Exploring Prosociality in Human-Robot Teams

Filipa Correia*, Samuel F. Mascarenhas*, Samuel Gomes*, Patrícia Arriaga†,
Iolanda Leite‡, Rui Prada*, Francisco S. Melo* and Ana Paiva*
*INESC-ID, Instituto Superior Técnico, Universidade de Lisboa
†Instituto Universitário de Lisboa (ISCTE-IUL), CIS-IUL
‡KTH Royal Institute Technology

***Abstract*—This paper explores the role of prosocial behaviour when people team up with robots in a collaborative game that presents a social dilemma similar to a public goods game. An experiment was conducted with the proposed game in which each participant joined a team with a prosocial robot and a selfish robot. During 5 rounds of the game, each player chooses between contributing to the team goal (cooperate) or contributing to his individual goal (defect). The prosociality level of the robots only affects their strategies to play the game, as one always cooperates and the other always defects. We conducted a user study at the office of a large corporation with 70 participants where we manipulated the game result (winning or losing) in a between-subjects design. Results revealed two important considerations: (1) the prosocial robot was rated more positively in terms of its social attributes than the selfish robot, regardless of the game result; (2) the perception of competence, the responsibility attribution (blame/credit), and the preference for a future partner revealed significant differences only in the losing condition. These results yield important concerns for the creation of robotic partners, the understanding of group dynamics and, from a more general perspective, the promotion of a prosocial society.**

***Index Terms*—Groups, Social Dilemma, Public Goods Game, Prosocial, Selfish**

I. Introduction

A large part of what constitutes human activity is conducted by teams rather than individuals on their own. Considering our social nature as a species, perhaps it is not that surprising that we excel at working together with others and often prefer to do so. With the rapid advances that are being made in the field of robotics, the fear that robots will eventually replace entire teams of humans in certain activities is one that has recently risen in popularity [1], [2]. However, a more optimistic possibility is that teams that mix both humans and robots in a successful manner will outperform exclusively robotic teams. Moreover, it is also possible that people will come to enjoy having robotic partners to collaborate with, assuming that those robots not only lead to an increase in the team's productivity but also have adequate social skills (e.g., fairness [3]) and are capable of fostering a sense of group trust and identification [4]. Indeed, recent studies have already shown that some behaviours such as expressing vulnerability [5] or having group-based emotions [6] do create a positive impact on human-robot teams. It is important to further understand and explore how people perceive robots when performing a shared activity with them and how their behaviour and disposition is affected by such collaboration.

In this article, we introduce a novel contribution to the research on teamwork in Human-Robot Interaction (HRI) and to the new area of Prosocial Computing. We present an empirical analysis on single individuals forming a team with two other fully autonomous robots. The analysis focuses on understanding how people respond to a prosocial robotic partner that always sacrifices its individual gains in favour of the group compared to a selfish robotic partner that cares about maximising its individual performance within the group. This notion of prosocial machines has been recently discussed in [7], where social robots can act as prosocial agents that promote beneficial actions for others at the cost of one's own in order to create a prosocial society of humans and machines.

To conduct the aforementioned analysis we developed a novel collaborative task that consists of a turn-based digital game named *For The Record*. The name is inspired by the game's musical theme as it involves a group of players that form a band together and then try to create and sell successful records. Despite its theme, the core mechanics of the game are not related to musical skills. Instead, the game works as a variant of a public goods game, which is one of the standard games that are used for analysing prosocial collaboration in fields such as experimental economics [8]. Essentially, a public goods game models a social dilemma in which there is a tension between contributing to the benefit of the group or acting selfishly and taking advantage of the other people that are contributing. While the purely rational decision is to act selfishly and free-ride, if everyone in the group does so, there is no collaborative gain to be shared. Studies have shown that when people play this game they usually contribute to the group, although, in iterated sessions they tend to reduce their contributions after witnessing others defecting [9].

The social dilemma that is present in *For The Record* is one where players have essentially to decide between acting in a prosocial manner by increasing the chances that the band records a good album or acting in a selfish manner by increasing the chances of maximising their individual profit on the band's successful records. One key distinction when compared to a standard public goods game, is that the outcome of both options is uncertain as players have to roll a certain number of dice to determine the outcome of their actions.

The authors show their gratitude to EDP and, in particular, to Pedro Fernandes for establishing this collaboration. This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT-UID/CEC/50021/2019), through the project AMIGOS (PTDC/EEISII/7174/2014). Filipa Correia acknowledges an FCT grant (Ref. SFRH/BD/118031/2016).

For the purpose of conducting our study on human-robot teamwork, the dice outcomes in *For The Record* were secretly manipulated. The goal was to create two experimental groups that participants would be randomly assigned to. Both groups would have identical outcomes except in the last round. In one group, the result is positive and the team wins, whereas in the other group, the outcome is negative and consequentially the team loses. In both groups, players play with the same robotic partners, one uses a prosocial strategy and the other uses a selfish strategy. Due to the fact that the latter robot is constantly outperforming the former in the amount of profit made, this could have led to the perception that it was the most competent out of the two robots. The results did not support this conclusion. In fact, the prosocial robot was seen as significantly more competent in the losing condition and there was no significant difference between the two in the winning group. Moreover, as the prosocial robot played in a more collaborative manner, we expected participants to choose it as a future partner rather than the selfish robot. However, this preference was only significant in the losing condition.

Overall, the obtained results suggest that, in a collaborative context, the perception of competence is more associated to how a robot contributes to the group rather than its individual performance within the group. However, when the group succeeds, the competence of a selfish robot is perceived as being similar to the competence of a prosocial robot. Additionally, another important result of the study is that the success of the team had a significant positive effect on group identification but not in group trust. To better understand this result, we conducted a regression analysis and discovered that the reported discomfort towards the selfish robot was a significant predictive factor of group trust but not its perceived competence as suggested in [10].

