TÉCNICO LISBOA

UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR TÉCNICO

Group Intelligence in Social Robots

Filipa Isabel Nogueira Correia

Supervisor: **Doctor Ana Maria Severino de Almeida e Paiva**

Co-Supervisor: **Doctor Francisco António Chaves Saraiva de Melo**

Thesis approved in public session to obtain the PhD Degree in

Computer Science and Engineering

Jury final classification: Pass with Distinction and Honour

2021

UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR TÉCNICO

Group Intelligence in Social Robots

Filipa Isabel Nogueira Correia

Supervisor: **Doctor Ana Maria Severino de Almeida e Paiva**

Co-Supervisor: **Doctor Francisco António Chaves Saraiva de Melo**

Thesis approved in public session to obtain the PhD Degree in

Computer Science and Engineering

Jury final classification: Pass with Distinction and Honour

Jury

Chairperson: Doctor José Carlos Alves Pereira Monteiro, Instituto Superior Técnico, Universidade de Lisboa

Members of the Committee:

Doctor Ana Maria Severino de Almeida e Paiva, Instituto Superior Técnico, Universidade de Lisboa

Doctor Malte Friedrich Jung, The Ann S. Bowers College of Computing and Information Science, Cornell University, EUA

Doctor Selma Šabanović, Luddy School of Informatics, Computer and Engineering, Indiana University Bloomington, EUA

Doctor Hugo Miguel Aleixo Albuquerque Nicolau, Instituto Superior Técnico, Universidade de Lisboa

Funding Institution

Fundação para a Ciência e a Tecnologia (FCT) - Portugal , grant SFRH/BD/118031/2016

2021

Abstract

Humans have a natural and biological tendency to live and organise themselves in groups, from families to workplaces and organisations. That tendency not only mirrors the way society is distributed but also shapes the interactions with technology in general. As a result, the field of Human-Robot Interaction has recently started to look into group interactions with robots. This type of interactions refers to multi-party settings that extend traditional dyads of one person and one robot to cases where there are multiple people and/or multiple robots. This PhD thesis lies in the intersection of human-robot teamwork and human-robot group interactions by exploring the challenges of creating social robots that sustain cohesive alliances with humans in multi-party team settings.

We start by developing two autonomous robotic characters that display different traits from the goal orientation theory. We validate the two robotic characters, and we study the impact of these traits on membership preferences and team formation in a multi-party team game. Using the same multi-party team game, we also extend an emotional agent architecture with an additional component to perform a self-categorisation process and, in turn, to allow the robotic agent to display group-based emotions. We evaluate the proposed model in a user study where two autonomous robots partnered with two humans to play the game in two competing teams, one of the robots displaying group-based emotions and the other individual-based emotions.

In a different game, portraying a collective risk dilemma in a team of three, we explore the impact of group-based decisions (to cooperate) and individual-based decisions (to defect) by robotic teammates. We run several user studies where two autonomous robots (one was a cooperator and the other was a defector) play the game with one human. We further analyse the impact of the game outcome (win vs. lose), and of the robot embodiment (embodied vs. disembodied), among other factors.

Finally, we propose two possible approaches for robots to perceive the gaze behaviours of their teammates in multi-party settings. In a silent coordination task between two humans and one robot, we analyse the perception of teamwork when the autonomous robot uses each of the two proposed approaches.

The contributions are threefold. First, we contribute with computational mechanisms to create autonomous social behaviour in multi-party settings, by proposing novel techniques to perceive, express and consider group patterns. Second, we introduce new approaches to evaluate people's behaviours and perceptions towards a mixed human-robot group. Third, this thesis sheds light on how cohesion can be fostered into a strong alliance between human(s) and robot(s), which constitutes a step forward

into the understanding of how mixed groups of humans and robots are established.

Keywords: Human-Robot Interaction, Human-Robot Teamwork, Group Interaction, Group Dynamics, Cohesion

Resumo

Os humanos têm uma tendência natural e biológica para viverem e se organizarem em grupos, que não só se reflecte na forma como a sociedade está distribuida, mas também molda as interacções com a tecnologia de uma forma geral. Como resultado disso, o campo que estuda Interações Humano-Robô começou recentemente a olhar para interacções de grupo com robôs. Este tipo de interacções diz respeito a configurações de grupo que estendem as tradicionais díades entre um humano e um robô para casos onde existem vários humanos e/ou vários robôs. Esta tese de doutoramento aborda os desafios de criar robôs sociais capazes de manterem alianças coesas em situações de grupo com humanos, e situa-se, portanto, na intersecção de trabalho de equipas humano-robô e interacções de grupo.

Começamos por devenvolver duas personagens robóticas autónomas que demonstram traços diferentes segundo a teoria da orientação dos objectivos. Validamos as duas peronagens robóticas, e estudamos o impacto desses traços nas preferências de grupo e formação de equipas num jogo de equipas em grupo. Usando o mesmo jogo de equipas em grupo, estendemos também uma arquitectura para agentes emocionais com um componente adicional que realiza o processo de auto-categorização e que, por sua vez, permite ao agente robótico expresar emoções de grupo. Avaliamos o modelo proposto num estudo com utilizadores em que dois robôs autónomos formam equipa com dois humanos para jogar o jogo em equipas diferentes, um dos robôs expressa emoções de grupo e o outro expressa apenas emoções individuais.

Num jogo diferente, que mapeia um dilema de risco colectivo numa equipa de três, exploramos o impacto de decisões focadas nos objectivos collectivos (cooperar) e nos objectivos individuais (trair) por parte de colaboradores robóticos. Corremos vários estudos com utilizadores em que dois robôs autónomos (um cooperador e outro traidor) jogam o jogo com um humano. Analisamos o impacto do resultado do jogo (ganhar vs. perder), e do corpo do robô (presente vs. ausente), entre outros factores.

Finalmente, propomos duas abordagens possíveis para os robôs percepcionarem o olhar dos seus colaboradores em cenários de grupo. Numa tarefa de coordenação silenciosa entre humanos e robôs, analisamos a percepção do trabalho de equipa quando o robô autónomo utiliza cada uma das duas abordagens.

Destacamos três contribuições. Em primeiro lugar, contribuímos com mecanismos computacionais para criar comportamento social autónomo em cenários de grupo, propondo técnicas novas para percepcionar, expressar e considerar padrões de grupo. Em segundo lugar, introduzimos abordagens novas para avaliar o comportamento humanos e as suas percepções em grupos mistos de humanos e

robôs. Em terceiro lugar, esta tese revela de que forma a coesão pode ser fomentada numa aliança forte entre humano(s) e robô(s), o que contribui para a compreensão de como é que equipas mistas de humanos e robôs se estabelecem.

Palavras-chave: Interacção Humano-Robô; Equipas Humano-Robô; Interacções de Grupo; Dinâmicas de Grupo; Coesão

Acknowledgements

In the past five years, while trying to accomplish and finish the PhD, I struggled with four different issues and I would like to take this opportunity to acknowledge the people and entities that helped me getting through them.

Firstly, there was an economic issue of paying the tuition, the costs related with doing research (e.g., computer, traveling, material), and, of course, living costs. Regarding those issues, I acknowledge the support of Fundação para a Ciência e Tecnologia (FCT-Portugal) through several projects and mainly the fellowship SFRH/BD/118031/2016. Additionally, I would like to thank GAIPS Lab for awarding me with a monetary bonus every year (due to my effort and dedication), which had a crucial impact on my economic situation. This is not negligible, considering the land boom that happened in the metropolitan area of Lisbon.

Secondly, the bureaucratic issues together with the rush of getting things done and published was supported by the help of Sandra Sá and Vanda Fidalgo, who I would like to acknowledge and thank.

Thirdly, growing as a young researcher involved learning from and watching many successful researchers around me. To overcome all the implications of this issue, I relied on my amazing supervisors, Ana Paiva and Francisco Melo, as well as the following people (sorted chronologically): Iolanda Leite, Patrícia Alves-Oliveira, Sofia Petisca, Tiago Ribeiro, Rui Silva, Fernando Santos, Samuel Mascarenhas, Diogo Rato, Kim Baraka, Miguel Vasco, Diogo Carvalho, and Joana Campos.

Last, but definitely not least, there are the emotional and psychological issues of finishing a PhD. Not only while trying to find a good balance in life, but also while having doubts and uncertainties. I am extremely thankful to the following people (sorted alphabetically): Beatriz Mano, Filomena Correia & Fernando Correia, Francisco Correia, and Nuno Maia.

All the people I have mentioned will receive personal letters with private acknowledgements.

Lisbon, April 2021

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Humans have a natural and biological tendency to live and organise themselves in groups, from families to workplaces and organisations. That tendency not only mirrors the way society is distributed but also shapes the interactions with technology in general [137]. Consequently, as robotic systems become part of our lives in various domains or environments, e.g., domestic [30], industrial [69], public spaces [79; 86], education [18], health [22], they are naturally expected to engage in group settings or even integrate teams along with humans. In those situations, *how can the social capabilities of a robot consider aspects related to the group setting?*

The field of HRI has recently been paying close attention to group interactions, also known as multi-party interactions [84]. Such settings extend dyads of one person and one robot, to cases where there are multiple people and/or multiple robots. Studying multi-party settings or group interactions between humans and robots opens a wide range of challenges according to the nature and structure of those groups. The possible configurations may require robots to (1) act in groups, (2) be part of groups alongside humans, or (3) be close to groups of humans. In the first case, multiple robots form a group to perform a task together and the challenges in those situations may include, for instance, exploring coordinated behaviour between robots [3; 61]. In the second case, the robots act collaboratively alongside humans with similar goals, which is also known as human-robot mixed groups. Their main challenge in those situations is how to enhance collaboration, which can be achieved, for instance, with signaling behaviours [82] or by supporting telepresence [140]. Lastly, a single robot may also be required to act near a group of humans, without being part of that group. In those cases, the robot can act as a mediator or a facilitator while having a strong impact on the interaction among humans [81].

Overall, in any multi-party setting, robots are required to perceive group patterns that might not occur in human-robot dyadic interactions [127]. Addressing the previously described research challenges contributes to successfully integrate robots in a society that is organised in groups. Robots will, therefore, be required to act in groups, to be part of groups alongside humans, or to be close to groups of humans. Jung and collaborators have proposed three broad research avenues to advance the study of group interactions in HRI [84]. Firstly, understanding how the dynamics of existing group settings are shaped by a robot. In other words, analysing the impact robots have on the group processes, e.g., conflict or power. Secondly, understanding how the actual human-human interaction is influenced by robots. Finally, the last research avenue matches the focus of this dissertation, which is how can robots mediate or enhance the effectiveness or performance of work groups or teams.

1.1 Group Dynamics

The field of HRI grows hand-in-hand with insights from behavioural sciences. Considering the scope of this thesis on group interaction between humans and robots, we have looked into the processes that occur within and between interpersonal groups. The scientific field that studies those processes is called *Group Dynamics* and it is being explored for the past century by several disciplines within the social sciences, e.g., organisational behaviour, social psychology, or management.

The social psychologist Forsyth defined a group as *"two or more individuals who are connected by and* *within social relationships"* [56]. He also pointed out the five main characteristics of groups—interaction, goals, interdependence, structure and cohesion—and the fact that each one of them provides complex and fascinating phenomena when analysed more closely. According to these key characteristics, a group interacts or sustains a relationship in order to achieve a certain purpose or outcome. It is usually structured in a way that its members have interdependent roles and relations. Lastly, a group remains together due to their cohesive alliance, or cohesion.

There is, however, a particular type of groups that intensifies these five characteristics to an extreme level—teams. Members of a team are usually committed to a common goal in which the individual success is only a consequence of the collective success. There is a strong interdependence of the members' efforts that requires a coordinated interaction. In terms of structure, teams usually present highly adaptive skills that allow a revision of their norms and procedures in order to improve their functioning. Finally, the cohesive alliance of a team is also a salient characteristic, as Forsyth defines teams as *"unified, cohesive groups"*.

1.2 Research Problem

Teams are a particularly strong type of groups, in the sense that members of a team experience intense relations towards their shared goal. Nevertheless, as teamwork and collaboration only require a team to have at least two entities, most HRI literature in this topic is focused on dyadic teamwork [74; 49; 78; 27; 131; 73]. Few examples explored teamwork and collaboration within groups of humans and robots (i.e., groups with more than one person and/or more than one robot) [82; 62; 122; 145].

Although robots can easily outperform humans when individually executing specific tasks [24; 31], achieving satisfactory and effective collaborations might not be straightforward [16]. As soon as robots start working close to humans or are required to collaborate with them, their social capabilities become not only useful but also necessary. Dautenhahn et al. extensively revisited the definition of socially embedded agents and their tight relation to the social environment [42]. The authors further proposed these agents can "benefit from possessing some degree of awareness of the social system, i.e., a means of perceiving and sensing structures within the social world, and means of acting within it". This thesis contributes to this notion of socially embeddedness, specifically in relation to the group dynamics that can occur in a social environment where the robot is required to be a team member.

An immediate question that might occur is: what exactly is *the effectiveness* of a mixed human-robot team, and how can it be measured? There might be several different ways of assessing effectiveness, depending on the purpose or the goals of the team. One possibility is maximising the performance achieved by the team, by looking at the final outcome. There are, however, other possibilities such as examining other aspects or factors that sustain the quality of the team and may, in turn, mediate performance [43]. One of those factors is certainly cohesion, pointed by many authors as the most important aspect of groups and, consequently, of teams. *Cohesion* is the integrity, solidarity and unity that maintains a group together or in a *cohesive alliance*. It is associated with satisfaction, performance, and productivity. For instance, Mullen and Cooper found an interesting bidirectional

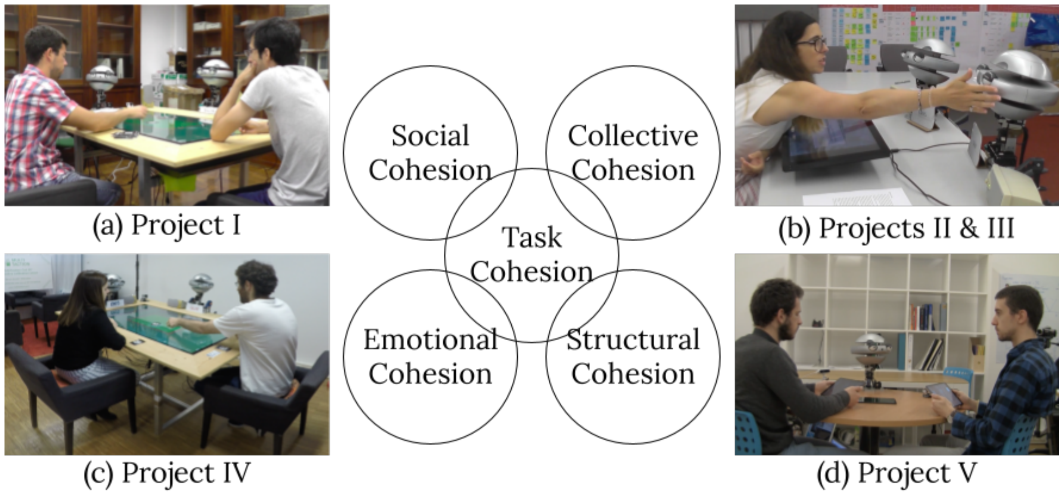


Figure 1.1: Diagram of the research plan to address the multi-dimensionality of the cohesion construct.

relationship between cohesion and performance, which suggests that not only cohesive teams tend to perform better, but also the success of the team leads to increased cohesion [104].

