Noname manuscript No.
(will be inserted by the editor)

Reciprocity in Human-Robot Interaction

A Quantitative Approach Through The Prisoner's Dilemma And The Ultimatum Game

Eduardo Benitez Sandoval · Jürgen Brandstetter · Mohammad Obaid · Christoph Bartneck

Received: date / Accepted: date

Abstract Reciprocity [48] is an important factor in human-human interaction (HHI), so it can be expected that it should also play a major role in Human-Robot Interaction (HRI). Participants in our study played the Repeated Prisoner's Dilemma game (RPDG) and the mini Ultimatum Game (mUG) with robot and human agents, with the agents using either Tit for Tat (TfT) or Random strategies. As part of the study we also measured the perceived personality traits in the agents using the TIPI test after every round of RPDG and mUG. The results show that the participants collaborated more with humans than with a robot, however they tended to be equally reciprocal with both agents. The experiment also showed the TfT strategy as the most profitable strategy; affecting collaboration, reciprocation, profit and joint profit in the game. Most of the participants tended to be fairer with the human agent in mUG. Furthermore, robots were perceived as less open and agreeable than humans. Consciousness, extroversion and emotional stability were perceived roughly the same in humans and robots. TfT strategy became associated with an extroverted and agreeable personality in the agents. We could observe that the norm of reciprocity applied in Human-Robot Interaction has potential implications for robot design.

Eduardo B. Sandoval, Jürgen Brandstetter, Christoph Bartneck
HIT Lab NZ, University of Canterbury
Christchurch, New Zealand
E-mail: eduardo.sandoval@pg.canterbury.ac.nz,
juergen.brandstetter@pg.canterbury.ac.nz,
christoph.bartneck@canterbury.ac.nz
www.sandoval.nz

Mohammad Obaid
t2i lab, Chalmers University of Technology,
Gothenburg, Sweden
E-mail: mobaid@chalmers.se

Keywords Human-Robot Interaction · Reciprocity · Game Theory · Prisoner's Dilemma · Ultimatum Game · Cooperation

1 Introduction

Companion robots are a subset of social robots and service robots which will become popular in the near future. Dautenhahn et al. described them as robots designed for personal use, capable of performing multiple tasks and interacting with the users in an intuitive way [13]. Several studies in social robotics propose the use of these robots in different scenarios. For example as educators, caregivers in nursery houses, nannies, housekeepers and assistants. In fact, important research consortia like The Cognitive Robot Companion [1] and Robot Companions for Citizens [2] are investing resources in the development of companion robots. Moreover, it is expected that users and robots develop short-term and long-term relationships if companion robots assume certain social roles in the life of the users. In HRI, it is commonly used Human-Human Interaction (HHI) as a reference to compare our robotic implementations. Besides, reciprocity is considered a cornerstone of human social interaction [27]. For these reasons we believe it could be valuable to develop studies about how humans and robots will interact in terms of reciprocity.

The simplest cultural reference for the concept of reciprocity is "If you do something for me, I will do something for you." Reciprocity is a very important factor in human social interaction, so it should be studied in order to know how it influences the relationships of

[1] www.cogniron.org
[2] www.robotcompanions.eu

humans and robots. Many of the interactions between humans and robots involving cooperation, persuasion, altruism, exchange of favors or mutual trust could depend on reciprocity. In addition, it should be considered that humans have a high capability to adapt to agents when they are interacting with them depending of their own personality traits. For instance, people could be reciprocal with a robot by paying it back for its services (taking care of the robot, giving technical maintenance, etc) if the robot encourages reciprocity via certain social strategy. Authors like Kahn et al. consider reciprocity as a benchmark in the design of Human-Robot Interactions [30] simply because reciprocity is present in other human social situations. In other words, humans tend to develop intricate relationships with pets, machines and artifacts, consequently, it is expected that reciprocity plays an important role in HRI. However, the question is; do people reciprocate towards robots in a similar way to how they reciprocate with humans?

We consider that an analysis of reciprocity in HRI could be useful in order to design more engaging and effective Human-Robot Interactions in different scenarios. Some studies report that users do not feel engaged enough with the robots and that they have high initial expectations of them which decrease over time [7,16]. On the other hand, companion robots have not so far had the expected impact in people's lives, particularly when they take care of particular users such as elderly people [7] or children. Dautenhahn et al. found that 40% of the users liked the idea of a companion robot in the home. In addition 96.4% of the users wanted a robot capable of doing the housework. However a robot playing a role in the human domain as friend or taking care of children was acceptable to only 18% of the participants [13]. We propose that in the future, robots could assume more social roles in the human domain if the Human-Robot Interaction would be more reciprocal.

In this paper we analyze the reciprocity in HRI compared with human-human interaction (HHI). We used Game Theory insights in our experiment because this is a powerful method to establish quantitatively a model of reciprocity in HRI. Repeated Prisoner's Dilemma (RPDG) and mini Ultimatum game (mUG) have been used to model different social situations in HCI and HRI, so we also used these decision games in order to compare with that studies. Our results show that people tend to be less cooperative with robots, however they tend to be equally reciprocal with humans and robots. Also we demonstrated that Tit for Tat (TfT) strategy used by the robot is the most profitable strategy; affecting collaboration, reciprocation, profit and joint profit in the Prisoner's Dilemma game. Most of the participants tended to be fairer with the human agent in mUG. Furthermore, robots were perceived as less open and agreeable than humans. Consciousness, extroversion and emotional stability were perceived roughly the same in humans and robots. TfT strategy became associated with an extroverted and agreeable personality in the agents. We can see that the Norm of Reciprocity applies in this Human-Robot Interaction, which has generalizable implications for robot design.

2 Reciprocity and HRI

In the sixties, Gouldner proposed the "Norm of Reciprocity", defined as "the compulsion to return a favor or gift in human relationships" [27,45]. However in this study we assume a more complete definition of reciprocity proposed by Fehr and Gachter: "Reciprocity means that in response to friendly actions, people are frequently much nicer and much more cooperative than predicted by the self-interest model; conversely, in response to hostile actions they are frequently much more nasty and even brutal" [20]. This definition is in line with the theory of reciprocity proposed by Falk and Fischbacher [18] based on experimental work. The theory explains a reciprocal action modeled as the behavioral response to an action that is perceived as either kind or unkind. In addition, reciprocity has been a well studied topic in humanities, philosophy, social sciences and psychology. Moreover reciprocity is connected with other phenomena such as persuasion [10] cooperation [4] , altruism [33], friendship [11], love [35] and compassion [55].

There are several studies about reciprocity in HRI with different approaches. Kahn et al., [31] discovered that children responded reciprocally and were more engaged with an AIBO robot which offered some motioning, behavioral and verbal stimulus than they were with a toy dog. Specifically, they claim that reciprocity is one of the benchmarks in the design of Human-Robot Interaction. Moreover, Kahn et al., speculate with an interesting question in HRI, which is: Can people engage substantively in reciprocal relationships with humanoids [30]? They argue that interactions involving reciprocity with anthropomorphic robots can be similar to human interactions [30]. Nass and Reeves have conducted ample research about how people tend to anthropomorphize objects such as computers [41]. Consequently, they conclude that humans could have similar attitudes toward inanimate objects and humans. In a very specific study Fogg and Nass demonstrated that users tend to be reciprocal with computers that had helped them previously [24].

More recently reciprocity has been very present in the debate of social robotics. The workshop: "Taking

care of Each Other: Synchronization and Reciprocity for Social Companion Robots" in the International Conference of Social Robotics 2013 discussed the importance of reciprocity in the design of companion robots. Several studies presented in the workshop reviewed concepts related to reciprocity as compassion, behavior imitation or social cognition mechanisms integrated to HRI [56] which could be the cornerstone in the development of future meaningful Human-Robot Interactions.

For instance Weiss presented the project Hobbit [55], a robot based in the "Mutual Care" paradigm proposed by Lammer et al. [36]. They, like us, propose that Human-Robot Interactions can improve if both parties take care of each other in a similar way to human human interactions. Furthermore, Lorenz claims that mutual compassion (understanding Compassion as the german word "Mitgefühl") should be considered as an important component in HRI due to this being a human ability based in synchronization and reciprocity. The benefits of mutual understanding based in a reciprocal relationship between humans and robots can improve the performance of social companion robots because of the resulting more intuitive behavior of the robot [38]. However, Broz and Lehmann claim that reciprocity is limited to certain HRI scenarios where robots assist humans in some activities and humans assist the robots in others. Although cooperation and reciprocity are closely related, they do not necessarily appear together. For instance in jobs as caregivers, which could be likely future roles for companion robots. The patients do not necessarily behave in a reciprocal manner with the caregiver [8]. It could be because the robot doesn't encourage reciprocal behaviour. Likely this lack of reciprocity in HRI can produce a depreciation of the services provided by the robot. Consequently the construction of a relationship will be degraded. We think that reciprocity is especially important if the users need the robot. If something happens to the robot but the user does not care, user will suffer later a negative impact on him/her because the robot could not do its work. In our opinion other roles for companion robots could require a more reciprocal behavior when a social interaction is developed.