Finally, rather than having participants from the academia, the study was conducted inside the offices of a large corporation in the energy sector. The employees who participated had little exposure to social robotics and we took this opportunity to ask their opinion on whether this type of robots can be a net good for society. Surprisingly, a significant difference was found between the losing and winning conditions.

II. Related Work

Mixed human-robot teams can vary quite substantially in the amount of autonomy that the robots possess. On one extreme, there are teleoperated robots who have little autonomy as their main purpose is to follow the instructions of a human controller [11]. On the other extreme, robots are fully autonomous with both shared and individual goals as is the case in the work presented here. As discussed in [4], this notion of robots as partners rather than tools requires that robots possess some level of social intelligence [12].

One of the first studies that investigated how people respond to robotic partners in a work-like setting was conducted by Hinds et al. [13]. In this study, participants had to move around a room and collect several objects with the assistance of either a human confederate, a human-like robot or a machine-like robot. The results showed that people were more likely to feel responsible for the task when they interacted with the machine-like robot that acted as a subordinate. Instead, when the robot acted as a supervisor, participants were more likely to blame it for any mistakes that occurred than when the robot acted as a peer or subordinate. In our study, we also conducted an analysis on how participants judged themselves and their robotic partners with regard to who was most responsible for the outcome of the collective task. However, rather than manipulating the perceived status of the robots and their appearance, we contrasted their decision-making strategy during the task.

One crucial factor that is needed to enable successful teamwork is a sense of group trust [14], [15]. This has led roboticists to explore and find different types of behaviours that robotic teammates can perform to increase how much people trust them. For instance, having a robot making vulnerable statements has been found to increase the amount of trust-related behaviours from its human partners. Such effect was discovered in a study conducted with a team of three individuals playing a collaborative digital game with a NAO robot [5]. In the study, the robot expresses vulnerability by admitting its mistakes to the group, which increases the amount of times that people will also admit to their mistakes. In a different study, group trust was shown to be higher when a robot expressed group-based emotions to its partner compared to a robot that expressed individual-based emotions instead [6]. Although the dyadic concept of trust towards a robot has been highly explored in the past years [10], the literature on trust towards a group of humans and robots is still scarce [16]. In our study, results provide new insightful considerations on the factors that impact the measure of group trust.

Social dilemmas have been a powerful tool to capture the social aspects of human behaviour and the altruism and prosociality of cooperation in Human-Computer Interaction and more recently, in HRI. More than 20 years ago, Kiesler et al. used a Prisoner's Dilemma to compare collaboration between a human and a computer with different human-like communicative channels. The results revealed that their best-liked computer partner was able to increase cooperation compared to other uncommunicative partners mentioned in the previous literature.

The later advances on social robotics led researchers to explore other influencing factors on the cooperation during Social Dilemmas. For instance, in the Ultimatum Game, participants reported higher rejection scores towards a computer opponent than towards a robot or a human opponent [17]. Another example by Terada & Takeuchi [18], inspired by the work of de Melo et al. on virtual agents [19], revealed that the display of emotions by a social robot can elicit altruistic behaviour in the Ultimatum Game. Additionally, Sandoval and collaborators investigated how humans perceive robots that apply different strategies [20]. They showed the Tit-For-Tat strategy is associated with the personality dimensions of extroversion and agreeableness. This finding exposes insightful concerns when designing behaviours and strategies for social

robots.

To the extent of our knowledge, Public Goods games are another type of Social Dilemmas that have not yet been explored in HRI. We believe this paper constitutes one of the first investigations to explore this inherently collaborative setting with important considerations for human-robot teams.

III. For The Record

For The Record is as a N-person threshold game with uncertain returns. In this game, there is a public good accessible by each team member independently of her contribution. The creation of such public good (their collective goal) requires that the sum of all contributions exceeds a threshold that is uncertain. Each player tries to maximise the collective goal by contributing to the public good. At the same time, individuals may opt to free ride on the efforts of others, while choosing to invest on their own individual goals. The game is set within an artistic context, in which players are musicians of a band. Even if framed within a specific context, this class of dilemmas is general enough to capture the non-linearity and uncertain nature of many Human collective endeavours, from group hunting to climate agreements. Introducing these type of social dilemmas in HRI, especially in group interactions, allows the analysis of prosocial collaboration and, in a more general perspective, the creation of new approaches in which robots can promote prosociality on humans.

The following description of *For The Record* considers the artistic context attributed to the game when introduced to the players. Each musician of the band has the goal of "maximising his/her revenue by contributing to the creation of successful albums and avoiding the collapse of the band".

The game is composed by R rounds and each round is the publication of an album on the market. Before detailing the stages of the album creation, consider that each player j has two distinct skills as a musician that are quantifiable in discrete levels: the musical instrument (li_j), and the marketing (lm_j). The instrument skill is used during the creation of an album, where each player j sequentially has to evaluate her individual performance by rolling li_j dice of 6 faces. Letting $D_f(n)$ denote the result of rolling n dice of f faces, the value of an album sums the value of each musician's performance, according to the following expression:

$$V_{album} = \sum_{i=1}^{N} D_6(li_j)$$

After creating each album, the market value determines whether that album succeeds or fails. The market value is calculated by rolling n dice of 20 faces. Additionally, *For The Record* includes two difficulty levels when publishing an album on the market, called national and international market that differ according to the following expressions:

$$V_{national\_market} = D_{20}(2)$$

$$V_{international\_market} = D_{20}(3)$$

Thereupon, each album is considered either a mega hit or a fail according to the following expression:

$$\begin{cases} \text{“}MegaHit\text{”} & \text{if} \quad V_{album} >= V_{market} \\ \text{“}Fail\text{”} & \text{if} \quad V_{album} < V_{market} \end{cases}$$

Each round ends with the players receiving their individual revenues. The revenue is 0 when the album has failed, however, in case of a mega hit, each player j has two options: to receive a default amount of 3000 or to use his/her marketing skill and receive according to the result of rolling lm_j dice of 6 faces. This second option is only available if $lm_j > 0$.