As we are particularly interested in this type of satisfactory and effective coalitions in mixed teams of humans and robots, an initial consideration is to understand of how humans establish cohesive alliances in their interpersonal groups or teams. The multi-dimensionality of the *cohesion* construct suggests that there are five possible courses that lead to a *cohesive alliance*: the task, the social bonds, the collective entity, the structure and emotions. *Task cohesion* emerges from the shared commitment towards a common goal, while *social cohesion* develops from the attractions among the members. The identification of the members with the group and the degree of belonging constitutes the *collective cohesion*. The *structural cohesion* derives from the norms, roles and relationships that link the members of the group. Finally, *emotional cohesion* is the intensity of the members to express group feelings. Considering the multi-dimensionality of this construct, we define a cohesive alliance between humans and robots as *a coalition in which the relation between group members, both humans and robots, emerges from at least one of the dimensions of cohesion and results in a shared sense of unity by all group members*. This thesis thus addresses the following research problem:

How can we endow a robotic teammate with social capabilities to improve the cohesive alliance in a multi-party setting with humans?

As a result, the approach to address our research problem was to tackle the dimensions of cohesion along five different projects, as Figure 1.1 shows. The first project, detailed in **Chapter 3**, focuses on membership preferences and team formation, which are related to aspects of social cohesion. The second and third projects explore how can robotic teammates portray different levels of collective cohesion, and how to do so through different degrees of embodiment, these projects are detailed in **Chapter 4** and **Chapter 5**. The fourth project, which is described in **Chapter 6**, focuses on the robot's expression of group-based emotions, which is related to emotional cohesion. The last and fifth project, which is presented in **Chapter 7**, explores different perceptual capabilities in robotic teammates that correspond to different structural team configurations. Additionally, task cohesion is also inherently present across all the projects, as we have chosen scenarios where humans are asked to form a team with robots.

1.3 Contributions

The five projects set three major contributions:

- C1 **Computational mechanisms to create autonomous social behaviour in multi-party settings** - By exploring how autonomous robots reason with a "shared sense of unity", we contribute with novel mechanisms to perceive, express or consider group patterns in their autonomous decision-making processes.

- C2 **Evaluation of people's behaviours and perceptions towards a mixed human-robot group** - By looking at how humans express a "shared sense of unity" in a mixed group with robots, we contribute with novel approaches to understand people's expectations, attitudes and predispositions towards robotic teammates in multi-party settings. Specifically, we pioneered in the assessment of some objective and subjective measures in HRI.

- C3 **Understanding relevant group processes in mixed-groups of humans and robots** - By exploring some well known theories from social sciences, we make a step forward into the understanding of how mixed-groups of humans and robots are established. In particular, we address aspects related to cohesion and show its relevance in human-robot multi-party teams.

The importance of these three contributions becomes clearer at a higher level due to the seamless relation among them. The interactive loop of designing, developing and evaluating robotic teammates for multi-party settings is mapped into C3, C1 and C2, respectively. Overall, we believe that reaching group intelligence in social robots can be strongly supported by the combination of these three contributions.

1.4 Roadmap

The current document contains 9 chapters and is organised as follows. In **Chapter 2**, we present the state-of-the-art on group interactions between humans and robots. The following five chapters, **Chapter 3-7** contain the core contributions of this thesis. Each of these chapters ends with a brief description of follow-up projects that were beyond the scope of this thesis, although related to multi-party or group settings. Finally, the document ends with a discussion and concluding remarks on the main contributions of this thesis, in **Chapter 8**, and with a summary of the achieved scientific publications.

Chapter 2

Related Work

In this chapter, we review the state-of-the-art of human-robot group interactions. Firstly, Section 2.1 covers papers that explored group phenomena in HRI. Secondly, Section 2.2 contains a significant scope of literature that analysed group identity or social categorisation, mostly involving mixed groups of humans and robots. Then, Section 2.3 includes papers that examined the human perceptions of robots in groups. Section 2.4 considered papers exploring the generation of group behaviour by social robots. Finally, Section 2.5 contributes with a general analysis of the literature on human-robot group interactions.

2.1 Group Phenomena in HRI

Within the literature of human-robot group interactions, we identified a line of research that tries to verify if well-known group phenomena from the social sciences also applies to HRI settings. Therefore, the commonality among the following user studies is a comparison between groups of individuals and groups of humans interacting with robot(s).

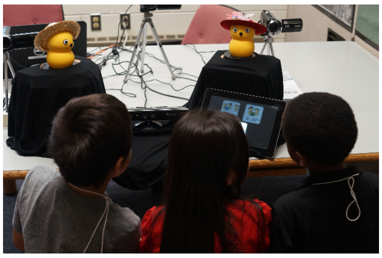
Chang et al. examined if the discontinuity effect translates into human-robot interactions [28]. This effect suggests that interactions among groups of humans are more competitive than interactions among individuals. The user study (see Figure 2.1a) manipulated the number of people in one team (either 1 or 2) and the number of robots in the adversarial team (also either 1 or 2). The interaction consisted of a board game with occasional social dilemmas between the teams to measure competitiveness. The results do not fully support the discontinuity effect in HRI.

In a similar experiment, Fraune et al. have also analysed the discontinuity effect [64]. Participants, individually or in a group of three, engaged in a social dilemma against a either a single robot or a group of three robots. The goal of this user study was to analyse the perceptions of the participants, as well as their competitiveness, in each of four experimental conditions. Contrary to human-human interaction, where increasing the number of members in at least one group would increase competitiveness, the results of this user study suggest a different pattern in human-robot interaction. In the two conditions where both teams had the same size, individual vs. individual and group vs. group, the competitiveness was higher. Additional results revealed that participants that played in groups of three had more negative emotions and competed more with the robot(s), compared to participants playing individually. Moreover, they found a positive correlation between the entitativity of the human group and the competitiveness, and also between the entitativity of the robot group and the reported fear.

Another interesting group phenomenon is conformity, which occurs in group interactions when a member changes his opinion or action in order to increase the consistency with other group members. This phenomenon illustrates how group processes may go beyond individual processes and how complex group interactions can be. Brandstetter et al. investigated if robots were also able to produce conformity behaviours in humans [20]. In their experiment, a group of 5 people was compared to a group of one person and four robots. The task was a linguistic quiz to identify past tenses. Although participants exhibited conformity with human peers, they did not display with the robots.

Solomons et al. have also explored conformity effect within a group of a human and three robots

(a) Chang et al. (2012) (b) Solomons et al. (2018)



(c) Leite et al. (2015) (d) Fraune et al. (2019)

Figure 2.1: Examples of literature analysing group phenomena in HRI.

playing the Dixit game [122; 121]. In their modified version of the game, a fourth robot was always the game master and would say a new word every round for the players to guess the corresponding card (see Figure 2.1b). There were two conditions: one where the participant could see which card was voted by the other robots and change his original decision; and a control condition in which the participant could not see the decisions of the robotic players. The results suggested that people have indeed conformed with the robots' responses in the experimental group. Additionally, the authors discussed a possible explanation for the incongruent result with Brandstetter et al. related to trust. They postulated trust plays an important role to perform conformity behaviours as robot started to choose incorrect answers and participants stopped conforming later in the game.

Interpersonal groups can outperform individuals in a variety of tasks and can, sometimes, even display higher learning gains. In order to explore this idea, Leite et al. have examined a learning activity between children and robots in an interactive storytelling scenario [96]. In particular, they have compared a single child with a group of three children in terms of story recall and the emotional interpretation of the story content (see Figure 2.1c). The results revealed that participants interacting alone presented higher levels of recall, compared to participants in the group condition. The authors speculated some possible reasons, for instance, group interaction may require children to work on their social standing, which may compromise their attention.

Finally, a recent in-the-wild experiment revealed several human factors affecting human-robot interactions, including the number of people within the group [58]. In this user study, Robovie was deployed in a Japonese shopping mall with the purpose of giving directions to people (see Figure 2.1d). Video recordings of 2714 participants and survey responses of 78 participants were collected and extensively analysed. The results showed that people interacting with the robot in groups, especially the entitative groups, (1) interacted more and for longer periods, (2) behaved more socially towards the robot, and (3) were more positive, compared to people interacting alone. Additionally, people in groups displayed compliance with social norms of the group, for instance, when someone in the group interacted with the robot, the others were likely to interact as well.

Overall, some of the well-known group phenomena from the social sciences partially apply to HRI. More investigation is needed to fully understand the particularities of these group interactions.

2.2 Group Identity, Membership and Social Categorisation

Group identity is one of the most fundamental aspects of a group as it defines its formation and membership. In other words, it is responsible for the categorisation of individuals as part of a larger identity or group. Group identification is also associated with many group processes and, therefore, researchers interested in human-robot group interaction have been playing close attention to it.

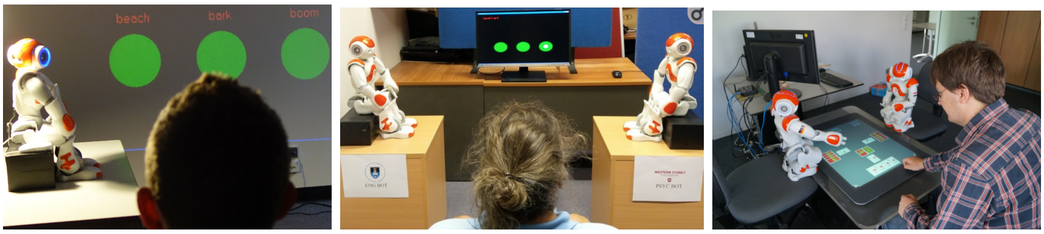
Eyssel & Kuchenbrandt pioneered the exploration of social categorisation in HRI. They manipulated social categorisation with the background information of the robot, i.e., nationality and name [55]. Participants reported their perceptions of the robot after seeing a picture of it. The salience of that group membership was enough for participants to rate the ingroup robot more favourably and to anthropomorphise it more.

In a similar experiment, Kuchenbrandt et al. manipulated group identity by simply associating colours to two groups, the minimal-group paradigm, and the robot would either be from the ingroup or the outgroup [93]. Participants were greeted with a short introduction by the robot and then rated their perceptions of it. Participants that interacted with the ingroup robot attributed significantly more anthropomorphism to it and even revealed higher willingness to interact with robots in general, compared to the ones that interacted with the outgroup robot.

One of the experiments presented by Deligianis et al. also explored the minimal group paradigm by simply priming participants about a presumed "robot condition" or a "computer condition" [44]. Participants had to play the shell game in a team with a robot that occasionally disagreed with the participant's answer to measure trust or compliance with that suggestion (see Figure 2.2a). Moreover, the difficulty of the game was manipulated in two levels, medium and hard. The results confirmed their first hypothesis stating that participants would trust the robot more often when the game difficulty was harder. Regarding the manipulation of the ingroup, it influenced the proxemics between the participants and the robot after the game, but no significant differences were found in trust to comply with the robot's suggestions.

The minimal group paradigm has not been always successful to induce the group membership. Sembroski et al. run a user study with a medical diagnosis task in which the robot would either be presented as a teammate (ingroup) or as a provider of additional information (outgroup) [129]. The goal of this user study was to analyse participants' willingness to follow the robot's instructions. Nevertheless, the manipulation check of group membership did not reveal significant differences between these two experimental groups suggesting the nature of the task might have conditioned this perception.

Recently, Steain et al. presented the results of a user study where they have analysed social



(a) Deligianis et al. (2017) (b) Steain et al. (2019) (c) Häring et al. (2014)

Figure 2.2: Examples of literature analysing group identity of social categorisation.

categorisation in the shell game [138]. The manipulation of identity was achieved by introducing the robots as “EngBot” and “PsycBot” to a population of students from psychology (see Figure 2.2b). The results of the experiment were focused on the perceptions of the robots and the degree of compliance with the answers of each robot i.e., trust. Although no significant differences were found for trust, participants rated the ingroup robot more favourably, kept a closer distance and preferred it more compared to the outgroup robot.

The identification towards a certain group can emerge from social aspects, as well as task-related aspects. As a result, several group identities can be present at the same time. In light of this idea, Haring et al. explored both task structure i.e., robotic partner or competitor, and the social membership, manipulated by the nationality of the robots [71]. In their user study, participants had to play a card game with two robots, where one of them was their teammate and the other was their opponent (see Figure 2.2c). Participants were told each robot was developed by a team of either their own nationality (ingroup) or another one (outgroup) and the names of the robots would also match their alleged nationality. The two experimental conditions mapped the congruency between social membership of the robots and their role in the game: ingroup partner and outgroup opponent; and outgroup partner and ingroup opponent. Results identified a main effect of the social membership on the perception of competence and on the cooperation index, with participants rating the ingroup robot as more competent and cooperated more with it compared to the outgroup robot. A surprising result of this user study was the fact that participants reported higher closeness to the ingroup robot when it was their partner compared to when it was their opponent but no similar difference occurred for the outgroup robot.

Finally, Fraune et al. presented the results of a user study whose goal was to compare the membership towards humans and robots. In their experiment, two teams of two people and two robots each had to competitively play a price-guessing game [62]. The main goal of this user study was to compare the perception and behaviour towards the ingroup and outgroup members, which was inherently set by the competitive setting. Additionally, it explored differences between robotic and human members. The results in terms of anthropomorphism and perceptions of cooperation support a strong impact of the ingroup setting compared to the outgroup, regardless of being a robot or a human. However, there was a surprising interaction effect on the aggression measure which was a noise blast that participants could use to damage partners or opponents during the game. Participants favoured the ingroup human over the ingroup robots, but they also favoured ingroup robots over outgroup humans. To a certain extent, this result indicates that people can favour robots over humans in some competitive settings.

Overall, the manipulations of group identity or the simple social categorisations were achieved by either the minimal group paradigm or by the structure of the group, i.e., people forming teams with the robots. Most results suggest that people can have strong ingroups with robots.

2.3 Perceptions of Groups of Robots

Robots may be required to perform or execute tasks in groups without human intervention or collaboration. Nevertheless, they will certainly be deployed in social settings and be surrounded by humans, whose perceptions and expectations might constrain their behaviour. As a result, this section overviews research on how people react to and perceive groups of robots.

Admoni et al. analysed different conformity levels in groups of robots and how they impact the perceptions of humans [3]. Participants saw videos of robots performing a dance routine where one of them was dancing independently while the others were coordinated and synchronised. They manipulated the size of the group to have either four or eight robots in order to increase the sense of majority and minority (3+1 and 7+1). Additionally, there was a control condition with only two robots with opposing dances. The dance type was also manipulated so that the minority robot could be dancing unique and distinct dance moves, similar moves but in a difference order, following the exact same moves but behind time, or leading the exact same moves ahead of time. Results showed the minority robot was consistently rated as less of a team-player and more anti-social, but more creative as well. This team-player perception was exacerbated when the group size increased. However, the differences across the four dance types were not statistically significant.