In terms of applications using the reciprocity concept in HRI; there are several examples of how the design of reciprocal behaviors could be applied in Children with Autism Spectrum Disorder [40] or elder care [36]. However a better understanding of reciprocity could help to improve the current use of companion robots in real applications like in the work presented by Broadbent et al., [7]. In that study robots did not have significant impact on the quality of life of the patients, depression or adherence. It is likely that a more reciprocal behavior of the robot could help to improve its performance with the patient.

Additionally, decision games such as those played in this study have been used to study different aspects of HRI [50]. To illustrate, Nishio et al. [42] have studied how the appearance of robots affects participants in an Ultimatum Game (UG). This game involves reciprocity because two players interact to decide how they will divide money or points in fair or unfair proposals. Nishio et al. conclude that people show changes in their attitude depending on the agents appearance. The agent (robot, human or computer) in the role of proposer influences the number of the rejections of the proposals. In particular an android appearance is associated with a higher number of rejections. Possibly not enough human likeness in the android's appearance is a main factor. In addition, Torta et al. used Ultimatum Game online to measure the perceived degree of anthropomorphism among a human agent, a humanoid robot and a computer. In that study, participants took more time to respond to the offer of a computer compared to that of the robot [53].

Despite the importance of reciprocity in HCI and HRI, the area has still not been explored enough. The research related with reciprocity is mainly focused in persuasion, negotiation and cooperation. Apparently the community of social robotics accepts reciprocity as a fact. However we consider that reciprocity should be measured and compared in order to have a reference to be used as a guideline in the design of new interactions.

Additionally, it is assumed that robots and other machines should be cooperative with humans but these studies have not considered reciprocity as a main factor in this phenomena. For instance, Fogg developed the concept of persuasive machines [23,22,21], considering that humans have an instinctive behavior towards devices that triggers feelings and emotions in response to their persuasiveness. These feelings and emotions are apparently reciprocal to the machines when they provide a good service or help. In other words, "If you are nice to me, in the future I will be nice to you".

Besides, negotiation is an activity which inherently involves reciprocity in order to obtain satisfactory results for negotiators. Several studies have been done with automated agents negotiating in different decision scenarios. Lin and Kraus offer an extensive review of these agents in [37]. The performance of the agents varies statistically significantly depending on the scenario and the internal design of the algorithms. Moreover, Kiesler et al. showed that humans show cooperative behavior towards computers [32] playing Prisoner's Dilemma when they have a chance to interact intensively with the agent. In this Prisoner's Dilemma the

cooperation is conditioned to the previous actions of the other participants; if a player was cooperative or defective that could condition the response of the opponent in the next round (reciprocal behavior), so "I will be nice with you now because in the future I expect that you will be nice to me too". De Melo et al. also used Moral Emotions (gratitude, anger, reproach, sadness) to elicit cooperation with a virtual agent in 25 rounds of Prisoner's Dilemma using a variety of Tit for Tat strategy [39]. However none of these studies consider reciprocity as a variable to be measured.

2.1 Game Theory as a research tool in HRI

To explore reciprocity we decided to use the insights of Game Theory. The definition we use is as follows: "an interdisciplinary theorist method that examines how people make decisions when their actions and fates depend on the actions of other people" [57]. We used Repeated Prisoners Dilemma (RPDG) and Ultimatum Game as the decision games that could offer us a quantitative reference of reciprocity in HRI. Both games are a common research tool used to investigate other related phenomena as cooperation or negotiation allowing simplification of different social situations. Additionally these games can be changed to model other scenarios. For instance, Prisoner's Dilemma could be adjusted without modifying the essence of the game for different situations where participants should take decisions such as in wars, law enforcement, or duopoly fights [52].

2.1.1 Prisoner's Dilemma

The Prisoner's Dilemma game is frequently used as a quantitative approach to study different phenomena. Since Rapoport and Chammah proposed the Prisoner's Dilemma in 1965 [47] there have been different versions of the experiment which differ in the terms of the defection and collaboration required of the players. In the original game two thieves are captured by the police and interrogated separately. They can cooperate with each other keeping quiet or they can defect confessing the crime, but the punishment of both thieves depends of the combination of cooperations or defections of each. The rules are: "There are two players. Each has two choices, namely cooperate or defect. Each must make the choice without knowing what the other will do. No matter what the other does, defection yields a higher payoff than cooperation. The dilemma is that if both defect, both do worse than if both had cooperated" [4]. One of the matrix versions of the game is shown in Table 1 [52].

In Table 1 the numbers represent time in prison for the participants in the game. The minus sign is a convention to indicate that this time is subtracted from the time of the criminal in the metaphor. To illustrate, if Criminal 1 and Criminal 2 both cooperate (keep quiet), both will spend just three months in jail. However, if Criminal 1 cooperates and Criminal 2 defects, Criminal 1 will spend 12 months in jail and Criminal 2 will be free. If both of them defect they will spend eight months in jail. The game represents situations where simultaneous decisions affect two parties.

		Criminal 2	
		Cooperate	Defect
Criminal 1	Cooperate	(-3,-3)	(-12,0)
	Defect	(0,-12)	(-8,-8)

Table 1 Basic Prisoner's Dilemma Matrix

Defect offers the highest profit for the players when the game is played once. Therefore strict dominance here is Defect. Spaniel defines a strict dominance when "We say that a strategy X strictly dominates strategy Y for a player if strategy X provides a greater payoff for that player than strategy Y regardless of what the other players do." [52]. In other words, when we have a strict dominant strategy in a decision game it should be clear for the participants what to decide in order to get the highest profit. For a single round of Prisoner's Dilemma, Defect is the strict dominance strategy because it allows a player to avoid punishment. However when many rounds are played, Cooperate or Defect are possible strategies to reduce the punishment of both players.

Diverse versions of Prisoner's Dilemma have been developed. For instance, Prisoner's Dilemma can also be played in consecutive rounds, which is called Repeated Prisoner's Dilemma Game (RPDG) modality. In this version, previous movements of the opponent become a factor for the next movement of the player, who is probably considering and recording the behavior of his opponent [57]. Furthermore, about 20 strategies have been tested in order to get a good score in the RPDG [14,2]. According to Axelrod the strategy designed by Rapoport, "Tit for Tat" (TfT) is the simplest and most effective strategy to follow in the RPDG [4]. Tit for Tat consists of cooperating in the first instance and then in the next movement copying the decision of the other participant did in the previous round. In two contests organized by Axelrod in the 1980s different strategies were tested. In both contests Tit for Tat was the winner [3].

2.1.2 Ultimatum Game

In this game, one of the participants (Proposer) decides how to distribute a certain amount of money. The second player (Acceptor) can decide to accept the distribution and both of them can keep the money. However, if the acceptor rejects the offer both of them lose the money. Like Prisoner's Dilemma, Ultimatum Game has different variants. One is the mini Ultimatum Game (mUG) in which participants decide upon a limited set of defined distributions of money, for example, 50%-50%, 20%-80%, 80%-20%, or other options [17]. For this study, we use the mUG version of the Ultimatum Game, and fixed the roles for the agent and the participant. Participant is always the proposer and Agent is the acceptor.

2.2 Studies of personality and reciprocity

Several researchers claim that human personality matters in games related to reciprocity such as Prisoner's Dilemma. Park et al. claim that the behavior in situations involving reciprocity is affected by personality and the interactions of the parties following the norm of reciprocity. In addition, they suggest that extroversion, agreeableness and neuroticism personality traits are related to cooperative strategies in conflict resolutions. [44]. Boone et al. conducted an experiment which deals with four personality traits: locus of control, self-monitoring, type-A behavior and sensation seeking [5]. In addition, Chaudhuri et al., performed the Repeated Play Prisoner's Dilemma (RPPD) researching trusting and reciprocal behavior [9]. They classified people with different propensities to cooperate showing differing degrees of trust and reciprocity. They found that people who chose to cooperate demonstrated higher levels of trust. In contrast, in reciprocal behavior, differences between cooperative subjects and defectors were not significant.

2.3 Research Questions

Our general research questions for this study are: Do people reciprocate differently towards other humans in comparison to robots? What consequences does the interaction strategy of the robot have on the humans' reciprocal behavior? In order to answer these questions we developed the following sub questions:

1. Do participants behave differently towards robots compared to other humans in terms of reciprocation, collaboration and the offer they make in the ultimatum game (Offer)?
2. Do participants behave differently towards agents that use the TfT strategy in comparison to how they behave with agents that use the Random strategy in terms of reciprocation, collaboration and the offer they make in the Ultimatum Game (Offer)?
3. Do participants win more money (Profit) when the agent uses the TfT strategy compared to when the agent uses the Random strategy?
4. Do participants and robots together win more money (Joint Profit) when the agent uses the TfT strategy compared to when the agent uses the Random strategy?
5. Is there any correlation between Collaboration, Reciprocation, Profit and Joint Profit?
6. Is the personality of the agent perceived differently when the agent uses the TfT strategy compared to when the agent is using the Random strategy and how is this relationship mediated by the participants' own personality?