In the beginning of each round, each player has to upgrade one of his/her skills by 1 point, between the instrument and the marketing skill. On the one hand, by increasing the level of the instrument, the player can roll one more dice during the evaluation of his/her performance and, therefore, increases the likelihood of producing a successful album. On the other hand, by increasing the level of the marketing, the player can roll one more dice during the revenue collection in case of a mega hit and, therefore, increases the likelihood of maximising the individual profit. In other words, each player has to choose between to cooperate, by contributing to the collective goal, or to defect, by contributing to his/her individual goal.

Another important rule is: during the R rounds, if the band achieves a limit L of failed albums, the game ends and each musician loses all the accumulated revenue. This is done in order to stress the importance of collaborating.

IV. User Study

We conducted a user study using the previously described *For The Record* game. The number of players, N, was 3 and the selected setting was one human participant playing together with two robotic players on a touch screen (Figure 1). Furthermore, we set the number of rounds, R, to 5 and the limit of failed albums, L, to 3. The band started to publish albums on the national market and changed to the international market on the 4^{th} round. The initial values for the levels of each skill were the same for all the players: 1 point in the instrument skill ($li = 1$), and no points in the marketing skill ($lm = 0$). Finally, players could upgrade their skills from the 2^{nd} round on, which means they had 4 decisions to make during the 5 rounds between improving their instrument skill (cooperate) or their marketing skill (defect).

One particular factor that is likely to influence how people perceive their robotic teammates is whether the team succeeds or fails in the shared task. As identified in [21], people are more sensitive to avoiding losses than to gains of equal monetary amount. This well-known cognitive bias is referred to as loss aversion. As a result, in this user study, we manipulated the game result in a between-subjects design, which produced two experimental conditions: winning or losing the game. In order to achieve these two deterministic outcomes, we scripted predefined orders for all the possible dice throws. Not only could we guarantee all the participants played the same number of rounds, we could also ensure that the dice rolls of the each robot were the same in both conditions.

Fig. 1: This interaction was captured during a session of the user study, where a participant is playing *For The Record* game with the two robotic partners.

Consequently, during the 5 rounds, participants got a failure, a victory, a failure, a victory and finally either a victory or a failure according to the condition.

The robotic players differ on the strategies they apply to play the game. One of them always defects by improving its marketing skill in every round, which we call the defector, while the other always cooperates by improving its instrument skill in every round, which we call the cooperator. Nevertheless, their verbal and non-verbal behaviours remained similar and we used two versions of the same embodiment for each character, the EMYS robotic head [22]. Regarding their speech acts, they encourage the team in the beginning of each album, they comment extreme luck or bad luck on the dice rolls for both themselves and the other players and, in the end of each album, they comment the round result with an emotional animation of either sadness or joy. The three game states that were used to emphasise the difference between their distinct game strategies were:

- The level up phase where each robot chooses to upgrade either its instrument skill (cooperate) or its marketing skill (defect) (e.g., Cooperator – "I will level up the instrument.", Defector – "I will improve the marketing.");
- The dice roll that corresponds to the individual performance for an album (e.g., Cooperator – "Wow, I added [N] points!", Defector – "[N] more points for our album!");
- The last decision of using or not the marketing skill to receive the revenue in case of success (e.g., Cooperator – "Here it comes the reward.", Defector – "I will use my [N] marketing skill points to see what I can get...").

To avoid having these autonomous robots speaking at the same time, the game engine randomly chooses which robot comments each game state. Finally, their non-verbal behaviours consists of gazing at: the other players when it is their turn; the other robot if it is speaking; or the touch screen by default.

A. Hypotheses

The following hypotheses state our expectations towards the differences on people's perceptions, judgements and preferences between a prosocial and a selfish robotic partners after teaming with them in a public goods game.

H1: The prosocial robot will be perceived more positively in its social attributes than the selfish robot.
H2: The prosocial robot will be perceived as less competent than the selfish robot.
H3: Group trust and group identification will be positively associated with the group performance.
H4: When the team wins, the main responsible factor will be the strategy of the prosocial robot.
H5: When the team loses, the main responsible factor will be the strategy of the selfish robot.
H6: The prosocial robot will be preferred as a future partner, rather than the selfish robot.

Our rationale behind these hypotheses is the following. Concerning **H1** and **H6**, we expect that participants see the prosocial robot in a more positive light as it acts in a fully collaborative manner, helping the team to succeed. The expectation behind **H2** lies in the fact that the selfish robot will always be ahead in terms of task performance (profit made). Additionally, in a study that asked participants to judge the competence of people playing the prisoner's dilemma, the results showed that those who were defected against were seen as less competent [23]. It is possible that the same effect occurs in our study given that prosocial robot's behaviour. The reason for **H3** is the aforementioned loss aversion bias [21] and finally, **H4** and **H5** are based on the assumption that people will correctly identify the main responsible actor behind the team's result.