Behavioural mimicry was also analysed with other two factors, appearance and eye gaze, by Nawroj and collaborators [105]. Participants of the user study saw videos of three robots performing a dancing routine (robots A, B and C) and were only asked to rate the groupiness of robot C. The mimicry variable was manipulated for the robot C to either mimic A, B or none. Similarly, appearance and gaze were also manipulated for the robot C to either look like and gaze at A, B or none. As a result, there were 27 different videos of the robots dancing and each participant saw and rate nine of them. The results highlighted a strong main effect of behavioural mimicry, even when it interacted with conflicting cues from the appearance and gaze. Additionally, there was also a main effect of the appearance variable while mutual-gaze was not a significant predictor of grouping patterns.

Fraune & Šabanović explored the perceptions of people towards a group of three robotic vacuum cleaners [61]. The user study manipulated in a between-subjects design the communication style among the robots, which could either be silent, loud, or no communication at all (see Figure 2.3a). The results revealed no significant differences between the attitudes towards robots, as well as on the perceived social attributes o the robots. Moreover, in the manipulation check questions, participants could not identify differences in the communication style of the robots. The authors speculate that the non-humanoid embodiment of the Roomba as well as the low sociability of the robots might have

(a) Fraune et al. (2014) (b) Fraune et al. (2015)

Figure 2.3: Examples of literature analysing robots in groups.

strongly influenced the lack of differences.

Later, the same authors and collaborators investigated the role of embodiment when comparing a single robot with a group of robots [63]. They run a user study in which videos of the robots were shown to people and their perceptions were assessed in subjective questionnaires. In a between-subjects design, the video displayed either a single robot or a group, which could in turn have a humanoid embodiment (NAO), zoomorphic (Pleo) or mechanomorphic (Roomba). The results revealed an interaction effect between the two variables, group size and embodiment type. Groups of Roomba robots increased the negative responses, such as threat, anxiety or fear, compared to a single Roomba robot. On the other hand, the groups of NAO robots were positively rated in terms of elicited affect, perceived threat and trust, compared to a single NAO robot. Finally, the comparison between a single Pleo and the group of Pleo robots did not reveal differences in affective states but rather in the description traits (i.e., stereotypes) and future work contexts.

In their demand to understand the factors influencing the perceptions of robots in groups, Fraune and collaborators conducted another experiment where sociable trash robot entered the a public space, the cafeteria of the university [59]. They assessed behavioural and subjective measures of participants interacting with the robots and manipulated the number of robots, either one or a group of three, and different robotic behaviours, social or functional (see Figure 2.3b). The social behaviours consisted of contingent behaviour towards participants, such as greeting with a bow or a nod and a nonverbal "thank you" for throwing trash. A main effect of the number of robots revealed that a group of robots induced more direct interaction and more trash being thrown. Similarly, the main effect of the behaviour also shown that functional behavioural elicited longer gazes by people and more trash being thrown. Overall people preferred single social robots and groups of functional robots.

Finally, the last contribution of Fraune and collaborators to this line of research includes an experiment on the impact of group entativity [60]. Participants engaged in an object-matching task and the robot, or group of robots (depending on the condition), would assert their ability to perform the task. There were three conditions: a single robot, an entitative group and a diverse group. Entitativity was manipulated by the appearance, motion, proximity and decisions of the robots. The entitative group was perceived as more threatening than both the diverse group and the single robots. Moreover, the entitative group was also perceived more negatively when compared to the single robots only. On the other hand, participants attributed more mind to the diverse group than single robots.

Overall, the perceptions people have of robots acting in groups can be affected by several factors, such as their embodiment, motion or even mimicry level. More investigation is needed to understand other factors influencing these perceptions as well as how can they influence human behaviour.

2.4 Robotic Group Behaviour

Providing autonomy to social robots in group interactions can be a challenging task as it may include designing and developing challenges. The findings of the following reviewed papers contribute to the scope of literature that understands how robots can perceive and act in group settings with humans.

Leite et al. have contributed to the enhancement of robotic perceptions in group interactions [97]. The authors collected data of children playing an interactive storytelling activity with two robots in two distinct conditions: one where children participated alone and another where children participated in groups of three. Then, they developed two predictive models of individual disengagement, one trained with data from single interaction condition and another from group interaction condition. The annotations of engagement were done by human coders. Their analysis is focused on the evaluation of the two predictive models in both datasets in order to investigate how well the model trained on single interaction performs in group interaction and vice-versa. The disengagement model trained from group interactions performed reasonably well in the single interaction dataset, while the opposite was not verified.

Jung et al. explored backchanneling behaviour in human-robot teamwork [82]. They run a user study where a team of 5 members (3 robots, 1 participant and 1 human confederate) had to retrieve items from a building that has allegedly collapsed (see Figure 2.4a). In a between-subjects design, they manipulated both the presence of backchanneling behaviour and the task complexity. Results showed that when robots employed the backchanneling behaviour, they were perceived as more engaged to the task and their human partners reported lower stress levels as well as lower cognitive load, especially for the complex task. However, the backchanneling behaviour decreased the perception of competence of the robots. Overall, it seems that backchanneling can positively influence human-robot teamwork at a certain cost, and it is mostly beneficial in complex or demanding task.

Another sort of nonverbal behaviour was also explored by Vázquez et al. that examined how a robot is perceived during a group conversation according to its body orientation and gaze direction [148]. They run a user study in which groups of three participants had to brainstorm about possible tasks or purposes for the robot (see Figure 2.4b). Participants were briefed and debriefed by the robot and during the group discussion, it pretended to be listening and paying attention to conversation. Both the orientation and gaze of robot were manipulated in a between-subjects design. Orientation could be towards the middle of the conversation group or towards the participant that had the conversational floor, while gaze could be random or also attentive towards the participant that had the conversational floor. Although participants in the four experimental groups reported similar feelings of inclusion and belonging to the group, the gaze affected participants' perception of the robot motion, as well as the orientation affected the perception of its gaze. These results suggest gaze and orientation should be jointly designed and controlled.

(a) Jung et al. (2013)

(b) Vazquez et al. (2017)

(c) Shen et al. (2018)

(d) Tennent et al. (2019)

(e) Jung et al. (2018)

(f) Sebo et al. (2020)

Figure 2.4: Examples of literature designing robotic group behaviour.

Conflict is believed to be an important stage of group development [56] and the resolution of conflict is associated with several measures of group performance [80]. Recently, Jung et al. have looked at how the emotional behaviour of a robot can have a positive impact on improving teamwork by employing emotion regulation strategies to diffuse conflict situations [83]. The results suggested that although the robot's repair interventions have increased the groups' awareness of conflict, they can aid the conflict regulating process of the team.

Similarly, Shen et al. analysed the impact of a robotic mediator during conflict resolution among children [132]. In their experiment, pairs of children played five activities during 50 minutes that were facilitated by the robot (see Figure 2.4c). In a between-subjects design, they manipulated whether the robot would or not display an additional conflict mediation when children engaged in object possession conflicts. The procedure for the robot to intervene consisted of three steps: (1) play a whistle sound and identify the conflict; (2) offer prompts for constructive conflict resolution; and (3) wrap up and move forwards from the conflict. Results have shown that children were more likely to solve the conflicts constructively in the condition where the robot mediated the conflicts compared to the control condition.

Social robots can have different roles in the the way they interact with humans, from partners to opponents, or even from having mixed-motive goals to be mere facilitators or mediators. Tennent et al. have recently investigated whether a peripheral robotic microphone can positively shape the interpersonal dynamics of a team during a problem solving task [142]. They run a user study in which the robot displayed one of three possible behaviours: no movement, random, and engagement. The engagement behaviour balanced between following who had the conversational floor and encouraging the participant who had spoke the least (see Figure 2.4d). Results showed that in the engagement condition, the conversational dynamics was more even compared to the no movement. Curiously, there were no significant differences between the random condition and the engagement, nor between the random and no movement. This suggest the behaviour of the robot might have to be somehow socially contingent in order to differ from the stationary robot. Additionally, they found a correlation between the unevenness measure and team performance. Overall, the engagement behaviour produced more balanced discussions among the team and led to better team performance.

Jung et al. have also explored a mediation role by a social robot, in which the robot was assisting a team of two humans in a tower construction task (see Figure 2.4e). In particular, they examined different resource allocation algorithms and their impact on the task execution [81]. In their user study, the robot could either equally distribute the blocks among the two team members or execute an unequal distribution by giving 65% of the blocks to one participant and 35% to the other. The results revealed that team members in equal distribution reported a significantly more positive interpersonal relationship than team members in the unequal distribution condition.

More recently, Traeger et al. examined groups of three humans and one robot playing a collaborative game in two different conditions: one in which the robot made vulnerable statements and in another where the robot made neutral statements [145]. They found the vulnerable comments influenced people not only to perceive their groups more positively, but also to converse substantially more and distribute their conversation more equally.

In a similar task, a modified version of the Desert Survival Problem (Figure 2.4f), Sebo and collaborators investigated how verbal support can affect human-human interactions in terms of psychological safety and inclusion [126]. They found that outgroup members received fewer verbal backchannels from ingroup members if their group received verbal support behaviours, compared to the control condition. They also found outgroup member participated more after receiving targeted support from the robot.

Finally, Gillet et al. developed an adaptive gaze behaviour so that the robot would not only react to the talking person, but also proactively shift its attention towards the least active person in order to encourage an even participation [65]. The results evidenced a balanced participation between the two humans when the robot employed the adaptive gaze behaviours.

Within the reviews papers, we identified a trend on most papers to explore scenarios with mixed groups of humans and robots, in which the robot had a mediating role.

2.5 Analysis of Human-Robot Group Interaction

The scope of papers we have reviewed highlights a diversity of topics within human-robot group interaction. Beyond the categorisation we have created in the previous sections, there are other interesting ways to analyse group interactions regarding the type of tasks involved, the structure of the group itself, and the role of the robot(s).

In order to perform such comparison, we have used the circumplex of group tasks proposed by McGrath [101], see Figure 2.5. This model for classifying group tasks has eight octants along the two dimensions of *generate-negotiate* and *choose-execute*. In the generate quadrant, tasks can either be *creative* if the goal is to generate ideas, or of *planning* type if the goal is to elaborate a plan. In the choose quadrant, the task is to solve a problem that can either have, or not have, correct answers,

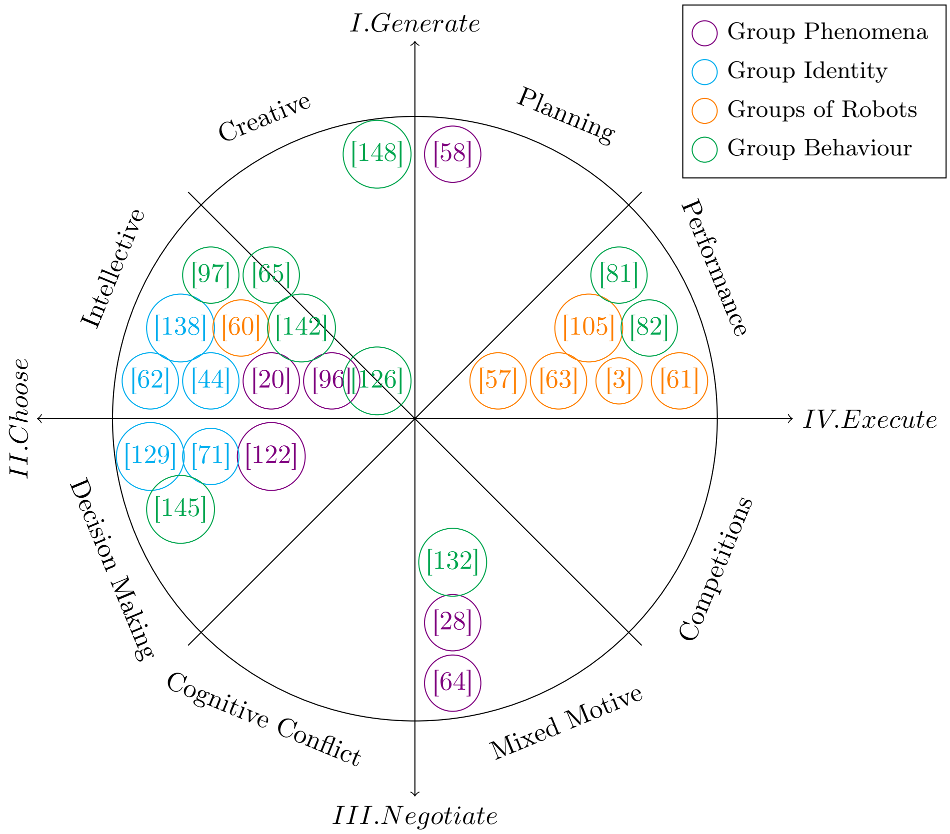


Figure 2.5: Reviewed papers in the Circumplex of Group Tasks [101].

intellective or *decision making*, respectively. In the negotiate quadrant, the tasks require conflict resolution and the octants depend on the nature of the conflict. It is a *cognitive conflict* in case of conflicting viewpoints and it is considered a *mixed motive* in case of conflicting interests. Finally, in the execute quadrant, tasks can either be *competitions* or of *performance* type if it involves resolving conflicts of power or execute a performance or psychomotor task, respectively.

Most of the works exploring group identity and group membership (blue references in Figure 2.5) are located in the *choose* quadrant, revealing the nature of those tasks requires people to perform some kind of decision. Additionally, we also noticed that the most common role for the robot(s) in this set of papers is the being a peer or a teammate. Consequently, the predominant structure or composition of these groups is human-robot mixed groups with collaboration.

The second group of papers we have analysed (violet references in Figure 2.5) corresponds to the exploration of group phenomena in HRI. There is no predominant type of task among the reviewed papers. Nevertheless, there seems to be a tendency to explore the role of a competitor. Most of the discussed scenarios attributed conflicting goals to both humans and robots in a way that they compete for resources either implicitly or explicitly.

Regarding the perceptions of groups of robots (orange references in Figure 2.5), we noticed that most tasks are located in the *performance* octant. The papers reviewed in this category discuss the perceptions people had seeing groups of robots executing a certain task, or videos of those performances, without requiring any human intervention. Therefore, it is hard to assess a specific role for robots in those situations. Nonetheless, as mere performers, such as the vacuum cleaner or the robotic trash bin, they can be seen as servants or subordinates.

Finally, the papers within the category of group behaviour cover several types of tasks (green references in Figure 2.5), including the uncommon creative type. Interestingly, we noticed a trend in the role of the robot(s) in this set of papers, which is a mediator or a facilitator. When the robot mediates or facilitates the interaction, its goal is somehow complementary to the humans and, therefore, it is strong enough to influence and shape the interaction.