3 Method

The aim of this paper is to model reciprocity with a quantitative approach in order to understand the reciprocal actions of the participants towards the robots. We used the Repeated Prisoner's Dilemma Game (RPDG). The participants played ten rounds similar to the experiment of Selten and Stocker who did a series of "super games" playing 25 times in periods of ten rounds [49]. Then the participants played as proposer and the agent as acceptor in the mini Ultimatum Game (mUG).

In our study the participants did not know how many times they would play against the agent. That means that their decisions would be conditioned by the possibility of interacting with the agent in an undetermined number of rounds. Apparently when people do not know the number of rounds they tend to be more reciprocal and collaborative due to the reputation of the opponent in the previous rounds [34, 1]. It is also necessary to have multiple interactions to be able to evaluate the personality of the opponent [9, 28]. That could have an impact in the long-term relationships between humans and robots. It takes several rounds of playing the game to get an impression of the strategy of the opponent [51]. However, cooperation is not stable along the RPDG and it tends to deteriorate when the game is played anonymously over ten rounds [19, 29].

In order to answer our research questions we developed a 2x2 mixed within/between experiment. The between factor was the agent, which could be either a human or a robot. The within factors were the strategies played by the agent, which could be either Tit for

Tat (cooperate in the first movement and then do whatever the other participant did in the previous move) or Random strategy.

We ran our experiment using robot agents and human agents in order to compare the behavior of the participants under the same controlled conditions. We used two robots, one of them customized with stickers, to avoid the possibility that the judgments of participants for the second within-condition would be influenced by the experiences made with the robot in the first within-condition. The participants would either first play with a robot that used the Tit for Tat strategy or with a robot that would use the Random strategy. In addition, we changed the robot every set of games, so either robot "A" or robot "B" would be the robot that used the Tit for Tat strategy. This comparison is a typical study of effectiveness of the strategies in Prisoner's Dilemma [4,2,45]. After one round of ten games of Prisoner's Dilemma the participants played one round of Ultimatum Game.

We followed the same setup for the human condition. Two male confederates were available to play versus the participants. We cannot control the physical appearance of the human agents; however, we asked them to be neutral and interact as little possible with the participant and avoid conversation. They would just respond nodding to the greeting of the participant at the very beginning and listening to the same instructions given to the participants. The participants did not know that they were playing with a confederate.

3.1 Measurements

One of the researchers recorded manually all the actions of the participants and the agent. The actions included the behaviors in every round of the Prisoner's Dilemma game (collaborate or defect). The record also contained how much money the participants were left with after each session. Participants and agents pointed out to the cards with the words "Cooperate" or "Defect". In addition, the log included the decision of the participants in the two Ultimatum games of each round.

The variables were the number of Cooperations and Reciprocations done in every set of Prisoner's Dilemma and the Offer made in Ultimatum Game. The number of Cooperations (frequency of cooperation) along the game was the variable that allowed us calculate the number of reciprocations (frequency of reciprocations). The number of reciprocal movements was calculated by counting the number of cooperative choices of the agent followed by the cooperative choice of the participant plus the number of defective choices of the agent followed by defective choices of the participant. See Figure 1

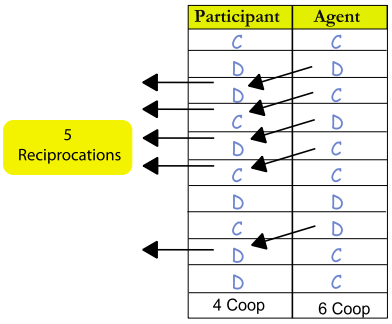

Fig. 1 Example of the computation of Cooperations and Reciprocations

A computer-based questionnaire recorded the demographic data. The same computer was used to apply the TIPI Test developed by Goslig et al [25] that was used to evaluate the Big Five traits of personality (extroversion, agreeableness, conscientiousness,neuroticism or emotional stability and openness)in the participant and the perception of personality of the agents. We chose this test because it could be answered by the participant in a short time provides reliable results.

Also, we tried to discover how humans and robots reach a goal. In this case the money used in the experiment is an outcome to measure how reciprocity affects joint tasks. A probable question of the reader about this experiment is: Why use money if robots do not need it? We must keep in mind that the original metaphor of the Prisoner's Dilemma describes a scenario avoiding spending time in jail. Money represents this time in jail. It is a token; a tangible representation of this metaphor. Robots don't need money; however, the coins used in the game are useful because can show us how humans and robots can perform a task together according to the degree of reciprocity between them. The less money humans and robots lose can be compared with the less time they spend in the hypothetical jail.

3.2 Development of the experiment

The experiment consisted of four phases which are shown in Figure 2. Participants were welcomed and taken to the experimentation room. In the case of the human condition actors arrived roughly at the same time and were in another room pretending to fill the same questionnaires as the real participant. Once in the room, the participants completed the consent form and filled

in the demographic and personality questionnaire (TIPI Test). Then, the metaphor of the Prisoner's Dilemma game was used to explain the structure of the RPDG used in this experiment. The rules of the game were stated before the participants played two trial rounds against the agent. The participants were informed that they could keep whatever money would be left at the end of the game.

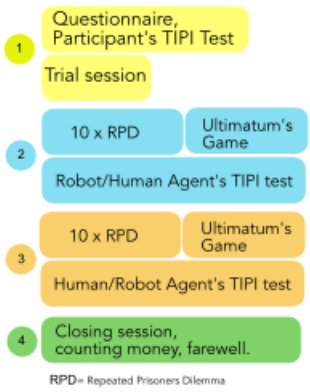

Fig. 2 Step-by-Step procedure for the participant.

After that, the experimenter explained the Ultimatum game and participants played one trial round with the same agent. The experimenter explained that the participant would be the proposer. The agents made the same pre-determined responses during the trials. The word "robot" was changed in the card by the word "agent" in the human condition.Three cards with different distribution of money were in the table. The participants chose one card and showed it to the agent. The participants were told that the agent would now make a choice whether to accept the offer or not. The agent was instructed to always accept the offer but the participants were not made aware of this fact.

After the practice session, participants continued with the second phase in which they played a first Prisoner game and started with NZ $6.50. Each session consisted of 10 rounds of Repeated Prisoner's Dilemma [28] against an agent followed by one round of Ultimatum Game in a common face-to-face configuration. At the beginning of each Prisoner's Dilemma round the referee rang a bell to signal the players to make their choice. After both players had chosen a card, the experimenter removed the board to allow both players to see each others decision. After that the participants gave the money they had lost following the matrix. The experimenter took the money from the robot. Then the participants played the Ultimatum Game with the agent. When the game was over, the participant completed an agent personality questionnaire on the computer. During that period, we changed the agent. This procedure was clearly visible to the participants and the experimenter informed the participants that in the next session they would be playing with a different agent. In the case of the human agent we pretended that he would fill in the questionnaire in other room.

In phase three, the participant then played a second Prisoner's Dilemma game and Ultimatum Game with the other agent. If the first agent played Tit for Tat then the second agent played the Random strategy. Afterwards the participants filled in the personality questionnaire for the new agent. Finally in phase four the participants were asked to count their money and we closed the experiment asking for their comments.

3.3 Setup

We used NAO Robots manufactured by Aldebaran [26]. One of the robots was customized with stickers. The robots performed programmed movements, controlled by a tele-operator hidden by a curtain. A hidden camera (not recording) provided a video of the situation and enabled the operator to enact both strategies. For the human condition the actors followed a script and tried to have a neutral behavior towards the participants. They used similar clothes and had limited interaction with the participants.

The experiment took place in a 3m x 3m area. In order to reduce the distractions for the participants we tried to keep the experimental area as minimalistic as possible. The participants were seated on a table opposite the agent, because face-to-face configuration increases collaboration amongst human players [51,29]. Oda claims that recognition of the opponent's face is a crucial factor when humans use a Tit for Tat strategy in social interactions [43].

A sliding board was used to allow the agent and participant to make private decisions in the Prisoner's dilemma game (see Figure 3). The referee was seated on the side of the table and was able to remove the sliding board in order to let the players see each others choice. A second table was located in the corner of the room for the computer with the questionnaires.