B. Procedure

Participation in this study was individual and started with a brief overview of each step. All the participants signed the consent form and then proceeded with the experiment. One researcher read the game rules one by one and answered participant's questions, while another researcher set up the robots and the video camera. Then, they played a training game without the robots to ensure the participant learned the game and to clarify any final doubts. The training game had a maximum of 5 rounds but the dice rolls were completely random. After that, the researcher initiated the game with the robots, alternating between conditions. Before the researchers left the room, they emphasised the goal of the study is to analyse their opinion of each robot and, therefore, they should pay attention to which is which and also to their behaviours during the game. Finally, after the interaction with the robots, each participant answered the questionnaire and was greeted by his/her participation.

C. Dependent Measures

The following dependent measures were used on the data analysis: **Competitiveness** level of the participant using a single-item question "How competitive do you evaluate yourself?"; **Group Identification** [24] using the Portuguese adaptation [25] with the dimensions of Solidarity, Satisfaction and In-Group Homogeneity; **Group Trust** [26]; **Robotic Social Attribute Scale (RoSAS)** [27] using its three dimensions of Warmth, Discomfort, and Competence towards each robot;

Choice of a Robotic Partner among the defector and the cooperator for a hypothetical future game; **Responsibility (blame/credit) attribution** of four different factors – randomness, participant's strategy, defector's strategy and cooperator's strategy – using single-item questions "The game result was mainly due to (...)". All the items in the questionnaire were assessed in a 7-points scale ranging from 1 ("Definitely not associated") to 7 ("Definitely associated") and the robots were always mentioned by their names.

D. Sample

The study was conducted at a company facility in order to collect a varied sample in terms of age, gender and background. There was a total of 70 participants (35 per experimental condition) with ages ranging from 22 to 63 ($M = 34.6, SD = 11.557$). Regarding gender, there were 32 females, 37 male, and 1 unknown.

V. Results

A. Social Attributes of the Robots

To analyse the impact of the game result on the perception of each robot, we used a Mixed analysis of variance (ANOVA) where the within-subjects factor is the robotic character and the between-subjects factor is the game result (winning or losing).

Regarding the perception of warmth, there was a significant main effect of the robotic character (Figure 2, $F(1, 67) = 17.366, p < 0.001, r = 0.454$) with the cooperator being rated with higher values of warmth ($M = 4.225, SD = 1.090$) compared to the defector ($M = 3.513, SD = 0.977$). However, the main effect of the game result and the interaction between the robotic character and the game result were not statistically significant ($F(1, 67) = 0.028, p = 0.869, r = 0.020$ and $F(1, 67) = 0.013, p = 0.908, r = 0.014$, respectively).

For the social attribute of discomfort, there was a main effect of the robotic character (Figure 2, $F(1, 67) = 30.982, p < 0.001, r = 0.562$), with the defector being rated with higher levels of discomfort ($M = 2.895, SD = 1.302$) than the cooperator ($M = 1.895, SD = 1.064$). Again, we have not found a significant main effect of the game result nor a significant interaction between the robotic character and the game result ($F(1, 67) = 0.525, p = 0.471, r = 0.088$ and $F(1, 67) = 1.141, p = 0.289, r = 0.129$, respectively).

These results suggest that the distinct strategies adopted by the robots affected the perception of the robot's warmth and the discomfort they felt regardless of the game result.

In terms of the perception of competence, there was a significant main effect of the robotic character ($F(1, 67) = 24.873, p < 0.001, r = 0.520$), with the the cooperator being rated with higher levels of competence ($M = 4.790, SD = 1.111$) than the defector ($M = 3.907, SD = 1.073$). Although we did not find a significant effect of the game result ($F(1, 67) = 0.966, p = 0.329, r = 0.119$), there was a significant interaction between the robotic character and the game result (Figure 3, $F(1, 67) = 4.095, p = 0.047, r = 0.240$). To understand this interaction, we compared the perception of

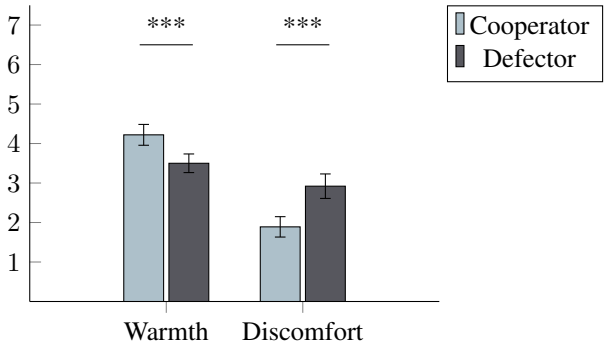

Fig. 2: Main effect of the robotic partner on the social attributes of warmth and discomfort.

competence attributed to each robot across the two possible game results using a Wilcoxon Signed-Rank test. In the case where the game result was winning, there was no significant difference between the competence attributed to each robot ($Z = -1.859, p = 0.063, r = -0.319$). However, in the case where the game result was losing, there was a significant difference between the competence attributed to each robot ($Z = -4.434, p < 0.001, r = -0.749$), with the cooperator being rated as more competent ($M = 4.876, SD = 0.958$) than the defector ($M = 3.624, SD = 0.896$).

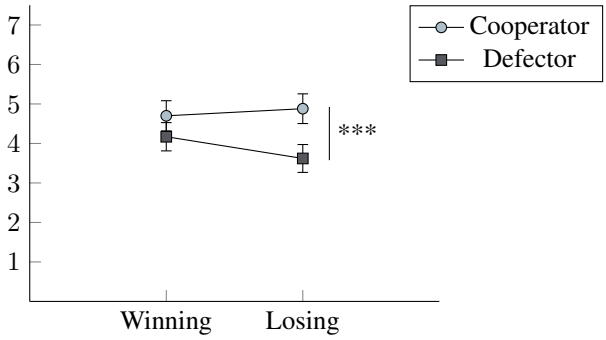

Fig. 3: Interaction effect between the robotic partner and game result on the attributed levels of competence.