Chapter 3

Membership preferences and team formation

Social cohesion emerges from the relations and attractions between team members, at an individual-level, or among the whole team, at a group-level. Nonetheless, it is not clear yet how those attractions develop when there is a social robot on the team. Moreover, as the interactions with robots become longer or occur in repeated events, we naturally attribute traits to these robotic partners based on their social behaviours. However, it is still not clear which traits we prefer in robotic teammates.

The goal orientation theory describes one of the traits that influences team interactions. At an individual level, people's goal orientations have a major effect on how they approach and respond to a task. Dweck extended the notion of goal orientation [50], initially introduced by Eison [52] and concluded that, during a task, people will present either a *learning goal* (*i.e.*, an interest in learning something) or a *performance goal* (*i.e.*, an interest in the result and what judgements will emerge from it). Teams consisting of individuals with a learning orientation are reported to show high levels of mutual support behaviours and high quality of interaction, team efficacy and commitment. By contrast, teams consisting of individuals with a performance orientation are negatively correlated with team efficacy and commitment [115].

As a result, we are particularly interested in exploring how membership preferences are influenced by the *goal orientation* of a robotic partner. To investigate such research question, we used a card-game scenario where two human-robot teams compete to collect more points and win the game. We developed two autonomous robotic partners displaying different goal orientations, which we describe in Sections 3.1 and 3.2. Then, we conducted two user studies: a first one to validate the perceived goal orientations on the robotic characters, detailed in Section 3.3; and a second study to assess membership preferences with those robots, detailed in Section 3.4.

3.1 Developing a Robotic Game Player

The two aimed robotic characters were embedded in similar robotic agents. The current section describes the development process for a robotic agent that acts autonomously. It begins with an analysis of a user-centred study that further inspired the definition of the agent's perceptions as well as its behaviours, which together constitute the main perception-reaction loop of each robotic game player.

3.1.1 User-Centred Design

We conducted a user-centred study to analyse and collect the behaviours of human players during the game interaction, as described in [33]. The video records of 10 independent games were annotated and converted into utterances of verbal and non-verbal behaviours. In addition to the dialogues, these utterances allowed us to analyse the relevant gaze directions during this card game, such as gazing at the table where the game was being played, at the partner, at opponents, at their own hand (own cards), or elsewhere. The initial coding scheme was semi-structured according to the following game stages that are commonly known: *session start*, *shuffle deck*, *split deck*, *deal cards*, *game start*, *next player turn*, *player played*, *trick end*, *game end*, and *session end*. Additionally, we realised a discrepancy in the competitiveness levels of some players between their interactions towards partners and towards opponents, especially during the game stages of *next player turn*, *player played*, *trick end*, *game end*, and *session end*, which we classified as the ones eliciting the competitiveness of the players.

Finally, we further examined the content of behaviours triggered during the game stages of *player played* and *trick end*, which revealed more complex appraisals. The participants of our user-centred study usually expressed how desirable such events are regarding their current scores in the game. Nevertheless, they were careful not to disclose information about their cards, which are unknown for the other players due to the fact that Sueca is a hidden-information game.

This detailed behavioural analysis was crucial to creating a social robotic player that acts in a natural and human fashion. Moreover, it guided the development of the two proposed characters (detailed in Section 3.2), where one is more performance-oriented and competitive, i.e., Emys, while the other is more learning-oriented and relationship-driven, i.e., Glin.

3.1.2 Perceptions of the Robot

After having analysed the relevant game stages that trigger the interactions among players during this card game, we were able to define the perceptions of each of the robotic players.

We decided to confine the perceptions of the robotic player solely to game events, without any further speech or image recognition. Therefore, a robotic game player requires the notification of all relevant game events, as well as their associated information, i.e., the identification of the player and/or team performing the event or the card being played. It is crucial that this communication occurs in run time so that the robotic player can autonomously react at the right moment.

Regarding the game events that elicit the competitiveness of a player, their perception process includes two additional appraisals. The first appraisal attempts to assess the immediate impact of a game event. It begins by identifying the team that benefits the most from it. In the case of a *trick end*, *game end*, or *session end*, this identification refers to the winning team. In the case of a *player played* event, it refers to the current trick winner. Although the ultimate winner of a trick can only be assessed after the four moves have been completed, human players usually acknowledge indefinite winners during the trick. Consequently, it is also important to assess the previous trick winner to check if the current move caused a change in the current scores. Finally, it is also relevant to quantify the impact of the play on the game scores. We have defined two boundaries based on the value of the highest cards of each suit in the Sueca game: *low* impact when the move/trick adds at least 3 points[1] and *high* impact when the move/trick adds at least 10 points[2]. The combination of these three factors is shown in Table 3.1 and is ordered according to how favourable the game event is at the moment for the agent's team.

The second appraisal process is an emotional appraisal related to the overall impact on the task. The agent uses an OCC model [110] with the appraisal variables of *Desirability*, *Desirability For Others*, and *Goal Likelihood*. Each agent has the goal of "winning the set of games". According to the following

[1] value of the 2^{nd} lowest card
[2] value of the 2^{nd} highest card

Table 3.1: Appraisal of game events ordered by the benefit they add to the agent's current trick score.

Previous Trick Winner	Current Trick Winner	Impact on Score	Valence
Other team	My team	High	+
Other team	My team	Low	
My team	My team	High	
My team	My team	Low	
Other team	Other team	Low	-
Other team	Other team	High	
My team	Other team	Low	
My team	Other team	High	

equations, the first two appraisal variables are updated after each *player played* event, while the last one is updated after each *trick end* event.

$Desirability = \frac{min(max(-15,TP),15)}{15}$

$DesirabilityForOthers = -Desirability$

$G.Likelihood = 0,5 \times \frac{MySP}{MySP+OSP} + 0,5 \times \frac{MyGP}{MyGP+OGP}$

$\|TP\|$ is the current trick points, and its valence may be positive if the current trick winner is the agent's team or negative if it is the other team. $MySP$ are the current session points of the agent's team, while OSP are the current session points of the other team. Similarly, $MyGP$ are the current game points of the agent's team, and OGP are the current game points of the other team. According to the OCC model, these appraisal variables will generate different types of emotions: well-being (i.e., joy and distress), fortune for others (i.e., resentment, gloating, happy for, and pity), and prospect-based (i.e., hope and fear).

The described appraisal mechanisms perform identically for our two robotic game players. Nevertheless, the role each player has on the game will produce different appraisals and, therefore, different perceptions. For instance, a player that adds more points to his/her team will be positively appraised by its robotic partner and negatively appraised by its robotic opponent.

3.1.3 Behaviours of the Robot

The final step in developing a robotic game player and closing the reaction-action loop is to define its behaviours. The goals of each robotic player include both socially interacting during the game and efficiently performing the task of playing the game, which results in both social and task-related actions, according to Figure 3.1 and described as follows.

Social Behaviours

The social behaviours of a robot are limited to the features of its embodiment. For the purposes of this card game scenario, we chose the stationary robotic head EMYS [87] capable of using gazes, animations, postures, and dialogues, which are expressed as follows.

The basic reaction to all the perceptions is a change in the gaze direction. For instance, game events such as *shuffle deck*, *split deck*, or *deal deck* cause an immediate shift in the gaze direction of

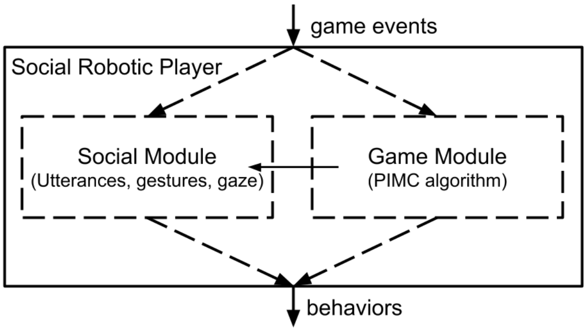


Figure 3.1: Social Robotic Player architecture

the robots towards the player performing these actions. In the same manner, game events referring to a new card on the table lead to a gaze towards the table centre. For the *next player turn* events, the gaze instructions depend on whether the next player is the robot itself or another player. In the former case, the robot immediately gazes at its hand (own cards). In the last case, it waits two seconds before looking to the next player to simulate that it is acknowledging the previous play.

Regarding other non-verbal behaviours, the robotic game player produces expressive facial animations and physical postures according to its activated emotional state. This is the behavioural response of the emotional appraisal previously described, which aims to display the perception of its performance on the task. As different emotions can be simultaneously activated by a single event (e.g., sadness and hope), the strongest emotion among the activated emotional states is used to drive the robot's physical posture and to select some of its animations. These non-verbal emotional behaviours perform similarly on each robotic game player, as their emotional appraisals are similar.

As for the robotic agent's more complex reactive behaviours in the form of utterances (which include dialogue, animations and/or gaze instructions), they appear as a reaction to a particular game event. Moreover, reactions will differ according to different perceptions of the game events, as shown in Table 3.2. Based on the user-centred study, competitive behaviours usually occur after extremely favourable or unfavourable game events, which, according to the appraisal presented in Table 3.1, occur in the highest and lowest situations. The authoring of the utterances is detailed in Section 3.2, as they were used to convey the manipulation of the goal orientation on the robotic players. Nevertheless, it is important to mention that all the utterances begin with a gaze towards the player the robot is interacting with and may end with a gaze towards the next player, if applicable.

Task-related Behaviours

For each robotic player to succeed in its task of playing the card game, it contains an algorithm able to perform online computations within the Sueca domain and to choose a suitable card to play.

Considering the main property of this card game, i.e., the imperfect information, we considered implementing the Perfect algorithm. It is based on the Monte Carlo method, which has recently successfully solved similar games, such as Bridge [66], Skat [23], or Hearts [141].

The algorithm 1 was further adapted according to the best parametrisation of the hybrid player for the Sueca domain, proposed and detailed in [34].

Algorithm 1 PIMC search pseudo-code.

procedure PIMC(InfoSet I, int N)
  for all $m \in$ Moves(I) **do**
    $val[m] = 0$
  for all $i \in \{1..N\}$ **do**
    $x =$ Sample(I)
    for all $m \in$ Moves(I) **do**
      $val[m]$ += PerfectInfoValue(x, m)
  return $\underset{m}{argmax}\{val[m]\}$

3.2 Creating Two Characters for Two Robotic Game Players

The two distinct characters will be identified by their names: Emys and Glin. Emys was given a *performance-driven goal orientation*, and as such, its behaviours and social actions are more aligned towards winning the game. Glin, by contrast, was given a *learning-driven goal orientation*; consequently, although Glin strives for its team to win the game, it also focuses on fostering team spirit and providing a good game experience.

The challenges associated with defining the two robotic characters were (1) how to reflect different goal orientations through the social interactions of two distinct robots and (2) how to guarantee, in the case of a group of two humans and two robots, that both robots are aware of and synchronised with the others, respect turn taking, and act naturally in a group of four. All the remaining aspects related to the technical development, tools and a detailed description about the rules of this card game are available in [39].

To address the first challenge, both robotic game players use the same agent. However, their utterances distinguish them as two different characters. In other words, their repertoire of dialogues was used to author the characters of Emys and Glin. Therefore, each robotic player has a unique set of utterances (420 per robot) for all the game events during a game session. The total amount is balanced to ensure that neither would be more repetitive than the other. Moreover, they produce behaviours with similar frequencies to ensure that neither would exceed the other in its interaction rate.

Table 3.2 exemplifies the differences between Emys' and Glin's interactions for the same perceived game states. For Emys, the utterances were built based on a competitive perspective, always in pursuit of the best score. For example, the emotion of joy is triggered when the situation reveals that its team is winning. At the same time, Emys will react with an angry emotion when losing and will consequently blame the others, either the partner or the opponents, for the game result. By contrast, Glin was built with different parameters, leading to a more relational perspective, verbalising more support towards its partner. When its team loses, Glin will respond with a sad emotion, encouraging its partner and fostering hope. Note that Glin also plays competitively, desiring its team to win but assuming more of a supportive role.

To produce natural interactions among the group of four (two humans and two robots) and considering the fact that both human and robotic players play certain roles (partner and opponent) in the game, the robotic players must be able to interact with each other in a manner as similar as

Table 3.2: Examples of utterances from Emys and Glin, the robotic partners with the performance and learning orientations, respectively.

Game State	**Emys**	**Glin**
Deal Cards	*“I only accept aces and sevens in my hand!"*	*“I hope there are good cards for everyone!"*
Self playing	*“Watch and learn how this is played."*	*“I am so proud to be on your team!"*
Partner played	*“Indeed, these points suit our team better."*	*“Our team is in sync!"*
Opponent's Turn	*“Play... or we will fall asleep."*	*“It's you, go ahead!"*
Partner's Turn	*“Don't disappoint me."*	*“Play with confidence!"*
Game End - Loss	*“This cannot continue like this! You have to play better!"*	*“No worries, next time we will do better!"*
Game End - Draw	*“With this score, I do not like to play."*	*“It's a draw... no worries, it's okay."*

possible to that in which they interact with human players.

Given that these autonomous robots do not have the capability to understand natural language, other mechanisms were implemented to achieve natural, believable, and human-like interactions. One fundamental capability required in this scenario is turn taking. For instance, humans use various sensory stimuli to perceive whether another person is going to speak, immediately establishing an order for the speakers according to each situation. Sometimes, a person will even step down from his or her intention to speak because someone else also started to speak or because there is no reason to speak anymore. To mimic this natural synchronisation process, we defined a two-phase handshaking protocol as an explicit communication interface. This protocol includes four messages: (1) to inform of an intention to speak, (2) to respond to an intention to speak, (3) to inform that an utterance has begun, and (4) to inform that an utterance has finished. Each robot can perform an utterance only when it receives a positive response. If it receives a negative response, it must wait and retry message (2) until it receives a positive response. A conflict may arise when a robot receives an intention to speak immediately after having sent the same message, as both robots will then receive a negative response and will both enter a retry loop. To avoid a communication deadlock, the two robots will retry their requests after different periods of time, which are randomly generated with values between 0 and 2 seconds. The next time, one of them may receive a positive response, and if not, they will continue retrying until a request receives a positive response or until a timeout period of 3 seconds has expired.

This simple mechanism enables a natural and fluid turn taking mechanism between the two robots. A similar mechanism with the human players would also improve the group interaction, but it was currently ignored due to its complexity. Nevertheless, we carefully avoided having explicit questions in the chosen dialogues. When necessary, we replaced them with rhetorical questions instead, as such utterances provide a rich feeling of interaction without requiring explicit answers. For instance, a robot may say *“Did we really lose this game?”* (Emys) or *“What am I going to play next?”* (Glin).

Figure 3.2: Experimental setting for Study 1.

3.3 Study 1: Character Validation

The first study was conducted to validate the differences between the two created characters, i.e., the more performance-oriented character, Emys, and the more relationship-oriented character, Glin. We expected that Emys would be perceived as more competitive, less helpful and less motivating and as providing less emotional security than Glin.