The Prisoner's Dilemma was based on the matrix shown in Table 2. The numbers are New Zealand dollars that the participant lost depending on whether he or she cooperated or defected. In this scenario defection is not punished and cooperative behavior is poorly rewarded. The distribution of the money keeps the configuration of the original Prisoner's Dilemma, with 30

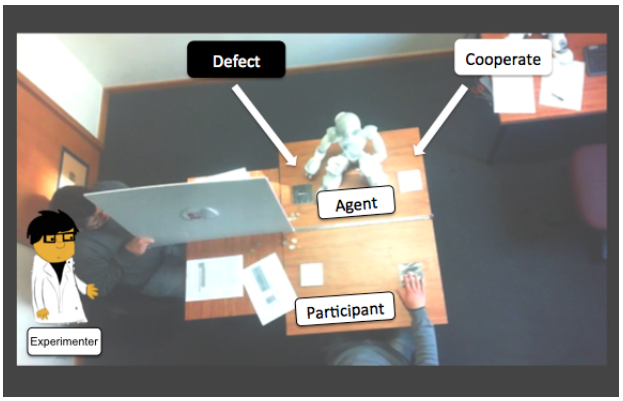

Fig. 3 Setup of the experiment.

cents, 50 cents and 1 dollar rewards depending on the combined actions. The participants received $6.50 in coins at the beginning of each Prisoner's Dilemma session. For the Ultimatum game the participants shared $2.

		Primes	
		Cooperate	Defect
Powers	Cooperate	(-0.3,-0.3)	(-1,0)
	Defect	(0,-1)	(-0.5,-0.5)

Table 2 Matrix used in the experiment. The values represent the dollars that participant lose.

The choices of the agents using Random strategy were based on four scripts of pseudo-random sequences of movements. Each script consisted of five collaborations and five defections. This quasi-random behavior ensured that the agent would not make an extremely low or high number of cooperations. The robot randomly picked one the four scripts. As we explained in 2.1.1, Tit for Tat strategy is based on the previous decision of the participants. For the first round that is not possible hence the agent always picked "cooperate" for its initial decision. The actors followed the same strategies, they could read the scripts of the random sequences during the game, and the script could not viewed by the participant.

Two cards with the labels "Cooperate" and "Defect" were placed in front of the participant and a second set in front of the agent. The participants and the agents had to choose their behavior in the game pointing to one of the two cards in front of them. In the Ultimatum Game participants used three pre-defined options printed on cards [42]. The three options were: (Robot 50% - Human 50%), (Robot 20% - Human 80%) (Robot 80% - Human 20%). For the human condition we changed the words on the cards to "Participant A" and "Participant B".

3.4 Participants

We used data [3] of sixty participants in the experiment: 30 in the robot condition and 30 in the human condition. All of the participants were recruited at the University of Canterbury and Facebook groups from Christchurch. The nationalities were diverse: 38.3% were from New Zealand, 18.3% Chinese and other Asian countries, 18.33% Latin Americans and Caribbeans, 5% Indians, 3.3% Middle East, 3.3% Russians and finally 13.3% from other Western Countries. Of the 60 participants, 39 were men. The average age was 26.5 years old (SD= 6.5); median 24.5. Only 40% of the participants had previous experience with a real robot.

In the robot condition the participants were 18 males and 12 females, whose ages averaged 28.27 years (SD = 6.73). Nine came from New Zealand; the rest from overseas. Half of them were in paid employment. Thirteen participants had previously interacted with a robot and seventeen had not. In the human condition the participants were 21 males and 9 females whose ages averaged 24.7 years (SD=5.96). Fourteen came from New Zealand and the rest from overseas. 73% were in a paid employment. Eleven participants had previously interacted with a robot and nineteen had not. All participants received an explanation of the procedure and signed the consent form. To raise their motivation, participants were told that their compensation would be how much they won in the games.

4 Results

We performed a mixed repeated measure ANOVA in which Agent was the between subject factor and Strategy was the within subject factor. The measurements were Cooperations, Reciprocations, Offer, Profit and Joint Profit. Figure 4 shows the medians and standard deviations of Cooperations and Reciprocation measurements across the four conditions. Figure 5 shows Profit, Joint Profit and Offer in Ultimatum game along the four conditions as well.

4.1 Differences between agents

Our first research question compares the agents in terms of reciprocation, cooperation, profit, joint profit and the

[3] Our data is available in http://goo.gl/NcKRBl as a .sav file

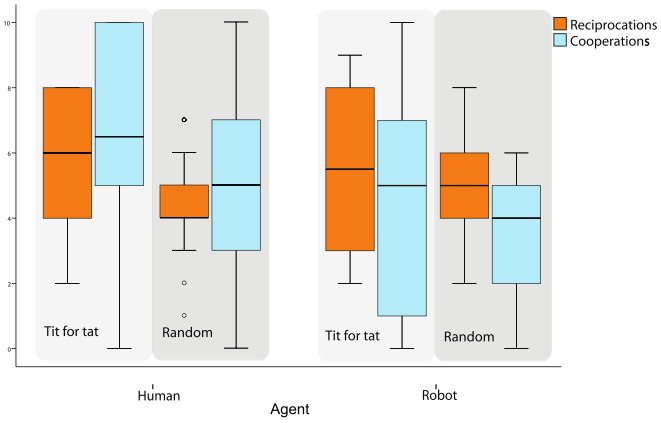

Fig. 4 Number of cooperations and reciprocations in the experiment.

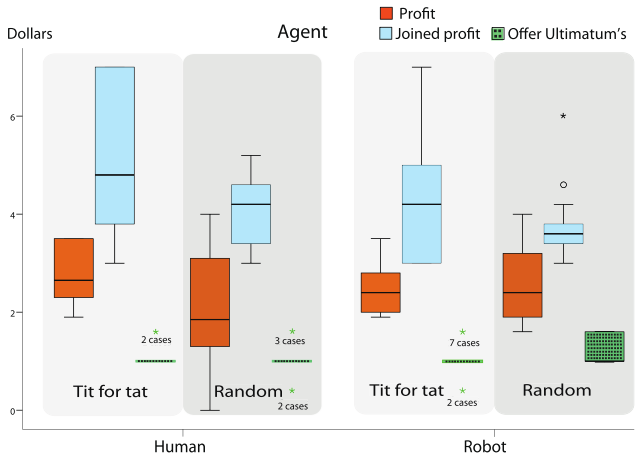

Fig. 5 Profit, Joint Profit and Offer made in Prisoner's Dilemma and Ultimatum Game.

offer that participants made in the Ultimatum Game. We observed that participants that interacted with a robot did not show significantly more reciprocations (m=5.3, SD=2.019), than when they interacted with a human agent (m=5.067, SD= 1.973), F(1,58)=0.349, p=0.557. Furthermore, Participants that interacted with a robot showed significantly fewer cooperations (m=4.15, SD= 2.72) than when they interacted with a human (m= 5.82, SD=3.13), F(1,58)= 6.889; p=0.011. Joint Profit was significant affected by the type of agent, F (1,58) = 6.418, p=0.014. Participants in the human condition had on average a Joint Profit of $4.64 (SD=1.31) and in the robot condition $4.05 (SD=1.11), not significant difference found. Profit of participants in the robot condition is in average $2.55, (SD=0.646) was not significantly higher than the average profit made in the human condition $2.39, (SD=0.976), F(1,58)= 1.778, p= 0.188.

We ran a chi-square analysis of the Offer in Ultimatum Game treating data as nominal variables. The frequency of the offers made to the human agent (f(50%) = 53, f(20%)= 5, f(80%)= 2) is significantly different from the offers made to the robot agent (f(50%)= 43, f(20%)= 15, f(80%)= 2), χ^2 = (2, N= 60)= 6.042 p= 0.039. In other words, reciprocations and profit were not significantly affected by the type of agent. There is a significant interaction effect between the agent and the strategy for the profit of the participant, F(1,58)=5.842, p=0.019. Participants who interacted with a human agent that used the Random strategy won less money than in the other conditions. A summary of the results can be found in Table 3 in the top of the next page.

4.2 Differences between strategies

Our second research question was if participants behave differently towards agents that use the TfT strategy in comparison to agents that use the Random strategy in terms of reciprocation, collaboration and the Offer they make in the Ultimatum Game.

Participants who played with the agent that used the TfT strategy collaborated (m=5.73, SD=3.39) significantly more than when they played with the agent that used the Random strategy (m= 4.23, SD=2.44), F(1,58)= 15.982, p <0.01. Furthermore, participants who played with the agent that used TfT strategy reciprocated (m=5.65, SD=2.31) significantly more than when they played with the agent that used the Random strategy (m=4.72, SD=1.497), F(1,58)= 9.019; p =.004.

We ran a chi square analysis in order to observe how the strategy affects the frequencies of the offer made to the agent in Ultimatum Game. The frequency of the offers made when Random strategy (f(50%)= 47, f(20%)= 11, f(80%)= 2) was played is not significantly different from the frequency of the offers made when TfT strategy was played (f(50%)= 49, f(20%)= 9, f(80%)= 2), χ^2 = (2, N= 60)=0.242 p=0.926.