Contrary to the previous social attributes, the competence attributed to each robot was affected by the game result. Participants have considered the cooperator as more competent only in the losing condition. This result suggests the negative effect of losing the game highlighted the difference in perceived competence between the robots.

B. Group Measures

To analyse the two dependent measures related to the group (Figure 4), i.e. group identification and group trust, between the two possible game results, we used Mann-Whitney U tests. Results showed a significant difference between the levels of group identification according to the condition ($U = 404.5, Z = -2.445, p = 0.014, r = -0.292$), with participants that won the game reporting higher levels of group identification ($M = 4.267, SD = 1.346$) than participants

who have lost the game ($M = 3.466, SD = 1.182$). Nevertheless, there was no significant difference between the levels of group trust according to the game result ($U = 535.5, Z = -0.715, p = 0.474, r = -0.086$).

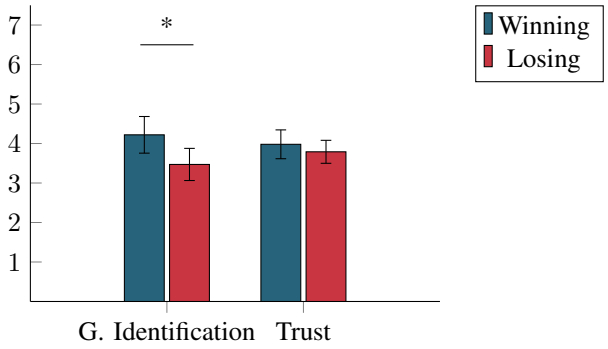

Fig. 4: Effect of the game result on the attributed levels of group identification and group trust.

These results revealed the group identification was affected by the game result, winning or losing the game, as we have predicted. However, the prediction about the measure of group trust was not confirmed, suggesting that other factors might have contributed to this outcome. Therefore, we conducted an additional analysis to interpret these surprising findings, by creating predictive models of both the group identification and the group trust levels. We used Stepwise regressions with the backward method to determine which variables could explain most of the variance of group identification and group trust levels. The initial seven predictor variables were the ones related with individual and group perceptions of the team members: defector's warmth, defector's competence, defector's discomfort, cooperator's warmth, cooperator's competence, cooperator's discomfort, and either group identification or group trust.

Regarding the group identification level, we found in the 5^{th} step that it can be significantly predicted ($F(3, 65) = 33.016, p < 0.001, R^2 = 0.604$) by - 1.652 + 0.843 (group trust) + 0.375 (defector's competence) + 0.158 (cooperator's competence), where variables are assessed with 7-points likert scales. Regarding the prediction of the group trust level, we found in the 6^{th} step that it can be significantly predicted ($F(2, 66) = 40.455, p < 0.001, R^2 = 0.551$) by 2.513 + 0.489 (group identification) - 0.174 (defector's discomfort), where variables are assessed with 7-points likert scales.

This exploratory analysis allowed us to understand that although there is a correlation between group identification and group trust, they were affected by other factors, after partialling out the shared explanatory effect of the other variables. Besides the strong relation of one another, group identification can also be predicted from the competence attributed to each of the team members, and group trust can also be predicted from the discomfort attributed to the defector.

C. Responsibility (Blame / Credit) Attribution

To analyse the responsibility attribution of the game result among the following four factors of (1) randomness, (2) participant's strategy, (3) defector's strategy and (4) cooperator's strategy, we used Friedman's ANOVA tests. In the winning condition (Figure 5), we found no significant differences on the credit attribution to the four factors ($\chi^2(3) = 7.142, p = 0.067, r = 0.070$). However, in the losing condition (Figure 6), the blame attribution was significantly different to the four possible factors ($\chi^2(3) = 33.264, p < 0.001, r = 0.326$).

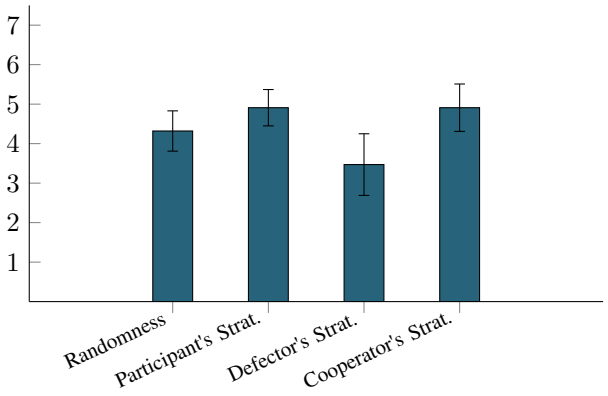

Fig. 5: Responsibility attributed to each factor in winning condition (credit).

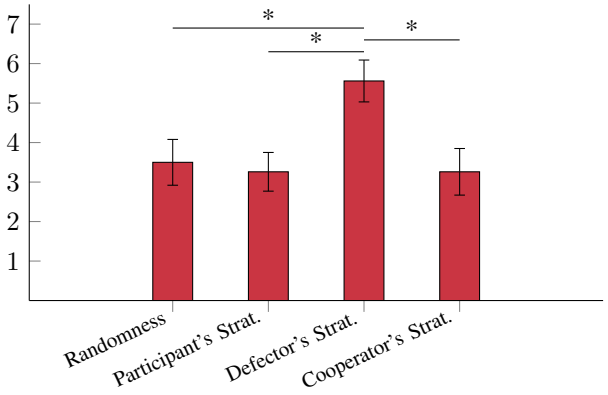

Fig. 6: Responsibility attributed to each factor in losing condition (blame).