3.3.1 Sample

We recruited a total of 30 university students (17 males and 13 females) with ages ranging from 19 to 42 years old ($M = 23.03$; $SD = 4.21$). Among the participants, 56.7% had a high level of expertise in the game, 40% had a moderate level of expertise, and only 3.3% had never played the game before. Regarding previous interactions with this robot, 24 participants had previously interacted with it, and 6 were interacting with it for the first time.

Each participant was randomly allocated to a session in which three human participants played either with Emys or with Glin. This session lasted approximately 1 hour, and the instruments used were an EMYS robotic head [87], two video cameras to record the interactions, a multi-touch table, and a deck of physical cards with printed fiducial markers that could be recognised by the table.

3.3.2 Procedure

The participants arrived at the room in groups of three. A researcher received them, explained the rules of the game, and conducted a test game to address any doubts that might arise regarding the game rules. After the explanation, the participants joined either Emys or Glin (chosen randomly) at the table and played a set of 3 games. When finished, the participants were administered a set of questionnaires, filled out the consent form and received a thank-you gift (a movie ticket). We presented the consent form at the end of the experiment so that the participants' interactions during the game would be as natural as possible. If any participant had not given consent, his or her data would have been erased. However, all participants signed the consent form.

3.3.3 Measures

To represent our sample, demographic information was requested in the questionnaires (gender, age, previous interaction with the robot and level of expertise in the game). In addition, all participants, independently of being the partner or an opponent of the robot, responded to the following questionnaires regarding the robot (Emys/Glin):

- *Competitiveness Index* [135], used to measure the level of competitiveness perceived in the robot. This measure is usually treated as being of a dichotomous true/false answer type; however, as our goal was to determine a range from the participants' answers, we measured it on a Likert scale ranging from "totally disagree" to "totally agree". An example of a statement would be *"I consider Emys a competitive individual"* or *"When Emys plays, he likes to keep an eye on the score"*.

- *McGill Friendship Questionnaire* [102], using three of its dimensions, namely, help (e.g., *"Emys helps me when I need it."*), motivation (e.g., *"Emys praises me when I do something right."*) and emotional security (e.g., *"If I was worried, Emys would make me feel better"*), with scales ranging from "totally disagree" to "totally agree".

- *Relationship Assessment Scale* [72], adapted to the context and used to ascertain the level of quality of the relationship with the robot, ranging from "few" to "a lot" (e.g., *"How good was Emys relationship with the players?"*).

- *Godspeed Questionnaire* [15], using the two dimensions of perceived intelligence and likeability to assess the level of intelligence thought to be given to the robot and its perceived likeability, measured as a semantic differential.

All dimensions were measured on a 6-point Likert scale, and when necessary, items were shuffled to mask their dimensions.

3.3.4 Results

To understand whether the two characters were perceived differently, statistical analyses were performed. When a normal distribution was present, we performed the Student's t-test for independent samples, and when the normality assumption was not met, we used the Mann-Whitney U test. The means and standard deviations are presented in Table 3.3.

For the *Competitiveness Index*, Emys was rated higher than Glin, with a statistically significant difference ($t(25) = -4.893$, $p < .001$). Notably, Glin also presented a certain level of competitiveness, which was expected since it also had the goal of winning the game. Regarding the *McGill Friendship Questionnaire*, there were statistically significant differences in the three measured dimensions of help ($t(28) = 2.312$, $p = .028$), motivation ($t(28) = 3.686$, $p = .001$), and emotional security ($t(28) = 3.218$,

Table 3.3: Means and ranks with standard deviations for the questionnaire dimensions comparing the evaluations of the Emys and Glin characters in Study 1. *$p \leq 0.05$

	Questionnaire dimensions	Emys	Glin
	Competitiveness Index *	4.57 ±0.40	3.86 ±0.33
McGill	Help *	3.78 ±0.89	4.51 ±0.81
	Motivation *	3.79 ±1.00	4.95 ±0.69
	Emo. Security *	3.26 ±1.09	4.37 ±0.77
	Relationship Quality *	4.41 ±0.52	5.32 ±0.38
Godspeed	Perc. Intellig.	4.59 ±0.74	4.93 ±0.49
	Likeability *	10.70 ±0.88	20.30 ±0.88

$p = .003$), with Glin presenting higher scores than Emys. On the *Relationship Assessment Scale*, Glin was rated higher than Emys, with a statistically significant difference ($t(28) = 5.514$, $p < .001$).

These results confirm that the behavioural manipulation of the goal orientations of both robots was perceived as intended: Emys was seen as more competitive, and Glin was seen as more relationship-driven, with a greater capacity to be helpful and motivating and the ability to provide more emotional security. Moreover, the relationship quality scores were also higher for Glin than for Emys. We additionally evaluated whether the roles of the participants (partner/opponent) had any influence on the scores given to the robots, and we found no statistical significance for all measures, suggesting that the role did not affect the evaluations.

Finally, concerning the findings of the *Godspeed Questionnaire*, there was no significant difference between the two robots in the perceived intelligence dimension ($t(28) = 1.511$, $p = .142$). This was somewhat expected since we equipped both robots with the same algorithm for solving the card game. Although the game includes an element of chance and each new game presents different winning probabilities for each team, we can conclude that the intelligence levels of both robots were similarly perceived. However, in the likeability dimension, we found a significant difference, with Glin receiving higher scores than Emys ($U = 40.50$, $p = .002$).

In general, it seems that our implementation was perceived by the participants as intended, and Glin was rated as more likeable than Emys. We could now move on to the implementation of both characters at the same time, using the two robots to test which would be the preferred partner.

3.4 Study 2: Choosing a Robotic Partner

The purpose of this study was to assess the participants' preferences regarding the choice of a robotic partner.

3.4.1 Sample

For the second study, we recruited a new sample consisting of a total of 61 participants (59 university students and 2 workers), 38 male and 23 female, with ages ranging from 17 to 32 years old ($M = 23.66$, $SD = 3.24$). The majority of the participants had never interacted with a robot before and had a

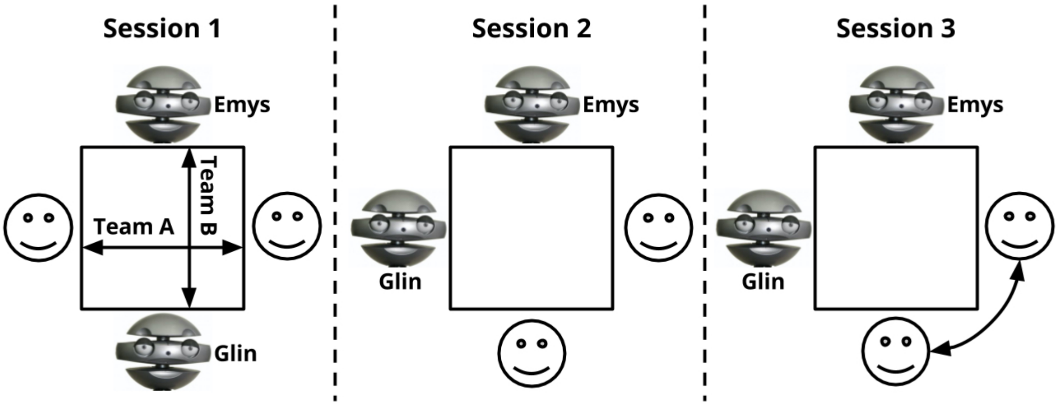


Figure 3.3: Experimental setting for Study 2.

moderate or high level of expertise in the game.

We measured the level of competitiveness of each participant using the Competitiveness Index [135]: 15 participants presented low levels of competitiveness (less than or equal to $M = 3.50$), 36 participants presented some level of competitiveness, and 10 participants showed high levels of competitiveness (higher than $M = 4.50$).

Each session was run with two human participants who did not know each other beforehand. We controlled for this factor to ensure that the participants were in the same position with respect to both each other and the robots. Each session lasted approximately 1 h 30 m, and the instruments used were the same as in the previous study except that two EMYS robotic heads were used simultaneously during the game interaction. A name tag was placed below each robot with its name, Emys or Glin, to allow the participants to easily identify them.

3.4.2 Procedure

The participants arrived at the room and responded to the first part of the questionnaire (see the Measures subsection below). Then, a researcher explained the game rules and conducted a test game to address any doubts that might arise. This study was divided into 3 consecutive sessions, as shown in Figure 3.3.

1st Session: The two participants partnered with each other and played a set of 3 games against the two robots (Emys and Glin), which acted as their opponents in the game. This session served to expose the participants to the two different characters while having the same role towards each one. After completion, the participants responded to the second part of the questionnaire.

2nd Session: Each participant partnered with one of the robots (see Fig. 3.4), and the group played another set of 3 games. The participants then responded to the third part of the questionnaire.

3rd Session: The participants played their last set of 3 games, now partnering with the robots with which they had not played before, and then responded to the fourth part of the questionnaire. At the end, they were given the consent form and were thanked for their participation with a movie ticket.

The balance between the orderings was ensured by the fact that participants attended in pairs and that while one participant was Glin's partner, the other one was Emys' partner. In the last two

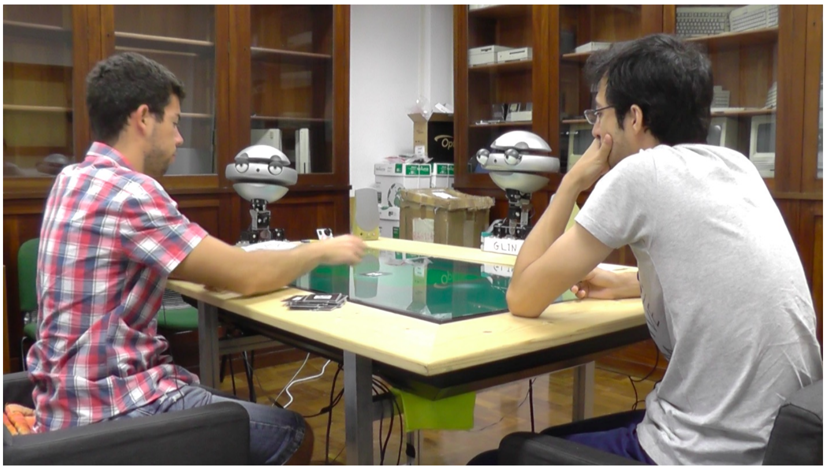


Figure 3.4: Experimental setting for Study 2 when each robot was partnering with a human.

sessions of each experiment, the participants had the opportunity to partner with both robots. We randomised which participant partnered with each robot first.

3.4.3 Measures

We used the same questionnaires as in the first study, organised in the following way:

First Part: The participants filled out some demographic questions and then an assessment of the *Competitiveness Index* related to themselves.

Second Part: The participants completed a questionnaire assessing the two *Godspeed* dimensions (perceived intelligence and likeability) for both robots and answered the following question: "If you could choose one of the robots as your partner, which one would it be? (Emys or Glin)".

Third Part: Each participant completed a questionnaire assessing the two *Godspeed* dimensions, the three *McGill Friendship* dimensions (help, motivation and emotional security) and the *Relationship Assessment Scale* with respect to the robot he or she had just partnered with.

Fourth Part: The same as the third part of the questionnaire but with respect to the new robotic partner. At the end, the participants were again asked to choose which robot they would prefer to be partnered with for future games and to justify their choice.

All dimensions were measured on a 6-point Likert scale, and when necessary, items were shuffled to mask their dimensions.

3.4.4 Results

Below, we present the results of this user study, beginning with how participants perceived each robot. Although we previously checked our manipulation, we repeated the analysis to check if the robots were perceived differently in the new 2-robot and 2-human players setting.

Then, we present the results of the participants' initial choice for the preferred robotic partner. We analysed the effect of participants' competitiveness index on this choice.

Finally, we present the participants' last choice after interacting with both robots. We analysed the effect of other measures, such as participants' competitiveness index and the team performance.

We also explored changes from the initial to the last choice and the participants' justifications for the chosen partner.

Due to the high number of statistical tests performed, we performed a Holm's sequential Bonferroni correction [76] to ensure that there were no false positives in our results, and this assumption was met for all the statistical tests.

Results (I) - Perception of the Robots

We began by analysing how the participants perceived each robot in their initial interactions. When the normality assumption was not met with the Shapiro-Wilk test, we used the Wilcoxon signed-rank test. The means and standard deviations are presented in Table 3.4.

Regarding the *McGill Friendship Questionnaire*, there were statistically significant differences in the help ($Z = -5.223$, $p < .001$), motivation ($Z = -6.066$, $p < .001$) and emotional security ($Z = -5.837$, $p < .001$) dimensions, with Glin being rated higher than Emys. For the *Relationship Assessment Scale*, there was also a statistically significant difference ($Z = -4.392$, $p < .001$), with Glin being rated higher than Emys, representing a higher relationship quality.

These latter two results confirm the successful behavioural manipulation of the robots. After interacting with both robots, the participants seemed to perceive Glin as having a greater capacity for being helpful and motivating and for providing more emotional security compared with Emys. Moreover, the participants perceived Glin as displaying a better relationship quality than Emys. Overall, these results seem to support the more relationship-driven characteristic with which we attempted to endow Glin, demonstrating the successful development and implementation of the two autonomous robots.

The participants assessed the two dimensions of the *Godspeed Questionnaire* for each robot twice, the first time before partnering with either of the robots and having only observed them as opponents and the second time immediately after having partnered with that robot. For the perceived intelligence dimension, we found no statistically significant difference between Glin and Emys in either the first measurement instance ($Z = -.733$, $p = .464$) or the second ($Z = -1.491$, $p = .136$). Thus, by using the same decision-making algorithm for both robots in this hidden-information card game, we achieved similar levels of perceived intelligence in both, as intended. For the likeability dimension, there was a statistically significant difference, with Glin receiving higher scores than Emys in both the first measurement instance ($Z = -3.451$, $p = .001$) and the second ($Z = -6.224$, $p < .001$).

Results (II) - Initial Choice of Robotic Partner

The participants were asked to choose which robot they would like to have as a partner immediately after the first session (in which they had both robots as opponents and had partnered only with another human participant). This allowed us to assess the first impressions people had of the robots and how these would guide their choice of partner. The results showed that 38 of the participants would prefer to have Glin as a partner, whereas 22 preferred Emys. Running a chi-square goodness of fit test, we found a statistically significant difference between the participants' choices ($\chi^2(1) = 4.267$, $p = .039$), with more people preferring Glin (63.3%) compared with Emys (36.7%). In this stage

Table 3.4: Means and ranks with standard deviations for the questionnaire dimensions comparing the characters Emys and Glin in Study 2. BP stands for "before partnering", and AP stands for "after partnering". *$p \leq 0.05$

	Questionnaire dimensions	Emys	Glin
McGill	Help *	3.35 ±1.08	4.42 ±1.13
	Motivation *	3.15 ±1.09	4.79 ±0.90
	Emo. Security *	2.58 ±1.14	4.29 ±1.19
Relationship Quality *		3.93 ±0.89	4.80 ±0.93
Godspeed	Perc. Intellig. (BP)	4.51 ±0.86	4.53 ±0.99
	Likeability (BP) *	3.70 ±1.19	4.28 ±0.94
	Perc. Intellig. (AP)	4.40 ±1.04	4.55 ±1.13
	Likeability (AP) *	3.51 ±1.35	5.25 ±0.75

of the experiment, the robots were on the same team, and as such, the performance of one robot could not be contrasted with the performance of the other. To better understand the participants' choices, we also compared the participants' competitiveness scores based on their chosen robots using the Student's t-test for independent samples, and we found that there was no statistically significant difference between the competitiveness scores of participants who chose Glin and those who chose Emys ($t(58) = 1.242$, $p = .219$). This suggests that at this stage, competitiveness did not influence the partnering choice. Therefore, the participants' choices seem to have been guided by the different social behaviours exhibited; in this case, the participants were more drawn to the relational robot (Glin), which, according to the Results (I) section, was perceived as more likeable than Emys. Thus, the findings support our hypothesis that people seem to prefer a friendlier and more relationship-oriented robotic partner. However, we also wished to investigate whether these characteristics would continue to drive the participants' preferences after they had interacted with both robots as partners.