In terms of money, the results show that participants who played with the agent that used TfT strategy made an average profit of $2.64, (SD=0.58) significantly higher than when they played with the agent that used the Random strategy m=$2.3, (SD=0.99), F(1,58) = 4.239; p=0.044. Also participants who played with the agent that used TfT strategy made an average Joint Profit of $4.80, (SD=1.5) significantly higher than when they played with the agent that used the Random strategy m=$3.83, (SD=0.66), F(1,58)=28.913; p <0.01. A summary of our analysis for question 2,3 and 4 is in Table 4.

Human vs Robot			Robot		Human	
Variable	F	p-value	Mean(SD)	SE	Mean(SD)	SE
Reciprocations	F(1,58)=0.349	0.557	5.3 (2.019)	0.261	5.067 (1.973)	0.255
Cooperations	F(1,58)=6.889	0.011	4.15 (2.717)	0.351	5.817 (3.133)	0.404
Profit	F(1,58)= 1.778	0.188	2.55 (0.646)	0.083	2.39 (0.976)	0.126
Joint Profit	F(1,58)= 6.418	0.014	4.05 (1.108)	0.143	4.64 (1.309)	0.169

Table 3 Number of reciprocations and Profit were not significantly different between the agents. Number of cooperations and Joint Profit were significantly different.

Tft vs Random			TfT Strategy		Random Strategy	
Variable	F	p-value	Mean(SD)	SE	Mean(SD)	SE
Reciprocations	F(1,58)= 9.019	0.004	5.65 (2.306)	0.298	4.717 (1.497)	0.193
Cooperations	F(1,58)= 15.982	<0.01	5.733 (3.394)	0.438	4.233 (2.438)	0.315
Profit	F(1,58)=4.239	0.044	2.645 (0.585)	0.075	2.3 (0.989)	0.127
Joint Profit	F(1,58)=28.913	<0.01	4.807 (1.501)	0.193	3.833 (0.657)	0.084

Table 4 In terms of strategy; Reciprocations, Cooperations, Profit and Joint Profit were significantly different between strategies.

4.3 Correlation between collaboration, reciprocation and money

We wanted to know if there was any correlation between Collaboration, Reciprocation, Profit and Joint Profit? We conducted a multiple regression analysis between Reciprocation, Collaboration, Profit, Joint Profit and Offer. The Pearson Correlation Coefficients are shown in Table 5. Reciprocation was significantly positively correlated with Collaboration, Profit and Joint Profit. Joint Profit is significantly positively correlated with Collaboration and Profit. Also, Offer is significantly positively correlated with Profit.

	Rec	Coop	Prof	Jprof
Coop	*0.182			
Prof	*0.241	-0.065		
Jprof	*0.405	*0.872	*0.281	
Offer	-0.019	0.008	*0.258	-0.033

Table 5 Pearson Correlation between Reciprocation and Collaboration, Profit, Joint Profit and Offer. The * sign indicates a significance level of p<0.05. Rep= Reciprocity, Coop= Cooperation, Prof=Profit, Jprof=Joint Profit

The regression equation is:

$$\begin{aligned}\text{Reciprocation} = 0.133 + (-0.68 \times \text{Collaboration}) + \\ (-0.557 \times \text{Profit}) + (2.211 \times \text{Joint Profit}) + \\ (0.754 \times \text{Offer})\end{aligned} \tag{1}$$

The model is able to explain 0.310% of the variance in the Reciprocation model.

4.4 The personality traits as factors in the experiment

We asked whether the personality of the agent is perceived differently when the agent uses the TfT strategy compared to when the agent is using the Random strategy, and how this relationship is mediated by the participant's own personality. We conducted a mixed repeated measure ANCOVA in which the agent was the between factor, strategy was the within factor and the personality traits of the participant were the covariants. The perceived personality traits of the agent were the dependent variables.

Our analysis shows that agent had a significant influence on the perception of the agent's agreeableness, F(1,58)=4.263, p=0.044. Participants who interacted with a robot agent perceived less agreeableness (m= 4.067, SD=1.361) compared to participants interacting with a human agent (m=4.517,SD=1.017). Also agent had a significant influence on the perception of the agent's openness. Participants who interacted with a robot agent perceived less openness (m=3.458, SD= 1.488) compared to participants interacting with a human agent (m=4.408, SD=0.95), F(1,58)=8.682, p= 0.005. However, agent did not have a significant effect on perceived extroversion of the agent (F(1,58)=0.102, p= 0.750), conscientiousness (F(1,58)= 0.113, p=0.738) or emotional stability (F(1,58)=0.005, p=0.944).

Participants that played with the agent that used the TfT strategy scored the agent significantly (F(1,58) = 4.865, p=0.032) lower on Extroversion (m=3.533, SD=1.1963) than when they played with the agent using the Random Strategy (m=3.558, SD= 1.1648). Also, participants that played with the agent that used the TfT strategy scored the agent significantly (F(1,58)= 3.586, p=0.064) higher on agreeableness (m=4.5, SD=

1.30, SE=0.168) than when they played with the agent using the Random Strategy (m= 4.083, SD=1.097, SE= 0.141).

However, strategy did not have a significant effect in perceived Openness (F(1,58)=1.94, p=0.17), Conscientiousness (F(1,58)=1.902, p=0.174), or Emotional Stability (F(1,58)=0.301, p=0.586). Interaction effect between strategy and participant conscientious appeared on the perceived extroversion (F(1,58)=6.047, p=0.017) and agreeableness (F(1,58) = 4.569, p=0.037) of the agent.

In summary, Agent had a significant influence on the perception of the agent's agreeableness and openness. The robot agent was perceived as less agreeable and less open than the human agent. Agent didn't have any influence in the perceived agent's extroversion, conscientiousness or emotional stability. Strategy had an influence in the perceived agents' extroversion and agreeableness, but not in the agents' perceived openness, conscientiousness or emotional stability. An agent using TfT strategy was scored lower in extroversion and higher in agreeableness compared with agents that used Random strategy. An agent's perceived extroversion and agreeableness were affected by an interaction effect between strategy and the participant's conscientious.

We also investigated the influence of the participants' personality traits on the perceived personality of the agents. We explored this relationship using the covariants. The results show that participants' extroversion had a significant effect on the perceived level of the agents' emotional stability (also called neuroticism) (F(1,58)= 7.907, p= 0.007). Also participants' agreeableness had a significant effect on the perceived level of the agents' openness (F(1,58)= 7.680, p= 0.008). Participants' openness had a significant effect on the perceived level of the agents' agreeableness (F(1,58)= 5.795, p= 0.020) and agents' emotional stability (F(1,58) = 5.192, p= 0.027). All the effects are positive correlated among them.

The influence of the personality traits in the participants as covariants for the perceived personality traits in the agent are shown in Table 6.

Participant's trait	**Perceived trait in the agent**
Extraversion	Emotional stability
Agreeableness	Openness
Openness	Agreeableness
	Emotional stability

Table 6 Covariants related with perceived personality traits in the agent.

4.5 Our Results compared with literature

We compared the results in both robot and human conditions using the Tit for Tat strategy with the results obtained from the study reported as the Flood-Dresher experiment in [46, 57] in terms of cooperation in RPDG. They reported that in 100 rounds of RPDG participants decided to collaborate in average 68% of the rounds. We performed a one-sample t-test to compare the data from our human and robot condition to this value. In both conditions, human agent and robot agent, there were fewer Cooperations. Participants cooperate significantly less (48.3% of the rounds) with the robot compared with 68% reported in the Flood-Dresher experiment, (t(29)= 7.095, p<0.01). Also, participants cooperate significantly less (66.3% of the rounds) with the human agent in our experiment compared with 68% reported in the Flood-Dresher experiment, (t(29)= 9.623, p<0.01). Although this is a significant difference it does not have practical implications. The difference between the means is minimal. In general terms we can say that our results are in line with the results shown in the Flood-Dresher experiment, and the slight difference can be attributed to uncontrolled variables in both experiments.

5 Discussion and Conclusion

Our results and the literature review show that people tend to cooperate more with a human agent than with robots. However, our results also showed no significant difference in the number of reciprocations in both agents. Apparently the participants tend to be similarly reciprocal with humans and robots. The Norm of Reciprocity seems to apply to Human-Robot Interaction using the Prisoner's Dilemma framework. Furthermore, our experimental results show that people are reciprocal with both cooperation and defection, which is in line with the definition of reciprocity proposed by Fehr and Gachter [20].

In terms of the strategy, participants reciprocated more with the agents who used TfT. That seems natural considering that other studies have shown that TfT strategy is intrinsically a reciprocal strategy. Participants also cooperate more with the agents playing TfT. However, it must be considered that cooperative behavior is the most profitable strategy in single Prisoner's Dilemma but not in RPDG. Dawes pointed out that subjects contribute in the game because they have high expectations about the contributions of others [15]. Therefore the number of interactions is a factor that should be considered carefully in the design of reciprocal behaviors for companion robots.