To follow up this finding on the attribution of blame, we conducted a *post hoc* analysis using Wilcoxon Ranks tests. Moreover, we applied a Bonferroni correction and all the effects are reported at a 0.008 level of significance. It appeared that all the pairwise comparisons involving the defector's strategy were significant. In the losing condition, participants attributed higher levels of blame to the defector's strategy ($M = 5.429, SD = 1.685$) when compared to the randomness factor ($M = 3.543, SD = 1.669; Z = -3.421, p < 0.001, r = -0.578$), to the participant's strategy ($M = 3.265, SD = 1.377; Z = -4.586, p < 0.001, r = -0.786$), and to the cooperator's strategy ($M = 2.743, SD =$

$1.669; Z = -3.909, p < 0.001, r = -0.661$). Regarding the remaining pairwise comparisons, there was no significant difference between levels of blame attributed to the randomness factor and to the participant's strategy ($Z = -0.745, p = 0.456, r = -0.128$), nor between the randomness factor and the cooperator's strategy ($Z = -2.201, p = 0.028, r = -0.372$), nor between the participant's strategy and the cooperator's strategy ($Z = -1.284, p = 0.199, r = -0.220$). These results reveal that there was no clear main responsible factor in the credit attribution of the winning outcome. However, participants clearly identified the defector's strategy as the main cause of the losing outcome.

D. Choice of a Robotic Partner

To analyse the choice of a robotic partner among the defector and the cooperator for a hypothetical future game, we used a Chi-Square Goodness-of-Fit test. Results indicated a significant difference in the preference for a robotic partner ($\chi^2(1) = 22.857, p < 0.001, r = 0.326$), with the cooperator being preferred (55 times, 78.6%) to the defector (15 times, 21.4%).

Additionally, we found a significant association between the preferred robot and the game result ($\chi^2(1) = 14.339, p < 0.001, \phi_c = 0.453$) and we have, therefore, also analysed preferences across conditions (Figure 7). In the losing condition, there was again a significant difference ($\chi^2(1) = 31.114, p < 0.001, r = 0.889$), with the cooperator being preferred (34 times, 97.1%) to the defector (1 time, 2.9%). However, in the winning condition, no significant difference was found ($\chi^2(1) = 1.400, p = 0.237, r = 0.040$), with the cooperator being chosen 21 times (60.0%) and the defector 14 times (40.0%).

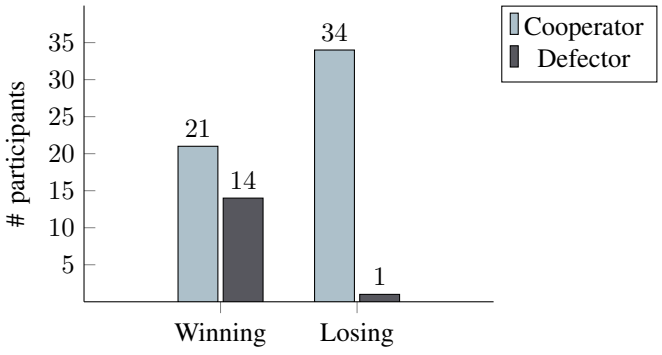

Fig. 7: Preferences for each robotic partner grouped by conditions.

When breaking down the choices of the participants across conditions, their preference is only clear when they lost the game. These results suggest that the negative impact of losing the game enhances the selfishness of the defector.

E. Strategy analysis

We analysed the playing strategies of the participants by looking at the number of times they have defected among their 4 decisions during the game. There were 8 participants that have never defected (11.4%), 21 that have defected once (30%), 33 that have defected twice (47.1%), 7 that have defected 3 times (10%), and only 1 that always defected (1.4%). Furthermore, out of the 33 that defected 2 times, 24 of them chose the strategy of "cooperate, defect, cooperate, and defect". Additionally, we found a weak positive correlation between the self-reported competitiveness level of the participants and the number of times they have defected ($r(70) = 0.235, p = 0.05$).

Finally, we did a correlation analysis to understand if the participants' perceptions of the robotic partners were associated with their competitiveness level or their playing strategy. In the winning condition, we found a moderate negative correlation between the rate of cooperation and the perceived impact of the defector's strategy ($r = -0.379, n = 34, p = 0.027$) and a moderate positive correlation between the rate of cooperation and the perceived impact of the self strategy ($r = 0.426, n = 35, p = 0.011$). No similar significant correlations were found in the losing condition. This suggests participants that cooperated more with team attributed more credit to their own strategy and less credit to the defector's strategy.

F. Societal Impact

Due to the diversity of our sample, we asked participants, at the end of the questionnaire, their agreement level on the sentence "Social robots will be relevant to the society", ranging between 1 ("Totally disagree") and 7 ("Totally agree"). Interestingly, we found a significant difference on their answers between conditions ($U = 435, Z = -2.143, p = 0.032, r = -0.256$), revealing a higher acceptance of social robots when they won the game ($M = 5.457, SD = 1.651$), compared to when they lost the game ($M = 4.686, SD = 1.676$).

VI. Discussion

According to **H1**, we have predicted that the prosocial robot would be perceived more positively than the selfish robot. We validated this hypothesis as the cooperator was rated as warmer and caused less discomfort. Our results suggest that the display of a prosocial strategy by the robotic partner enhanced the perception of its social attributes.