Results (III) - Final Choice of Robotic Partner

When asked to choose a robotic partner in the last questionnaire session (after having partnered with both robots), 35 of the participants preferred Glin and 25 preferred Emys (one participant refrained from choosing). Running a chi-square goodness of fit test, we found no statistically significant difference between the participants' choices ($\chi^2(1) = 1.667$, $p = .197$). We then investigated the factors driving the participants' choices at this stage of the interaction.

Looking at the levels of competitiveness of the participants and comparing them according to their final choices, we found a statistically significant difference ($t(58) = 2.953$, $p = .005$), indicating that the participants who chose Emys also tended to have higher competitiveness scores ($M = 4.21$, $SD = 0.67$) compared with the scores of the participants who chose Glin ($M = 3.73$, $SD = 0.58$). This implies that a participant's own characteristics (being more or less competitive) played a role in his or her choice of robotic partner after interacting with each robot on his or her team over repeated interactions.

Since the participants partnered with both robots, we also considered the possibility that the performance of the team formed with each robot (winning or losing) also affected the partner choice. To investigate this, we calculated the performance of each human-robot team using the summed results of the sessions, i.e., the sum of the points that Glin's team earned in Session 2 + Session 3, independently of its human partners, compared with the points earned by Emys' team. We observed that based on this criterion, Emys' team won 16 times and Glin's team won 12 times (4 draws occurred). Although this difference was not statistically significant ($\chi^2(1) = .571$, $p = .450$), we found a significant association with the partnering preference using Fisher's exact test ($p = .008$). It seems that the participants aligned their choices with the robot that was winning more. However, we must be careful with this assumption; each robot was always playing on a team, so if a particular robot won, its win was due not only to its own performance but also to its human partner's performance. Therefore, we can speak of the team performance as a factor influencing the partner choice.

Looking only at the participants who changed their choices of robotic partner between the first session and the last, we found a statistical association between the last chosen robot and that robot's team performance according to Fisher's exact test ($p = .002$). By contrast, for the participants whose choices did not change, no significant association was found according to Fisher's exact test ($p = .409$). This suggests that the participants who changed their choices did so because of the robot's team performance, thereby solidifying the conclusion that the team performance was indeed one factor accounting for the partner choice, but not the only one.

To clarify whether the robot's character had any influence on the participants' choices at this stage, we analysed their justifications for preferring their chosen robots. For this purpose, two coders (who were completely unaware of the purpose of the study) coded the participants' phrases according to the following coding scheme: they coded a response as *relational* if the justification for the choice of robot was more closely related to team spirit or the robot showing a warmer, more motivating, or more supportive attitude toward its partner, and they coded a response as *competitive* if the justification was based on the robot being the best robot, earning more points, or being more competitive either on its own or towards its opponents. This coding scheme was based on the development objectives for the two different characters. The Cohen's kappa value was k=.73 ($p < .001$), revealing good agreement between the coders. We found from the analysis that Glin was chosen 26 times with relational justifications and only 9 times with competitive justifications. By contrast, Emys was chosen 21 times with competitive justifications and 4 times with relational justifications. These results suggest that the robots' characters were also perceived by the participants and used to justify their choice, although this was not the only factor considered.

Overall, these results suggest that *team performance*, a *person's level of competitiveness*, and the *robot's character* play a role in a person's choice of a robotic partner after having previously partnered with it.

3.5 Concluding Remarks

We explored preferences regarding robotic partners in mixed teams of humans and robots. Moreover, we studied the factors driving the human participants' partnering choices. For this purpose, we developed two autonomous social robots with different characters, i.e., Emys and Glin, a more competitive robot and a more relational robot, respectively. These two autonomous robots interacted in a group with two humans while playing a competitive game. We began by validating that the two robotic characters were, in fact, differently perceived by the participants. Then, we investigated which of them would be chosen by the participants as a partner for future games. The participants were asked which robotic character (Emys or Glin) they preferred at the two following points in time: (1) before having partnered with either robot and (2) after having played with both robots as partners.

The partner choices seemed to be guided by different factors depending on the context of the participants. In the first session, when the participants had experienced both robots as opponents and had not yet created a partner relationship with either, they seemed to choose their partners based solely on character (either the relationship-driven or competitive robot). At that time, Glin, the relational robot, was the preferred partner. This finding confirms our hypothesis, consistent with the study of [115], that teams whose members prioritise relational features are perceived more positively (e.g., reporting higher levels of supportive behaviour and higher-quality interactions).

However, at the end of the final session, when they had experienced a partner relationship with each robot, the participants' choices became less clear, calling attention to other factors that came into play. It seems that *personal characteristics* and *team performance* took higher precedence when participants had experienced partner-partner relationships with the robots. The participants seemed to be affected by their *own characteristics* in their partner choices, as we observed that participants with higher levels of competitiveness tended to choose the more competitive robot (Emys), whereas the less competitive participants tended to choose Glin. At the same time, although both autonomous robots played the game using the same algorithm and the difference between the numbers of victories achieved by Emys' and Glin's teams was not significant, there was an association between the team performance and the chosen robot. It was observed that the change in participants' choices between the first and last sessions showed a significant association with team performance. Reinforcing this observation, the performance of the team was also a factor in the final choice of the preferred partner. The same association was not observed for the participants who maintained their choices. In addition, the robot's character also seemed to have influenced the choice, as the participants' justifications of their choices were related to the robots' characters. For example, Glin was chosen because it was much more relational, whereas Emys was chosen because it was more competitive.

The second user study, in particular the first session where both participants were opponents to both robotic characters, was carefully designed to expose the characters to the users on an equal footing. We note, however, that the subsequent user choices and preferences might have been different without this initial session. Moreover, our results do not explore ordering effects, which might be interesting to explore in the future.

Nevertheless, these results have important implications for the creation of robotic teammates who can adapt to their human partners' specific characteristics. Consistent with recent findings [60] showing that people perceive multiple robots that act and look the same as more threatening than a diverse group of robots, people's preferences also need to be considered in the creation of mixed human-robot teams. Indeed, as we move towards scenarios featuring interactions among multiple robots and multiple users, the "diversity" of the robots should not only be investigated but also engineered.

3.6 Other follow-up works

With the data collected in the second user study (described in Section 3.4), we performed a behavioural analysis. Using a coding scheme based on Bales Interaction Process Analysis [10], the video-recorded interactions were analysed in terms of socioemotional positive, negative and task oriented behaviours. A marginal multilevel modelling analysis yielded significant interactions between the robotic addressee and the role the robot displayed in the socioemotional and task-oriented behaviours. Overall, our main results demonstrated the following: (1) Participants directed more behaviours towards partners than opponents, although most of these behaviours occurred between humans when they were partners. (2) When comparing players in the role of opponents, participants directed more socioemotional behaviours towards robots than towards the other human player. (3) No difference in task-oriented behaviours was observed among any of the players in this condition. These results suggest the occurrence of different behavioural patterns in competitive and collaborative interactions with robots that might be useful to inform the future development of more socially effective robots. The full description of this follow-up work can be found in [107; 109].

Chapter 4

Prosociality in human-robot teams

Collective cohesion reflects the degree of identification with the group. It is usually mentioned as the replacement of the "I" by the "we" and, when this feeling becomes stronger, it may even affect one's actions with the collective goals taking priority over the individual ones. As a result, we developed a new scenario where both the outcome could be controlled and the actions of the team members would express either a low or high degree of cooperation towards a collective goal. The scenario is called "For The Record", a collective risk dilemma mapped into an entertaining game, and its description is presented in Section 4.1.

We conducted a user study where a team constituted by two social robots and one person would play a session of "For The Record", as detailed in Section 4.2. We manipulated not only the outcome of the game to either be a victory or a loss, but also the degree of cooperation on the two robots, high (prosocial robot) and low (selfish robot). This experimental set-up allowed us to investigate the following three research questions:

- How do people perceive pro-social and selfish actions of robotic teammates?
- How can the perception of those robotic teammates be affected by the outcome of the team?
- Does the outcome of the team affect how humans identify with the team and trust it?

The chapter proceeds with the results of the user study in Section 4.3 and a detailed discussion of our hypotheses in Section 4.4. Finally, some concluding remarks are presented in Section 4.5.

4.1 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 pro-social collaboration and, in a more general perspective, the creation of new approaches in which robots can promote pro-sociality 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:

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

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.

4.2 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 4.1). Furthermore, we set the number of rounds, R, to 5 and the limit

Figure 4.1: 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.

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 [85], 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. 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 [87]. 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.

4.2.1 Hypotheses

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

H1: The pro-social robot will be perceived more positively in its social attributes than the selfish robot.

H2: The pro-social 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 pro-social robot.

H5: When the team loses, the main responsible factor will be the strategy of the selfish robot.

H6: The pro-social 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 pro-social 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 [92]. It is possible that the same effect occurs in our study given that pro-social robot's behaviour. The reason for **H3** is the aforementioned loss aversion bias [85] and finally, **H4** and **H5** are based on the assumption that people will correctly identify the main responsible actor behind the team's result.

4.2.2 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.

4.2.3 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** [94] using the Portuguese adaptation [117] with the dimensions of Solidarity, Satisfaction and In-Group Homogeneity; **Group Trust** [4]; **RoSAS** [26] 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.

4.2.4 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.

4.3 Results

4.3.1 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 4.2, $F(1, 67) = 17.366, p < 0.001, r = 0.454$) 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).

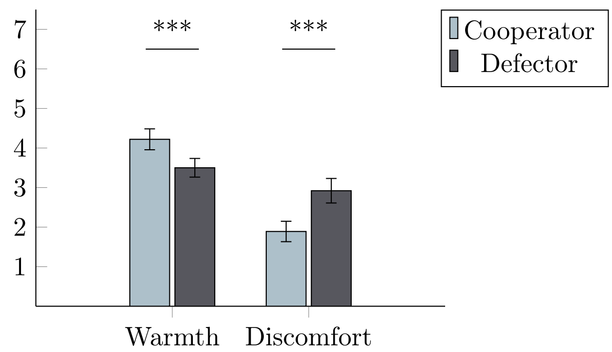


Figure 4.2: Main effect of the robotic partner on the social attributes of warmth and discomfort.

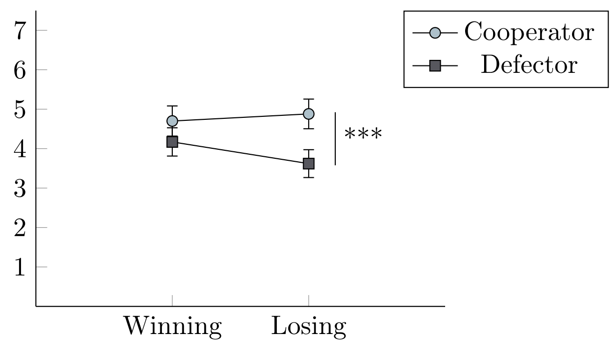


Figure 4.3: Interaction effect between the robotic partner and game result on the attributed levels of competence.

For the social attribute of discomfort, there was a main effect of the robotic character (Figure 4.2, $F(1, 67) = 30.982, p < 0.001, r = 0.562$), 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 4.3, $F(1, 67) = 4.095, p = 0.047, r = 0.240$). To understand this interaction, we compared the perception of 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 ($M = 3.624, SD = 0.896$).

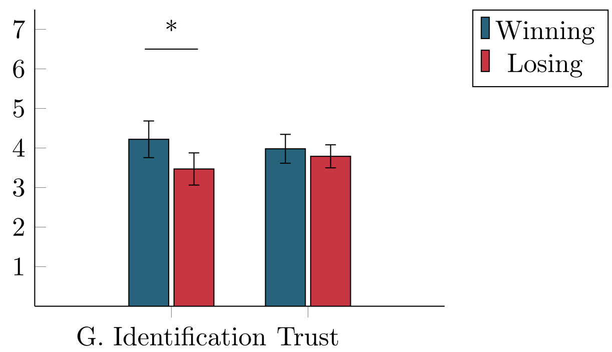


Figure 4.4: Effect of the game result on the attributed levels of group identification and group trust.

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. In terms of the team's evaluation, the identification with the group was indeed positively associated with the final outcome, although the same association was not verified for group trust.

4.3.2 Group Measures

To analyse the two dependent measures related to the group (Figure 4.4), 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 ($U = 535.5, Z = -0.715, p = 0.474, r = -0.086$).

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

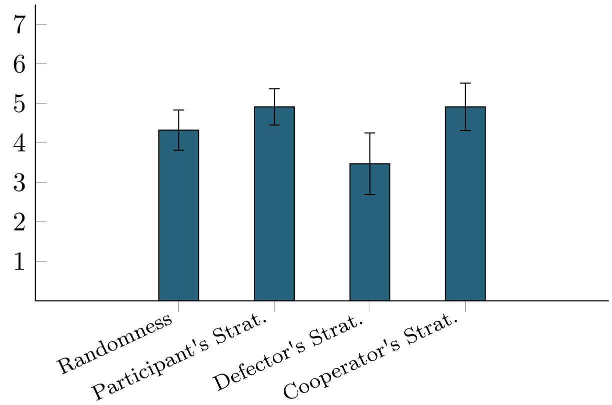


Figure 4.5: Responsibility attributed to each factor in winning condition (credit).

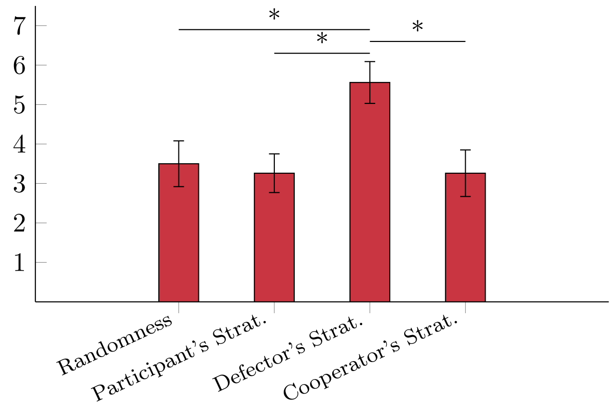


Figure 4.6: Responsibility attributed to each factor in losing condition (blame).

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.