In addition, TfT strategy increases the cooperations of the participants (m=5.733) compared with the Random strategy (m=4.233). TfT strategy encourages cooperation in the participants with an initial cooperation that can be perceived as a cooperative attitude. This strategy had an effect in the Profit and Joint Profit due to the number of cooperations and reciprocations. A higher number of cooperations reduces the loss of money per participant. A combination of cooperative behaviors in both participant and agent allows both to increase their own profits. Consequently a higher individual profit amounts to a higher Joint Profit. Participants tended to have a higher Joint Profit with a robot agent than with the human agent. However the participants profit was not significantly affected by the agent. The higher Joint Profit can be explained by the combination of agent-strategy in every stage of the experiment. In other words, participants would be guessing the strategy of the agent before seeing a pattern in the first round of games, and then they could define a stable strategy in the second round.

Also, we compared the number of cooperations using the Tit for Tat strategy with the results reported as the Flood-Dresher experiment in [46,57]. They reported that in 100 rounds of RPDG participants decided to collaborate in average 68% of the rounds. In our study participants cooperate with the robot agent in 48.3% of the rounds and with the human agent in 66.3% of the rounds. On the other hand, de Melo et al. reported in [39] that participants cooperate more with a virtual agent that shows moral emotions (66.28%, 12.57 of 25 rounds) rather than agent that doesn't shows any emotion (51.57%, 12.893 of 25 rounds). The agent used Random strategy in rounds 1 to 5 and TfT strategy in rounds 6 to 25. These results are very close to the results obtained in our study. This could be consistent to fact that participants perceive moral agents as more human-like as de Melo et al. reported. In our study robot agents didn't show any emotion and we trained human agents in order to reduce any emotional expression.

Besides, participants offer significantly less money in average in the Ultimatum game to the robot than to the human agent. Furthermore, according to our chi-square analysis participants made 50%-50% offers more infrequently to the robot than to the human agent. We expected that the offer in the Ultimatum Game would be affected by the strategy performed independently of the agent in the Prisoner's dilemma. Humans are known to typically reject offers that are 80%-20% [42] Thus players play safe most of the time, offering a 50%-50% offer to the agent. However, according to the final comments of some participants playing with the robot, they wanted to experiment with different offers just to see the reaction of the robot.

People perceived higher openness and agreeableness in the human agent. However the agent did not have a significant effect in the other personality traits. This can be explained by the personality of the actors playing human agents. Although we asked to the actors to keep themselves neutral and reduce the communication to minimal; we could not control the subtle body language and the gaze that could affect the perception of the participants.

When the agents played TfT strategy it was perceived as more extroverted and agreeable than when they played Random strategy. Probably participants perceived a subtle pattern playing TfT that they related with these two personality traits. If the agents started the game cooperating it is probable that people recognized that their agreeableness and extroversion related to a higher number of collaborations, reciprocations, profit and join profit.

Relationships between personality traits, agents and strategy can be useful as guidelines for the robot designers. Robot designers could make efforts in the design of robot behaviors and strategies matching with the personality of the users and triggering reciprocity in the user. We could say that under certain social situations extroverted people would tend to work in a better way with robots. Hirsh and Peterson have studied the influence of extroversion and neuroticism, personality traits in the Big Five test using the Prisoner's dilemma. They found that extroversion and neuroticism traits predict a greater likelihood of cooperation [28].

5.1 Contribution of our study

Results of our study suggest that reciprocity exists in Human-Robot Interaction under Prisoner's Dilemma scenario. Certainly Prisoner's Dilemma can be adapted to other social situations which involve interactions and decisions between different agents. This study helps us to understand the importance of the strategy used by the agent in order to receive a reciprocal treatment. The implications in the design of companion robots can be significant in terms that robot designers should consider that the behavior of their robots (independently of other variables as embodiment or anthropomorphism) must be aimed to follow a similar pattern as the Tit for Tat strategy. It is easy to imagine different scenarios in which this pattern could appear in HRI. For instance, companion robots in the role of an assistant could offer their services and then predict the actions of the users. If the user wants a companion, the robot would also show itself keen to offer companionship; if the user

rejects the presence of the robot then the robot would also indicate that it did not require the user. However, this raises questions about predictability, such as: What is the the threshold to be reciprocal with the user? Do humans expect some unpredictability in robots in order to maintain attention on them?

In general terms, we can explain our results with the media equation theory [41] and the natural identification of patterns. Humans tend to treat objects as other social actors; therefore, they tend to be similarly reciprocal with them. Furthermore, Turkle in [54] claims that actual users are focused on the outcomes of the experience rather than on the agent, and for the youngest people it does not matter if the player of a certain social activity is a robot or a sentient being if this agent reaches the goal to entertain or do something else for the users . Thus, we can consider that robots will receive a reciprocal treatment similar to what humans receive in scenarios similar to the Prisoners Dilemma and Ultimatum Game. However we can even raises the question Why do the participants actually reciprocate equally to humans and robots? Because they treat the robot as a human, or because they think that this is the most promising strategy. Certainly these questions should be require further study.

Aditionally, we can go back to the question: Do people reciprocate towards robots in a similar way to how they reciprocate with humans? We can say that if it were possible to situate Prisoner's Dilemma and Ultimatum Game in different social situations people would be reciprocal with robots. Although people tended to be less collaborative with robots than with humans in our experiment; reciprocation is similar. If robots show a cooperative behavior people would tend to respond in the same way, and would tend to respond with the same attitude. Of course, the social situations involving HRI are more complex than that. For instance, scenarios involving negotiation between robots and humans require the analysis of other variables.

Finally if we try to answer the hypothetical question of Kahn et al. of whether people can engage substantively in a reciprocal relationship with robots, we can say that it is possible if the robot first shows a reciprocal behavior toward humans like in Prisoner's Dilemma. Furthermore, we can discuss how companion robots can engage in a positive reciprocal relationship with humans if the companion robots have an efficient strategy like TfT. Robot designers should work on designing reciprocal strategies that increase the collaboration in HRI to the same level as in HHI. However more studies should be done in order to explain all the future social implications in the field. This studio should be a first step towards a better understating of the importance of reciprocity in the use of companion robots.

We consider that there will be many activities in which companion robots and humans would need to work cooperatively. However this cooperation could be closely related to reciprocal behavior. Although Broz and Lehnman claim that we would not feel any reciprocal feeling towards robots such as compassion [8], there are other studies that claim that people naturally tend to be reciprocal with machines (computers, mobile devices, cars) in terms that these objects offer a benefit to the user and the user takes care of them. Logically the user takes care of his/her objects to keep them working offering service, help or benefit to the user. Indeed, a critical future work is the development of companion robots capable of showing the proper actions, behaviors and social clues to encourage a reciprocal behavior in the users. As Breazel claims, the development of sociable robots involves interpretation of intentional and unintentional acts, subjectivity, (showing rudiments of intentional behavior), proto-dialogue, consistency and expressive characteristics of emotion in voice, face, gesture and posture [6]. Furthermore, Dautenhahn claims that social robots would be socially evocative, socially situated, sociable and socially intelligent [12]. All these robotic skills involve reciprocity.

5.2 Limitations and future work

As so often in HRI studies, the participants had only very limited previous experiences with robots. 56.7% (17 of 30) of the participants in the robot condition had never interacted with a robot before. This may have lead to a novelty effect that could have substantiated itself in a tendency of the participants to explore this new experience rather than focusing on winning the game.

Reciprocity is a very complex social phenomenon. As a future work we will study HRI scenarios in which it is not clear how the decisions are clearly taken; for instance scenarios involving bribery or unfair behaviors. Moreover, deeper studies should be conducted to explore whether reciprocal interactions generate more engaging interactions.

Acknowledgements The authors would like to acknowledge the support from NZi3 and NEC NZ Corp., in particular to Hamish House and Glen Cameron. We also want to thank the support of the UC International Doctoral Scholarship, CONACYT Scholarship and John Templeton Foundation (Award ID 36617). Additionally, we wish to thank the other members in the HIT Lab NZ: Mark Billinghurst, Jakub Zlowtoski and Anthony Poncet for their useful advice and good ideas. Thanks to Philippa Beckman for her extensive

proofreading. This experiment was approved by the Human Ethics Committee of the University of Canterbury under the reference 2013/23/LR-PS.