We have also predicted in **H2** that the selfish robot would be perceived as more competent, which was not confirmed. In fact, the opposite result was found, although only in the losing condition. This hypothesis was based on the fact that the defector uses the optimal strategy of maximising its profit on the efforts of the others, commonly called the free rider. One possible explanation is that participants construed the notion of competence as one that necessitates the absence of exploitation of others and, therefore, even though selfish acts are highly profitable, they are deemed as incompetent. It is also the case that, in the long run with multiple iterations of the game being played, the higher return obtained by a selfish strategy will diminish when considering the results obtained for the future partner choice in the losing condition. Another possible contributing factor is that participants were highly sensitive

to the risk involved in the uncertainty threshold of this game. Consequently, when participants lost the game, the evidence of a risky strategy became blameworthy and unreasonable.

Our results partially support **H3** as group identification was indeed positively associated with the performance of the group, although the same association was not verified for group trust. This surprising difference led us to analyse more carefully which factors were predicting both measures. According to our regression analysis, the best predictors of group identification were the group trust and the competence of each team member. Considering the discussion about H2, the competence attributed to the defector was significantly different across conditions. This can be the reason why there was also a significant difference on the levels of group identification.

On the other hand, the regression analysis for the group trust revealed that its best predictors were group identification (as they were highly correlated) and the discomfort attributed only to the defector. As the discomfort attributed to the defector remained similar in the two conditions, it seems to have strongly influenced the level of trust to follow the same pattern. Interestingly, literature on human-robot trust has previously suggested that performance is one of the most influencing factors to develop trust [10], which only occurred for group identification rather than for group trust.

Our results do not support **H4**, which predicted that, when the team wins, the main responsible factor would be the strategy used by the prosocial robot. There was no main responsible factor on the credit attribution of the winning outcome. Only 8 participants (11%) used the same prosocial strategy of cooperating 4 times and most participants defected at least 2 times (58.5%). Although most participants were more selfish than the prosocial robot, they attributed credit similarly between their own strategy and prosocial strategy.

According to **H5**, we have predicted that when the team loses, the main responsible factor would be considered the strategy of the selfish robot. Our results supported this hypothesis as the blame attribution to selfish robot were significantly higher than all the other 3 factors: randomness, the strategy of the participant, and the strategy of the prosocial robot.

Finally, **H6** hypothesised that the prosocial robot would be preferred as a future partner, which was only partially verified from our results. The preference for the prosocial robot was only clear in the losing condition. It seems that their preferences of a future partner were aligned with the responsibility attributions they mentioned and their perceptions of competence. The negative impact of losing the game might have stressed participants' judgements, which was denoted by significant differences on this choice.

VII. Conclusions and Future Work

We are moving towards a society in which robots are increasingly present and able to work with us. In this paper, we explored the role of prosociality as a contributing factor to establish cohesive collaborations with robots.

We conducted a user study where each participant formed a team with two autonomous robots to play a public goods game. In this type of social dilemmas, players have essentially to decide between acting in a prosocial manner by opting for the collaborative goal (cooperate) or acting in a selfish manner by choosing the individual goal (defect). The two robotic players used opposite strategies during the game: the selfish robot always defected while the prosocial always cooperated. Moreover, we manipulated the outcome of the game to either result in winning or losing.

Results showed that a prosocial partner can be perceived more positively in terms of its social attributes regardless of the game result, which generally reveals the importance of group-oriented decisions by social robots. Additionally, the differences between the participants' perception of competence, responsibility attribution and preferred robot were only significant when the participants lost the game. In particular, the portrayal of selfish behaviours by a robotic partner was negatively identified only when the performance of the team was compromised. More broadly, loosing outcomes seem to increase the people's awareness of what decisions players took throughout the game, and what impacts such decisions have for the success of the group.

This paper also shed some light on the development of trust and group identification towards mixed human-robot teams. In fact, many authors working on this topic have focused on trust, given that it is a critical element for group collaboration. Interestingly, in this study we found that the success of the team produced an increase in group identification but not in group trust. This has a broad implication that suggests these two measures can vary independently of one another. Furthermore, we provided some evidence on which social attributes of a robotic team member play a role on the levels of trust and group identification. These findings contribute not only to the understanding of these measures, but also to enhance human-robot collaboration.

An important consideration of our user study was the fact it took place at the facility of a large company and, therefore, our sample is more balanced in terms of ages and backgrounds than the most commonly reported samples that consist of young adults from universities [28].

As future work, it would be interesting to analyse the influence of the embodiment on the current findings, by replicating this user study with non-embodied agents. We would also like to explore the ingroup/outgroup relations of humans and robots, similar to what was done in [29], [30], by for instance changing the proportion between the number of human and robotic team players. Another aspect we are keen to work on is the impact of using different game strategies that are neither purely prosocial nor purely selfish. Finally, it would also be interesting to analyse the inclusion of additional social mechanisms such as punishments. This could be done either by (1) approaching the notion of altruistic punishment or (2) implying punishment in the agents' social behaviours, similar to [31].

REFERENCES

[1] C. Dirican, "The impacts of robotics, artificial intelligence on business and economics," *Procedia-Social and Behavioral Sciences*, vol. 195, pp. 564–573, 2015.

[2] M. Decker, M. Fischer, and I. Ott, "Service robotics and human labor: A first technology assessment of substitution and cooperation," *Robotics and Autonomous Systems*, vol. 87, pp. 348–354, 2017.

[3] F. P. Santos, J. M. Pacheco, A. Paiva, and F. C. Santos, "Evolution of collective fairness in hybrid populations of humans and agents," in *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*. AAAI Press, 2019.

[4] G. Hoffman and C. Breazeal, "Collaboration in Human-Robot Teams," *AIAA 1st Intelligent Systems Technical Conference*, pp. 1–18, 2004. [Online]. Available: http://arc.aiaa.org/doi/10.2514/6.2004-6434

[5] S. Strohkorb Sebo, M. Traeger, M. Jung, and B. Scassellati, "The Ripple Effects of Vulnerability," *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18*, pp. 178–186, 2018. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3171221.3171275

[6] F. Correia, S. Mascarenhas, R. Prada, F. S. Melo, and A. Paiva, "Group-based emotions in teams of humans and robots," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2018, pp. 261–269.