4.3.3 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 4.5), 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 4.6), the blame attribution was significantly different to the four possible factors ($\chi^2(3) = 33.264, p < 0.001, r = 0.326$).

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

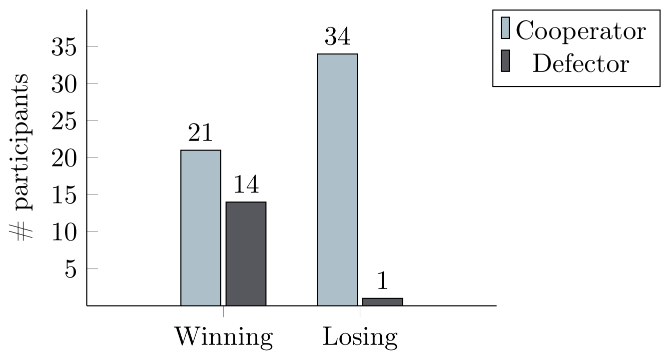


Figure 4.7: Preferences for each robotic partner grouped by conditions.

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.

4.3.4 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 4.7). 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%).

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.

4.3.5 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.

4.3.6 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 ($M = 4.686, SD = 1.676$).

4.4 Discussion

According to **H1**, we have predicted that the pro-social 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 pro-social 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 [70], 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 pro-social robot. There was no main responsible factor on the credit attribution of the winning outcome. Only 8 participants (11%) used the same pro-social strategy of cooperating 4 times and most participants defected at least 2 times (58.5%). Although most participants were more selfish than the pro-social robot, they attributed credit similarly between their own strategy and pro-social 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 pro-social robot.

Finally, **H6** hypothesised that the pro-social robot would be preferred as a future partner, which was only partially verified from our results. The preference for the pro-social 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.

4.5 Concluding Remarks

We are moving towards a society in which robots are increasingly present and able to work with us. In this project, we explored the role of pro-sociality as a contributing factor to establish cohesive collaborations with robots and, in particular, the impact of the outcome on the establishment of those alliances.

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 pro-social 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 pro-social always cooperated. Moreover, we manipulated the outcome of the game to either result in winning or losing.

Results showed that a pro-social 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 project also sheds 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.

Finally, 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 [17].

4.6 Other follow-up works

The work presented in this chapter had three follow-up projects. Two of them will be briefly described as follows, while the third one is detailed in the next chapter.

Motivated by the vision of prosocial computing and the challenges of engineering prosociality with autonomous agents [112], this follow-up project explored how robotic teammates can enhance and promote cooperation in collaborative settings. In particular, we looked at the display of verbal criticism by a robotic teammate towards noncooperator human partners, and we were interested in analysing such behaviour in two particular situations: one where the robot acts according to its own criticism and another where the robot hypocritically adopts the strategy it criticises. It run a user study in which participants engaged with two fully autonomous robotic partners to play "For The Record". Each participant collaborated with two robotic partners that adopt opposite strategies to play the game: one of them is an unconditional cooperator (the pro-social robot), and the other is an unconditional defector (the selfish robot). In a between-subjects design, we manipulated which of the two robots criticizes behaviours, which consists of condemning participants when they opt to defect, and it represents either an alignment or a misalignment of words and deeds by the robot. We found that (1) the misalignment of words and deeds may affect the level of discomfort perceived on a robotic partner; (2) the perception a human has of a robotic partner that criticizes him is not damaged as long as the robot displays an alignment of words and deeds. The full description of this follow-up work can be found in [35].

The second follow-up project was inspired by the result in Figure 4.7, which showed people only preferred cooperative partners when they lost a previous game. We developed a simplified evolutionary game theoretical model that sheds light on these results, pointing the evolutionary advantages of selecting cooperative partners only when a previous game was lost. We showed that this strategy constitutes a convenient balance between strictness (only interact with cooperators) and softness (cooperate and interact with everyone), thus suggesting a new way of designing agents that promote cooperation in Collective Risk Dilemmas [124]. Later, these theoretical results were confirmed through computer simulations considering a more complex strategy space and more empirical evidences providing additional support to the human predisposition to use outcome-based partner selection strategies in human–agent interactions [123].

Chapter 5

Teaming Up With (Dis)Embodied Agents

Considering the same scenario of the previous chapter (the "For The Record" game), which is inspired by the game theory to study human behaviour in collective situations, we wondered to what extent the same results could be obtained when agents lack embodied features. While several findings were both observed in interactions with virtual agents and with robots [13; 95; 128; 75], there is also evidence this interchangeability might not be straightforward [45]. This question becomes even more relevant during the current pandemic of SARS-CoV-2[1], during which several user studies had to be adapted to virtual solutions and the validity of transferring the results is still being analysed [8].

This project explores the impact of embodied affordances of socially interactive teammates in a cooperative task. If previous findings support the general idea that people comply more with social and moral norms around robots (compared to virtual agents), we also expect that people would cooperate, trust and identify more with a team of robots (compared to a team of virtual agents). Consequently, we also analyse how people perceive an artificial teammate that does not comply with collective norms and how the perception of such teammate is affected by its embodied affordances.

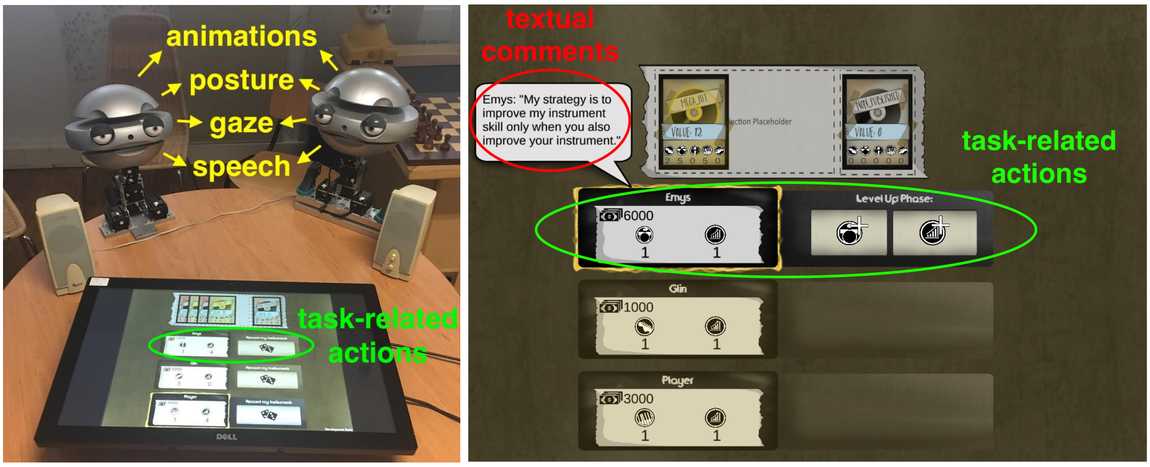
We conduct a user study using a mixed-design, where we manipulate the embodiment (between-subjects) and the degree of cooperation employed by each agent (within-subjects). As a result, each participant engages in a team setting with two autonomous agents that display opposite strategies to play a Collective Risk Dilemma: one is prosocial (high degree of cooperation) and the other is selfish (low degree of cooperation). Generally, we hypothesise that the embodied condition increases the degree of cooperation of the participants and would improve their subjective evaluation of the team. Moreover, we also expect that, in the embodied condition, the perception of the prosocial agent would be more positive and the perception of the selfish agent would be more negative, compared to the disembodied condition.

5.1 Background on Embodiment

There is no consensual definition of embodiment, especially when artificial agents come into play. For instance, Ziemke identified six notions of embodiment [151]. Although most of these notions try to discern whether a living body is required, there is a particular one that Ziemke identified as an orthogonal perspective – the *social embodiment*. It was initially proposed by Barsalou et al. and is focused on the interplay of embodiment and the social interaction [12]. The *social embodiment* requires an instantiation that mirrors "the state of the body" and has a central role during the interaction. However, such instantiation does not have to necessarily occur in the real world, as the notions of physical-, organismoid- or organismic embodiment do, and therefore considers as well virtually-embodied agents.

This project adopts the aforementioned *social embodiment* notion and identifies as "disembodied" any agent that is not instantiated in a body, i.e. without any virtual or physical shape. Consequently, it identifies any robot or virtual agent as embodied (physically- or virtually-embodied, respectively), as long as they possess some external and visible shape or representation for the user. Nevertheless,

[1]Severe Acute Respiratory Syndrome CoronaVirus 2, also known as COVID-19



(a) Embodied condition (b) Disembodied condition

Figure 5.1: Manipulation of the embodiment

we would like to emphasise the close relation between *social embodiment* and the *structural coupling* defined by Quick et al. [116], in which an agent is considered embodied as long as mutual perturbations between it and its environment are possible. Later, on top of this notion, Dautenhahn et al. have proposed that the degree of embodiment can be measured by quantifying those mutual perturbations between the agent and the environment [42]. Interestingly, this notion does not require the agent to possess a body and the authors have actually raised the question "*(...) what, if anything, is special about material embodiment?*". In that perspective, our definitions of embodied and disembodied agents could also be mapped into a higher and lower degree of embodiment, respectively. Due to clarity and practical reasons, we will keep the notion of *social embodiment*.

5.2 Experimental Setup

This user study aims at exploring the effects of embodiment on human perceptions of robotic teammates, as well as their degree of cooperation with the team of robots. In our experiment, participants were asked to team up with two autonomous agents and play a collaborative game. We manipulated, in a mixed-design (Figure 5.1), the embodiment of those agents (disembodied vs. robot), the outcome of the game (team wins vs. team loses), and the strategy of each agent (selfish vs. prosocial).

The agents are fully autonomous and interact in a social manner during the game. Those interactions were developed with the SERA toolkit [119] and are reactive to events from the game engine. The verbal and non-verbal utterances were previously scripted but they are chosen by an emotional decision making framework [98]. The code of the game and the scripted social behaviours of the agents are openly available[2].

[2] `https://github.com/SamGomes/for-the-record`

5.2.1 Task

The main task was to play a collaborative game, called *For The Record*, which is extensively described in [124]. The game frames a Collective Risk Dilemma and uses a musical metaphor making it more entertaining and pleasing. Players form a "a band of musicians" and their goal is to record as many successful albums together. During the game, each player has two individual skills: *instrument* and *marketing*. The *instrument* is used to help the team create a good album, while *marketing* is used to collect individual revenue in the game. In the end, players are ranked according to their individual profit, but only if the team won the game (reached the threshold), otherwise all players lose their accumulated profits. The skills of *instrument* and *marketing* are used to throw dice and, consequently, to determine the contribution to the album (collective goal) and receive individual profit (individual goal), respectively. In each phase of the game, each player can throw as many dice as their level on that particular skill. The mixed-motive decision occurs when each player has to decide which skill they want to upgrade, which happens once per round. By upgrading the *instrument*, a player increases the likelihood of team's success. By upgrading the *marketing*, on the other hand, a player increases the likelihood of his individual success.

Another important aspect of this game is the uncertainty factor due to the outcomes being determined by digital dice rolls. Players may increase their probability to succeed collectively or individually (by cooperating or defecting, respectively), but the outcome of their actions is ultimately controlled by an uncertain event. For the purpose of our experimental study, the outcomes of the dice were secretly manipulated to ensure a particular team outcome, given the experimental condition assigned to the participant.

5.2.2 Independent Variables

Embodiment (between-subjects)

The embodiment was manipulated in two levels. Participants would either play with two physically-embodied and co-present robots (Figure 5.1a) or with two disembodied agents (Figure 5.1b). The agents were fully autonomous in both conditions. In terms of behaviour, the robots in the embodied condition performed verbal and non-verbal behaviour (gaze, posture, animated expression of emotions). The disembodied agents only had verbal behaviour in the form of silent speech bubbles and used the exact same speech acts as the robots in the embodied condition. The speech bubbles of each agent would always be displayed in distinct positions to distinguish which one is speaking.

Agents' Playing Strategy (within-subjects)

The two agents played the game with distinct strategies: the *prosocial* agent chooses to upgrade the *instrument* skill in each new round (cooperation); the *selfish* agent chooses to upgrade the *marketing* skill in each new round (defection). As each participant was exposed to the two strategies, this was a within-subjects variable. The agents were distinguished by their names in the game interface and, accordingly, in the questionnaires (*Emys* - selfish strategy; *Glin* - prosocial strategy).

Outcome (between-subjects)

The manipulation of the outcome had two levels in a between-subjects design: each participant would either win or lose the game. This was achieved by secretly manipulating the value obtained in the (digital) dice throws during the game. The rationale for this control is based on previous findings suggesting that the outcome of a collaborative task significantly affects the perception of robotic teammates [37]. Note that by scripting the outcomes of the both dice, the *instrument* and the *marketing*, we can control both if the team wins or loses the game, as well as the ranks of each player, respectively. The *selfish* agent was always at the first place and the *prosocial* agent at last. The participant's punctuation could be closer to one of the two agents, according to strategies employed by the participant, but would always end up at second place in the game.

5.2.3 Hypotheses

Our hypotheses are based on the premise that the interaction people have with robots is closer to the one they have with humans when compared with virtual or virtually-displayed agents. Overall, it seems that either the social presence or the physical presence of a robot leads people to attribute more agency, authority and realism to it, as well as they seem to comply more with social and moral norms around robots [9; 13; 91]. Consequently, we expect people to tolerate less the artificial teammate that is more selfish (and to praise more the prosocial teammate) when it is embodied compared to when it is disembodied. The following hypotheses are aligned with these results, in the sense that they assume that the presence of embodiment improves the perception of the team, the social attributes of the agents, and the degree of cooperation of the participants.

H1 Participants will perceive the team more positively in the embodied condition compared to the disembodied condition.

H2a Regardless of the embodiment, the perception of the prosocial agent will be more positive than the selfish agent.

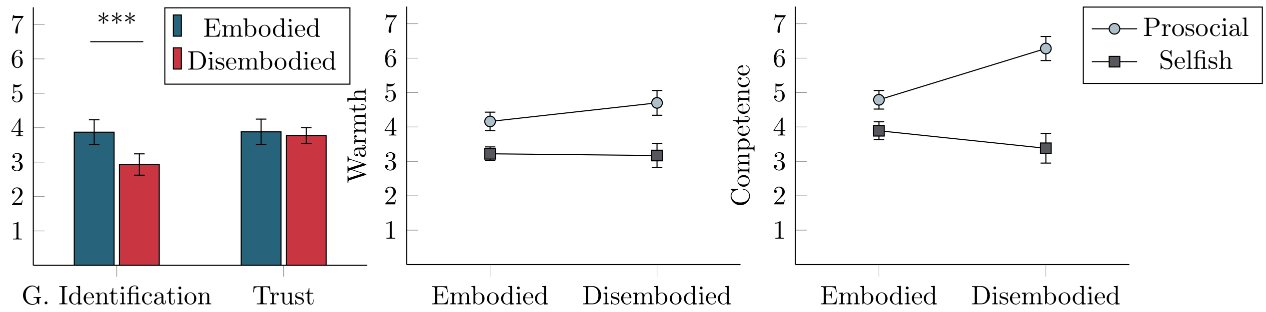
H2b The perception of the prosocial agent will be more positive when it is embodied than when it is disembodied.