References

1. Andreoni, J., Miller, J.H.: Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. The Economic Journal **103**(418), 570–585 (1993). DOI 10.2307/2234532. URL http://www.jstor.org/stable/2234532
2. Axelrod, Robert: Effective choice in the prisoner's dilemma. Journal of Conflict Resolution **24**(1), 3–25 (1980)
3. Axelrod, R.: More effective choice in the prisoner's dilemma. Journal of Conflict Resolution **24**(3), 379–403 (1980). DOI 10.1177/002200278002400301. URL http://jcr.sagepub.com/cgi/doi/10.1177/002200278002400301
4. Axelrod, R.M.: The evolution of cooperation. Basic Books, New York (1984)
5. Boone, C., De Brabander, B., van Witteloostuijn, A.: The impact of personality on behavior in five prisoner's dilemma games. Journal of Economic Psychology **20**(3), 343–377 (1999). DOI 10.1016/S0167-4870(99)00012-4. URL http://www.sciencedirect.com/science/article/pii/S0167487099000124
6. Breazeal, C.L.: Designing sociable robots. Intelligent robots and autonomous agents. MIT Press, Cambridge, Mass.; London (2002)
7. Broadbent, E., Peri, K., Kerse, N., Jayawardena, C., Kuo, I., Datta, C., MacDonald, B.: Robots in older people's homes to improve medication adherence and quality of life: A randomised cross-over trial. In: M. Beetz, B. Johnston, M.A. Williams (eds.) Social Robotics, *Lecture Notes in Computer Science*, vol. 8755, pp. 64–73. Springer International Publishing (2014). DOI 10.1007/978-3-319-11973-1_7. URL http://dx.doi.org/10.1007/978-3-319-11973-1_7
8. Broz, F., Lehmann, H.: Do we need compassion in robots? In: A. Weiss, T. Lorenz, B. Robins, V. Everes, M. Vincze (eds.) International Conference of Social Robotics Proceedings, Taking Care of each Other: Synchronisaton and Reciprocity for Social Companion Robots, pp. 15–18. Springer International Publishing (2013). URL http://workshops.acin.tuwien.ac.at/ISCR2013/
9. Chaudhuri, A., Sopher, B., Strand, P.: Cooperation in social dilemmas, trust and reciprocity. Journal of Economic Psychology **23**(2), 231–249 (2002). DOI 10.1016/S0167-4870(02)00065-X. URL http://www.sciencedirect.com/science/article/pii/S016748700200065X
10. Cialdini, R.B.: Influence: science and practice, 3rd ed edn. HarperCollinsCollegePublishers, New York (1993)
11. Clark, M.L., Ayers, M.: Friendship Expectations and Friendship Evaluations: Reciprocity and Gender Effects. Youth & Society **24**(3), 299–313 (1993). DOI 10.1177/0044118X93024003003. URL http://yas.sagepub.com/cgi/doi/10.1177/0044118X93024003003
12. Dautenhahn, K.: Socially intelligent robots: dimensions of human-robot interaction. Philosophical Transactions of the Royal Society B: Biological Sciences **362**(1480), 679–704 (2007). DOI 10.1098/rstb.2006.2004. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2346526/
13. Dautenhahn, K., Woods, S., Kaouri, C., Walters, M., Werry, I.: What is a robot companion - friend, assistant or butler? 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems pp. 1192–1197 (2005). DOI 10.1109/IROS.2005.1545189. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1545189
14. Davis, W.: Strategies in iterated prisoners dilemma. (2013). URL http://www.iterated-prisoners-dilemma.net/prisoners-dilemma-strategies.shtml
15. Dawes, R.M., Messick, D.M.: Social dilemmas. International journal of psychology **35**(2), 111–116 (2000). URL http://www.tandfonline.com/doi/abs/10.1080/002075900399402
16. Draper, H., Sorell, T., Bedaf, S., Syrdal, D., Gutierrez-Ruiz, C., Duclos, A., Amirabdollahian, F.: Ethical dimensions of human-robot interactions in the care of older people: Insights from 21 focus groups convened in the uk, france and the netherlands. In: M. Beetz, B. Johnston, M.A. Williams (eds.) Social Robotics, *Lecture Notes in Computer Science*, vol. 8755, pp. 135–145. Springer International Publishing (2014). DOI 10.1007/978-3-319-11973-1_14. URL http://dx.doi.org/10.1007/978-3-319-11973-1_14
17. Falk, A., Fehr, E., Fischbacher, U.: On the nature of fair behavior. Economic Inquiry **41**(1), 20–26 (2003). URL http://onlinelibrary.wiley.com/doi/10.1093/ei/41.1.20/abstract
18. Falk, A., Fischbacher, U.: A theory of reciprocity. Games and Economic Behavior **54**(2), 293–315 (2006). DOI 10.1016/j.geb.2005.03.001. URL http://www.sciencedirect.com/science/article/pii/S0899825605000254
19. Fehr, E., Fischbacher, U.: The nature of human altruism. Nature **425**(6960), 785–791 (2003). DOI 10.1038/nature02043. URL http://www.nature.com/nature/journal/v425/n6960/full/nature02043.html
20. Fehr, E., Gaechter, S.: Reciprocity and economics: The economic implications of homo reciprocans. European Economic Review **42**(3–5), 845–859 (1998). DOI 10.1016/S0014-2921(97)00131-1. URL http://www.sciencedirect.com/science/article/pii/S0014292197001311
21. Fogg, B.J.: Persuasive technologies. Communications of the ACM **42**(5), 26–29 (1999)
22. Fogg, B.J.: Persuasive technology: using computers to change what we think and do. Ubiquity **2002**(December), 5 (2002). URL http://dl.acm.org/citation.cfm?id=763957
23. Fogg, B.J.: Persuasive computing: technologies designed to change attitudes and behaviors. Morgan Kaufmann; Elsevier Science, San Francisco, Calif.; Oxford (2003)
24. Fogg, B.J., Nass, C.: How Users Reciprocate to Computers: An Experiment That Demonstrates Behavior Change. In: CHI '97 Extended Abstracts on Human Factors in Computing Systems, CHI EA '97, pp. 331–332. ACM, New York, NY, USA (1997). DOI 10.1145/1120212.1120419. URL http://doi.acm.org/10.1145/1120212.1120419
25. Gosling, S.D., Rentfrow, P.J., Swann Jr., W.B.: A very brief measure of the big-five personality domains. Journal of Research in Personality **37**(6), 504–528 (2003). DOI 10.1016/S0092-6566(03)00046-1. URL http://www.sciencedirect.com/science/article/pii/S0092656603000461