[7] A. Paiva, F. P. Santos, and F. C. Santos, "Engineering pro-sociality with autonomous agents," in *AAAI*, 2018.

[8] U. Fischbacher, S. Gächter, and E. Fehr, "Are people conditionally cooperative? evidence from a public goods experiment," *Economics letters*, vol. 71, no. 3, pp. 397–404, 2001.

[9] J. Andreoni, "Why free ride?" *Journal of Public Economics*, vol. 37, no. 3, pp. 291–304, 1988.

[10] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human Factors*, vol. 53, no. 5, pp. 517–527, 2011.

[11] B. Stoll, S. Reig, L. He, I. Kaplan, M. F. Jung, and S. R. Fussell, "Wait, can you move the robot?: Examining telepresence robot use in collaborative teams," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2018, pp. 14–22.

[12] K. Dautenhahn, "Socially intelligent robots: dimensions of human–robot interaction," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 362, no. 1480, pp. 679–704, 2007.

[13] P. J. Hinds, T. L. Roberts, and H. Jones, "Whose job is it anyway? a study of human-robot interaction in a collaborative task," *Human-Computer Interaction*, vol. 19, no. 1, pp. 151–181, 2004.

[14] G. R. Jones and J. M. George, "The experience and evolution of trust: Implications for cooperation and teamwork," *Academy of Management Review*, vol. 23, no. 3, pp. 531–546, 1998.

[15] L. M. Ma, T. Fong, M. J. Micire, Y. K. Kim, and K. Feigh, "Human-Robot Teaming: Concepts and Components for Design," *Field and Service Robotics*, pp. 649–663, 2018.

[16] A. Freedy, E. DeVisser, G. Weltman, and N. Coeyman, "Measurement of trust in human-robot collaboration," in *Collaborative Technologies and Systems, 2007. CTS 2007. International Symposium on*. IEEE, 2007, pp. 106–114.

[17] E. Torta, E. van Dijk, P. A. Ruijten, and R. H. Cuijpers, "The ultimatum game as measurement tool for anthropomorphism in human–robot interaction," in *International Conference on Social Robotics*. Springer, 2013, pp. 209–217.

[18] K. Terada and C. Takeuchi, "Emotional expression in simple line drawings of a robot's face leads to higher offers in the ultimatum game," *Frontiers in Psychology*, vol. 8, no. MAY, pp. 1–9, 2017.

[19] C. M. de Melo, P. Carnevale, and J. Gratch, "The effect of expression of anger and happiness in computer agents on negotiations with humans," in *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*. International Foundation for Autonomous Agents and Multiagent Systems, 2011, pp. 937–944.

[20] E. B. Sandoval, J. Brandstetter, M. Obaid, and C. Bartneck, "Reciprocity in Human-Robot Interaction: A Quantitative Approach Through the Prisoner's Dilemma and the Ultimatum Game," *International Journal of Social Robotics*, vol. 8, no. 2, pp. 303–317, 2016.

[21] D. Kahneman and A. Tversky, "Choices, values, and frames," in *Handbook of the Fundamentals of Financial Decision Making: Part I*. World Scientific, 2013, pp. 269–278.

[22] J. Kędzierski, R. Muszyński, C. Zoll, A. Oleksy, and M. Frontkiewicz, "Emys—emotive head of a social robot," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 237–249, 2013.

[23] J. I. Krueger and M. Acevedo, "Perceptions of self and other in the prisoner's dilemma: Outcome bias and evidential reasoning," *The American journal of psychology*, pp. 593–618, 2007.

[24] C. W. Leach, M. Van Zomeren, S. Zebel, M. L. Vliek, S. F. Pennekamp, B. Doosje, J. W. Ouwerkerk, and R. Spears, "Group-level self-definition and self-investment: a hierarchical (multicomponent) model of in-group identification." *Journal of personality and social psychology*, vol. 95, no. 1, p. 144, 2008.

[25] M. R. Ramos and H. Alves, "Adaptação de uma escala multidimensional de identificação para português," *Psicologia*, vol. 25, no. 2, pp. 23–38, 2011.

[26] K. Allen and R. Bergin, "Exploring trust, group satisfaction, and performance in geographically dispersed and co-located university technology commercialization teams," in *In Proceedings of the NCIIA 8th Annual Meeting: Education that Works*, 2004, pp. 18–20.

[27] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The robotic social attributes scale (rosas): development and validation," in *ACM/IEEE Int. Conf. on Human-Robot Interaction*, 2017.

[28] P. Baxter, J. Kennedy, E. Senft, S. Lemaignan, and T. Belpaeme, "From characterising three years of hri to methodology and reporting recommendations," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press, 2016, pp. 391–398.

[29] W.-L. Chang, J. P. White, J. Park, A. Holm, and S. Šabanović, "The effect of group size on people's attitudes and cooperative behaviors toward robots in interactive gameplay," in *RO-MAN, 2012 IEEE*. IEEE, 2012, pp. 845–850.

[30] M. R. Fraune, S. Sabanovic, and E. R. Smith, "Teammates first: Favoring ingroup robots over outgroup humans," in *RO-MAN 2017. The 26th IEEE International Symposium on Robot and Human Interactive Communication, Submitted*, 2017.

[31] A.-L. Vollmer, R. Read, D. Trippas, and T. Belpaeme, "Children conform, adults resist: A robot group induced peer pressure on normative social conformity." American Association for the Advancement of Science, 2018.