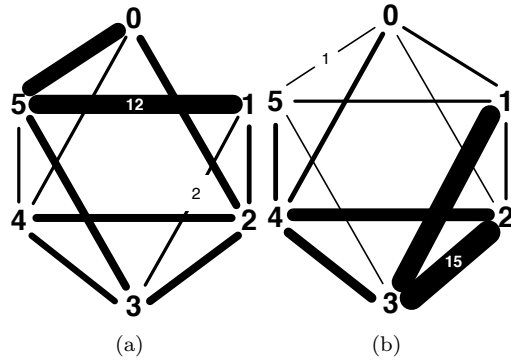


Figure 3: Examples of social graphs describing trust relationships among participants in two games

### **Captions**

1. Snapshots of the CT GUI
2. Factors of importance for subjects' decisions according to post-study questionnaire
3. Examples of social graphs describing trust relationships among participants in two games

**Figures in Color** All figures are intended for color reproduction on the Web and in print, except for figure 3.

## Tables

Table 1: Details of the CT Package Delivery Game

---

<b>participants</b>	humans & agents
<b>package pickup</b>	scattered large (LP) and small (SP) packages
<b>package delivery</b>	one central depot
<b>delivery payoff</b>	SP: 3, LP: 60 or 180
<b>forming teams</b>	negotiate over split of reward
<b>defection initiator</b>	at any given time
<b>defection member</b>	accept outside offers
<b>choose members</b>	ask anyone with different capability

---

Table 2: Statistics for formed and successful teams

	<b>formed</b>	<b>successful</b>	
	frequency	frequency	% of formed
2-member teams	122	83	68%
3-member teams	205	114	56%

Table 3: Fairness measure of offers made to humans vs. agents and in 2- vs. 3-member teams

	Humans / Agents	2-member/3-member
Fairness	94% / 82%	82%/99%

Table 4: Hypotheses discussed and the results confirming/refuting them.

<b>hypothesis</b>	<b>confirmed / refuted</b>	<b>results</b>
1. Participants prefer to form teams over individual activity.	confirmed	1, 2
2. Larger teams are more likely to fail than smaller teams.	confirmed	3
3. Human team initiators offer agents less benefit to join their teams than they offer to humans.	confirmed	5
4. Human team members are more likely to defect from agent-led teams than from human-led teams.	refuted	6
5. Team members defect because they receive outside offers that are significantly higher than the allocated reward of their current team.	refuted	7
6. Team members are more likely to defect when the received outside offer originates from a participant with whom they have worked with successfully in the past.	confirmed	7, 8
7. Team initiators prefer to invite those with whom they have established positive relationships in the past.	confirmed	8
8. Participants who are more likely to defect suffer in performance, because former team members are less likely (i) to invite them to join teams, or (ii) to agree to join teams initiated by defectors.	refuted	9