26. Gouaillier, D., Hugel, V., Blazevic, P., Kilner, C., Monceaux, J., Lafourcade, P., Marnier, B., Serre, J., Maisonnier, B.: Mechatronic design of NAO humanoid. In: IEEE International Conference on Robotics and Automation, 2009. ICRA '09, pp. 769–774 (2009). DOI 10.1109/ROBOT.2009.5152516
27. Gouldner, A.W.: The norm of reciprocity: A preliminary statement. American Sociological Review **25**(2), 161–178 (1960). DOI 10.2307/2092623. URL http://www.jstor.org/stable/2092623
28. Hirsh, J.B., Peterson, J.B.: Extraversion, neuroticism, and the prisoner's dilemma. Personality and Individual Differences **46**(2), 254–256 (2009). DOI 10.1016/j.paid.2008.10.006. URL http://www.sciencedirect.com/science/article/pii/S0191886908003772
29. Kagel, J.H., Roth, A.E.: The handbook of experimental economics. Princeton University Press, Princeton, N.J. (1995)
30. Kahn, P., Ishiguro, H., Friedman, B., Kanda, T.: What is a human? - toward psychological benchmarks in the field of human-robot interaction. pp. 364–371 (2006). DOI 10.1109/ROMAN.2006.314461. URL http://dx.doi.org/10.1109/ROMAN.2006.314461
31. Kahn Jr., P.H., Friedman, B., Perez-Granados, D.R., Freier, N.G.: Robotic pets in the lives of preschool children. Interaction Studies **7**(3), 405–436 (2006). DOI 10.1075/is.7.3.13kah
32. Kiesler, S., Sproull, L., Waters, K.: A prisoner's dilemma experiment on cooperation with people and human-like computers. Journal of Personality and Social Psychology **70**(1), 47 – 65 (1996)
33. Kolm, S.C.: Chapter 6 reciprocity: Its scope, rationales, and consequences. In: Serge-Christophe Kolm and Jean Mercier Ythier (ed.) Handbook of the Economics of Giving, Altruism and Reciprocity, vol. Volume 1, pp. 371–541. Elsevier (2006). URL http://www.sciencedirect.com/science/article/pii/S1574071406010062
34. Kreps, D.M., Milgrom, P., Roberts, J., Wilson, R.: Rational cooperation in the finitely repeated prisoners' dilemma. Journal of Economic Theory **27**(2), 245–252 (1982). DOI 10.1016/0022-0531(82)90029-1. URL http://www.sciencedirect.com/science/article/pii/0022053182900291
35. Kunz, P.R.: Romantic Love and Reciprocity (1969). URL http://www.jstor.org.ezproxy.canterbury.ac.nz/stable/10.2307/582223?Search=yes&resultItemClick=true&searchText=romantic&searchText=love&searchText=and&searchText=reciprocity&searchUri=/action/doBasicSearch?Query=romantic+love+and+reciprocity&acc=on&wc=on&fc=off&
36. Lammer, L., Huber, A., Weiss, A., Vincze, M.: Mutual care: How older adults react when they should help their care robot. In: AISB2014: Proceedings of the 3rd international symposium on New Frontiers in Human-Robot interaction (2014)
37. Lin, R., Kraus, S.: Can automated agents proficiently negotiate with humans? Commun. ACM **53**(1), 78–88 (2010). DOI 10.1145/1629175.1629199. URL http://doi.acm.org/10.1145/1629175.1629199
38. Lorenz, T.: Synchrony and reciprocity for social companion robots: benefits and challenges. In: A. Weiss, T. Lorenz, B. Robins, V. Everes, M. Vincze (eds.) International Conference of Social Robotics Proceedings, Taking Care of each Other: Synchronisaton and Reciprocity for Social Companion Robots, pp. 10–14. Springer International Publishing (2013). URL http://workshops.acin.tuwien.ac.at/ISCR2013/
39. Melo, C.M.D., Zheng, L., Gratch, J.: Expression of Moral Emotions in Cooperating Agents *. Intelligent Virtual Agents (2009)
40. Muscolo, G.G., Recchiuto, C.T., Campatelli, G., Molfino, R.: A robotic social reciprocity in children with autism spectrum disorder. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 8239 LNAI, pp. 574–575 (2013). URL http://www.scopus.com/inward/record.url?eid=2-s2.0-84892383760&partnerID=tZOtx3y1
41. Nass, C., Reeves, B.: The media equation: how people treat computers, televisions, and new media like real people and places. CSLI Publications; Cambridge University Press, Stanford, Calif.: New York; Cambridge (1996)
42. Nishio, S., Ogawa, K., Kanakogi, Y., Itakura, S., Ishiguro, H.: Do robot appearance and speech affect people's attitude? evaluation through the ultimatum game. In: 2012 IEEE RO-MAN, pp. 809–814 (2012). DOI 10.1109/ROMAN.2012.6343851
43. Oda, R.: Biased face recognition in the prisoner's dilemma game. Evolution and Human Behavior **18**(5), 309–315 (1997). URL http://www.sciencedirect.com/science/article/pii/S1090513897000147
44. Park, H., Antonioni, D.: Personality, reciprocity, and strength of conflict resolution strategy. Journal of Research in Personality **41**(1), 110–125 (2007). DOI 10.1016/j.jrp.2006.03.003. URL http://www.sciencedirect.com/science/article/pii/S009265660600033X
45. Perugini, M., Gallucci, M., Presaghi, F., Ercolani, A.P.: The personal norm of reciprocity. European Journal of Personality **17**(4), 251–283 (2003). DOI 10.1002/per.474. URL http://onlinelibrary.wiley.com.ezproxy.canterbury.ac.nz/doi/10.1002/per.474/abstract
46. Poundstone, W.: Prisoner's Dilemma. Anchor (2011)
47. Rapoport, A.: Prisoner's dilemma: a study in conflict and cooperation. University of Michigan press, Ann Arbor, Mich
48. Sandoval, E.B., Brandstetter, J., Bartneck, C.: Can a robot bribe a human? the measurement of the negative side of reciprocity in human robot interaction. In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 117–124 (2016). DOI 10.1109/HRI.2016.7451742
49. Selten, R., Stoecker, R.: End behavior in sequences of finite prisoner's dilemma supergames a learning theory approach. Journal of Economic Behavior & Organization **7**(1), 47 – 70 (1986). DOI http://dx.doi.org/10.1016/0167-2681(86)90021-1. URL http://www.sciencedirect.com/science/article/pii/0167268186900211
50. Short, E., Hart, J., Vu, M., Scassellati, B.: No fair. an interaction with a cheating robot. In: Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on, pp. 219–226 (2010)
51. Sophister, L.D.J.: Public goods and the prisoners dilemma: Experimental evidence URL http://www.tcd.ie/Economics/SER%20New/Archive/2000/DILEMMA.PDF
52. Spaniel, W.: Game Theory 101: The Complete Textbook
53. Torta, E., Dijk, E., Ruijten, P., Cuijpers, R.: The Ultimatum Game as Measurement Tool for Anthropomorphism in Human–Robot Interaction. pp. 209–217. Springer International Publishing (2013). URL http://dx.doi.org/10.1007/978-3-319-02675-6_21
54. Turkle, S.: Alone together: why we expect more from technology and less from each other. Basic Books, New York (2011)

55. Weiss, A.: Grounding in human-robot interaction: Can it be achieved with the help of the user? In: A. Weiss, T. Lorenz, B. Robins, V. Everes, M. Vincze (eds.) International Conference of Social Robotics Proceedings, Taking Care of each Other: Synchronisaton and Reciprocity for Social Companion Robots, pp. 7–10. Springer International Publishing (2013). URL http://workshops.acin.tuwien.ac.at/ISCR2013/
56. Weiss, A., Lorenz, T.: Icsr 2013 workshop 3: Final report and results. taking care of each other:sincronization and reciprocity for social companion robots. In: A. Weiss, T. Lorenz, B. Robins, V. Everes, M. Vincze (eds.) International Conference of Social Robotics 2013 Proceedings, Taking Care of each Other: Synchronisaton and Reciprocity for Social Companion Robots, pp. 1–7. Springer International Publishing (2013). URL http://workshops.acin.tuwien.ac.at/ISCR2013/
57. Williams, K.C.: Introduction to game theory: a behavioral approach. Oxford University Press, New York (2013)

Eduardo B. Sandoval is currently PhD student at the Human Interface Technology Laboratory (HIT Lab NZ), which is part of the University of Canterbury, New Zealand. He is developing a model of reciprocity in Human Robot Interaction (HRI). This model could be applied in research related to influence, altruism and fair treatment of robots. Eduardo has a Master degree in Industrial Design by the National Autonomous University of Mexico (UNAM) and a Bachelor degree in Bionic Engineering by the National Polytechnic Institute of Mexico (IPN). He did an academic exchange in the Laboratory of Intelligent Robotics in Osaka University (2011). He was awarded by the UNAM with the Alfonso Caso medal in 2012.

Jürgen Brandstetter is currently PhD student at the HIT Lab NZ and works on humanoid robots and its influence on humans when using the oral language. He studied at the Vienna University of Technologies and Copenhagen University and holds a master in computer science. In his master thesis he looked at a new computer game type called "Positive Impact Game" and worked on a rehabilitation games for humans who need to learn how to work with an arm prosthesis. He is also a full stack developer, android developer, prototyping enthusiast, User Experience Designer, entrepreneur and design thinking tutor.

Dr. Mohammad Obaid is a postdoctoral fellow at the t2i Lab, Chalmers University of Technology, Sweden. He gained his MSc.(2007) with First Class Honours and PhD (2011) degrees from the Computer Science and Software Engineering Department, at the University of Canterbury. He started his career with a postdoctoral fellowship at both the Human Centered Multimedia Lab (Augsburg, Germany) and the HIT Lab NZ (2011-2014). He collaborated and worked at several international research institutes including CNRS at LTCI, Telecom ParisTech (Paris, France), Australian National University (Canberra, Australia), Institute for Computer Graphics and Vision at the Graz University of Technology (Graz, Austria), and the Digital Media department at the Upper Austria University of Applied Sciences (Hagenberg, Austria). His research interests fall in the areas of Human-Computer Interaction and Human-Robot Interaction.

Dr. Christoph Bartneck is an associate professor and director of postgraduate studies at the HIT Lab NZ of the University of Canterbury. He has a background in Industrial Design and Human-Computer Interaction, and his projects and studies have been published in leading journals, newspapers, and conferences. His interests lie in the fields of Social Robotics, Design Science, and Multimedia Applications. He has worked for several international organizations including the Technology Centre of Hannover (Germany), LEGO (Denmark), Eagle River Interactive (USA), Philips Research (Netherlands), ATR (Japan), Nara Institute of Science and Technology (Japan), and The Eindhoven University of Technology (Netherlands). Christoph is a member of the New Zealand Institute for Language Brain & Behavior, the IFIP Work Group 14.2 and ACM SIGCHI. He has an outstanding publication record, including leading journals and conferences. Christoph is serving as an associate editor for the International Journal of Social Robotics and the International Journal of Human Computer Studies. Furthermore, he is the editor in chief of the Entertainment Robotics section of the Entertainment Computing journal. He has been a member of the program committee of the Human-Robot Interaction conference for several years. He organized several workshops at the CHI conference and is frequently invited as a speaker for symposia and conferences. He has been invited to present his work by Carnegie Mellon University, the Future University Hakodate, the University of Venice, the Stedelijk Museum and the Pictopia Festival. The press regularly reports on his work, including the New York Times, New Scientist, Natuurwetenschap & Techniek, Volkskrant and Dutch national television and radio stations.