H2c The perception of the selfish agent will be more negative when it is embodied than when it is disembodied.

H3 Participants will be more prosocial in the embodied condition compared to the disembodied condition.

5.2.4 Dependent Measures

We assessed participants' subjective evaluation of each agent in terms of their social attributes: warmth, competence and discomfort from RoSAS [26]. They were also asked to select only one agent for hypothetical future games, between the prosocial and the selfish one. Participants evaluated the quality and satisfaction with the team at a group level by reporting their subjective trust [4] and identification towards the team [94]. Finally, we analysed the objective cooperation rate of the participants by



(a) Main effects of the embodiment on group identification ($p < 0.001$) and group trust ($p = 0.517$).

(b) Interaction effect between embodiment and the agents' strategy on the perception of warmth ($p = 0.035$).

(c) Interaction effect between embodiment and the agents' strategy on the perception of competence ($p = 0.001$).

Figure 5.2

counting how many times they chose to cooperate, out of 4 decisions.

5.2.5 Procedure

The procedure of the experiment consisted of: (A) a briefing about the experiment, game rules and a demonstration of how to play the game; (B) a session of 5 rounds of the collective dilemma with two autonomous agents on the team; (C) a final self-assessment questionnaire; and (D) a debriefing. However, the embodied condition was collected in the lab, while the disembodied was run on an online platform. The main difference on the procedures was in the initial briefing (A): for the embodied condition, a researcher explained the game rules in a verbal manner; and for the disembodied condition, it was an online tutorial that participants had to read. Participants in the embodied condition took approximately 30 minutes to complete the experiment while in the disembodied condition they took approximately 40 minutes. Also, the participation reward, given at the debriefing phase (D), was a fixed amount of 5$ for participants in the disembodied condition and a cinema ticket of a similar price for participants in the embodied condition, regardless of their individual outcomes in the game.

5.2.6 Sample

In the embodied condition, 70 participants were recruited in person at the facilities of an energy company. For the disembodied condition, a total of 89 participants were recruited on MTurk. However, after validating the attention checks, out of those 89, only 41 valid participants remained. As such, out of the initial sample of 159, the statistical analysis was executed on a data set of 111 participants. Regarding the independent variable of the outcome, we collected 57 that won the game (35 in the embodied condition and 22 in the disembodied condition) and 54 that lost the game (35 in the embodied condition and 19 in the disembodied condition). Overall, in terms of gender, there were 47 females, 63 males and 1 participant chose "other". The ages ranged from 22 to 63 ($M = 36.060, SD = 10.816$).

As our two embodiment conditions were collected separately i.e., one in the lab and the other on MTurk, we checked for differences in the distributions of demographic variables. Both the age and gender of participants was similar across the four experimental groups, revealing identical demographics in our conditions. Furthermore, several experimental studies analysed the validity of comparing MTurk with laboratory subjects [143; 99; 14], even for economic games [5], showing comparable results in these two samples. Overall, given this body of evidence we are convinced that it is quite unlikely that different results would be obtained if we opted for not using MTurk for the disembodied condition.

5.3 Results

5.3.1 Group measures

Group Identification

The reliability of the group identification scale was $\alpha = 0.813$, which indicates a good internal consistency. Through a two-way ANOVA, we observed that embodiment had a significant medium-sized main effect on the participants' group identification perception ($F(1, 106) = 15.589, p < 0.001, r = 0.358$). Participants perceived higher group identification when collaborating with the embodied agents ($M_{Embodied} = 3.866, SD = 1.320$; $M_{Disembodied} = 2.932, SD = 1.482$) revealing that group identification was positively influenced by the presence of embodiment (Figure 5.2a). Game result also had a significant medium-sized main effect on the participants' group identification perception ($F(1, 106) = 24.913, p < 0.001, r = 0.436$). Participants perceived higher group identification when winning the game ($M_{Victory} = 4.056, SD = 1.367$; $M_{Loss} = 2.956, SD = 1.321$). No significant interaction effect was perceived between embodiment and game result ($F(1, 106) = 3.769, p = 0.055, r = 0.184$).

Group Trust

The reliability of the group trust scale was $\alpha = 0.945$, which indicates an excellent internal consistency. A two-way ANOVA showed no significant main effect of the embodiment ($F(1, 106) = 0.423, p = 0.517, r = 0.063$, Figure 5.2a), nor of the game result ($F(1, 106) = 3.201, p = 0.076, r = 0.170$). Moreover, no significant interaction effect was found between embodiment and game result ($F(1, 106) = 0.733, p = 0.394, r = 0.084$).

5.3.2 Perception of the agents

Warmth

The internal consistency for the warmth dimension was close to acceptable $\alpha = 0.693$. No main effects were observed for embodiment ($F(1, 106) = 2.574, p = 0.112, r = 0.155$) nor game result ($F(1, 106) = 1.422, p = 0.236, r = 0.114$). However, a large-size main effect of the agents' strategy could be observed ($F(1, 106) = 75.661, p < 0.001, r = 0.645$). The warmth attributed to the prosocial was significantly ranked higher than the warmth attributed to the selfish over all conditions ($M_{prosocial} = 4.364, SD = 1.152$; $M_{Selfish} = 3.201, SD = 0.941$).

A two-way interaction effect (see Figure 5.2b) between strategy and embodiment was also found ($F(1, 106) = 4.566, p = 0.035, r = 0.202$). This indicates that the warmth values of the two strategies, prosocial and selfish, were affected differently by the embodiment. We performed a contrast analysis between different levels of embodiment for each agent, the prosocial and the selfish, using Mann-Whitney U tests. For the selfish agent, no significant difference was found on its warmth when it was embodied ($U = 1407, Z = -0.171, p = 0.864, r = 0.016; M_{Selfish-E} = 3.218, SD = 0.833$) and disembodied ($M_{Selfish-D} = 3.171, SD = 1.110$). However, the attribution of warmth to the prosocial agent was significantly different when it was embodied ($U = 1089, Z = -2.015, p = 0.044, r = 0.192; M_{prosocial-E} = 4.162, SD = 1.115$) and disembodied ($M_{prosocial-D} = 4.703, SD = 1.147$). It reveals the warmth of the prosocial agent increased when that agent was disembodied, increasing as well the discrepancy with the warmth attributed to the selfish agent.

No significant two-way interaction effect was observed between strategy and the game result ($F(1, 106) = 1.243, p = 0.267, r = 0.110$) and no three-way interaction effect was observed between strategy, embodiment and game result ($F(1, 106) = 1.964, p = 0.164, r = 0.134$).

Competence

The competence dimension had a good internal consistency ($\alpha = 0.809$). A small-sized main effect was observed for embodiment ($F(1, 106) = 9.376, p = 0.003, r = 0.285$) indicating that the disembodied agents' competence was significantly ranked higher than the embodied agents' competence ($M_{Disembodied} = 4.827, SD = 1.231; M_{Embodied} = 4.342, SD = 1.093$). A small-size main effect was also observed for game result ($F(1, 106) = 4.405, p = 0.038, r = 0.200$), indicating that participants who won the game significantly ranked the competence of the agents higher than participants who lost the game ($M_{Victory} = 4.672, SD = 1.320; M_{Loss} = 4.367, SD = 1.144$). Additionally, a large-sized main effect of strategy could be observed ($F(1, 106) = 145.730, p < 0.001, r = 0.761$) indicating that the competence attributed to the prosocial was significantly higher than the competence attributed to the selfish over all conditions ($M_{prosocial} = 5.344, SD = 1.318; M_{Selfish} = 3.701, SD = 1.208$).

A large-size two-way interaction effect was also observed (see Figure 5.2c) between strategy and embodiment ($F(1, 106) = 41.909, p = 0.001, r = 0.532$), revealing that the competence attributed to embodied and disembodied agents was affected differently by their strategies. More precisely, the competence attributed to the selfish agent was significantly lower when it was disembodied ($U = 1049, Z = -2.373, p = 0.018, r = 0.225; M_{Selfish-D} = 3.378, SD = 1.357$) than when it was embodied ($M_{Selfish-E} = 3.894, SD = 1.075$). However, the attribution of competence to a prosocial agent significantly increased when it was disembodied ($U = 449.5, Z = -5.974, p < 0.001, r = 0.570; M_{prosocial-D} = 6.276, SD = 1.104$) compared to when it was embodied ($M_{prosocial-E} = 4.790, SD = 1.111$).

A medium-sized two-way interaction effect between strategy and game result was also found ($F(1, 106) = 11.633, p = 0.001, r = 0.315$). This indicates that the competence values of the two strategies, prosocial and selfish, were affected differently by the game result. When participants won the game, the competence attributed to the prosocial was significantly higher than to the selfish ($Z = -4.065, p < 0.001, r = 0.539$). However, that difference was even more noticeable when they lost the game ($Z = -5.999, p < 0.001, r = 0.816$).

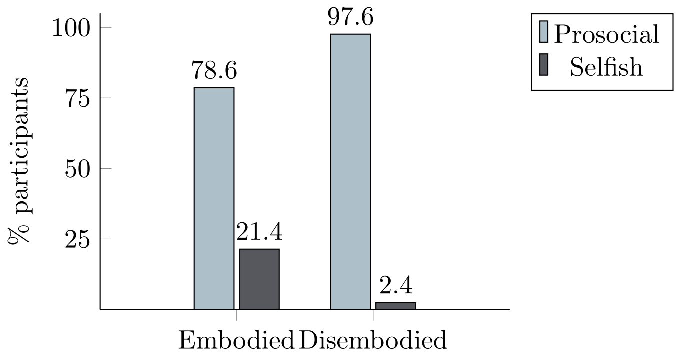


Figure 5.3: Association between the embodiment and the number of participants that preferred each agent ($p = 0.005$).

Finally, no significant three-way interaction effect was observed between strategy, embodiment and game result ($F(1, 106) = 1.298, p = 0.257, r = 0.110$).

Discomfort

The items of the discomfort dimension had a good internal consistency ($\alpha = 0.817$). No main effects were observed for embodiment ($F(1, 106) = 0.004, p = 0.949, r < 0.032$) nor game result ($F(1, 106) = 2.697, p = 0.104, r = 0.158$). Nevertheless, a small-size main effect of strategy was found ($F(1, 106) = 59.160, p < 0.001, r = 0.598$), indicating that the discomfort attributed to the prosocial was lower than the discomfort attributed to the selfish ($M_{Selfish} = 2.946, SD = 1.364; M_{prosocial} = 1.846, SD = 1.042$).

Additionally, we neither observed a significant two-way interaction effects between strategy and embodiment ($F(1, 106) = 0.555, p = 0.458, r = 0.071$), nor between strategy and game result ($F(1, 106) = 2.008, p = 0.159, r = 0.138$), nor a significant three-way interaction effect between strategy, embodiment and game result ($F(1, 106) = 0.007, p = 0.933, r < 0.032$).

Partner Selection

Using a Fisher's Exact test, we found a significant association between the level of embodiment and the number of times participants selected each agent for future games (Figure 5.3, $\chi^2(1) = 7.558, p = 0.005$). In the embodied condition, 78.6% of participants (55 out of 70) selected the prosocial robot for future games, while in the disembodied condition, that preference was even more salient with 97.6% of participants (40 out of 41) choosing the prosocial agent.

5.3.3 Cooperation Rate

Finally, through a Mann-Whitney U Test, we analyzed the actions performed by the participants when playing the game alongside embodied and disembodied agents. No significant differences were observed ($Z = -0.565, p = 0.572, r = 0.055$), meaning that the amount of times players decided to cooperate did not significantly varied when playing alongside embodied ($M = 2.400, SD = 0.875$) or disembodied agents ($M = 2.439, SD = 0.950$). Overall, the average cooperation rate was 2 times out of 4.

5.4 Discussion

In **H1**, we have hypothesised that *participants would perceive the team more positively in the embodied condition.* The results of our study only partially validate this hypothesis, since a significant difference in the predicted direction was only found for group identification, but not for group trust. It seems the degree of identification towards the team can be positively affected by the embodied presence of the robots and the richer social interaction they might provide. The group trust, on the other hand, seems to have been less affected by our manipulation of embodiment.

Due to the close relation between these two measures, which are usually correlated, it was surprising that their results were not aligned. However, the subjective nature of these group perceptions can reflect both the execution of the task or a certain degree of satisfaction with the social interaction of the team. Therefore, our findings suggest group identification might be more associated with the social interaction, while group trust might reflect more the actions performed for the task. We speculate that in our collaborative game, due to the uncertainty factor, participants have evaluated the task execution by considering the strategies of all members. For instance, the fact that the team had an unconditional defector in both conditions could have damaged participants' trust.

Furthermore, considering the results we obtained in these group measures, we would like to highlight their importance as an additional consideration for future research on human-robot group interactions. Due to the known relation between trust and the performance of a robot [70], it would be interesting to also analyse the relation between group trust and the perception of competence of the team as a unified entity. In our user study, this group measure can be hard to infer only from the individual perceptions of competence of each agent. Not only because we have not assessed participants' perceptions of self competence, but also because this group perception might not necessarily follow a linear structure. In the future, we would like to further explore this idea by considering more group measures.

We predicted in **H2a** that *regardless of the embodiment, the perception of the prosocial agent will be more positive than the perception of the selfish agent.* We found support for this hypothesis as the prosocial agent was indeed rated with higher levels of both warmth and competence, and lower levels of discomfort when compared to the selfish agent.

In **H2b**, we have hypothesised that *the perception of the prosocial agent will be more positive in the embodied condition.* Our results do not support this hypothesis and we actually verified the opposite situation. The prosocial agent was rated with higher levels of both warmth and competence when it was disembodied than when it was embodied.

A similar situation occurred with **H2c**, in which we hypothesised that *the perception of the selfish agent will be more negative in the embodied condition.* The attribution of competence revealed an opposite result to what we have predicted, with the embodied selfish agent being rated as more competent than the disembodied selfish agent. Regarding the other social attributes, warmth and discomfort no significant differences were found.

Our hypotheses regarding the perceptions of the agents, both **H2b** and **H2c**, were based on the assumption that the social behaviour of the embodied robots would increase the expectations for them to have an adequate behaviour regarding social and moral norms. This assumption is supported by the actual behaviour people display around physically-embodied agents compared to virtually-embodied agents [9; 91]. As a result, we expected that the higher degree of embodiment would actually highlight the strategies of the agents, by improving the perception of a prosocial strategy and by worsening the perception of a selfish strategy. On the contrary, our results suggest that the robotic embodiment reduced, rather than increased, the gap between the perceptions of the two strategies. It seems that the lack of an embodiment made participants more focused on the agents' strategies and, conversely, in the embodied condition participants did not pay as much attention to the strategies due to other factors being considered. In fact, the social behaviours of the embodied agents were designed in a neutral tone so that their only difference was the strategy to play the game. However, it might be the case that participants felt that there was a disconnect between their strategy and the way they utilized its embodiment to convey that strategy. Another possible interpretation is that the embodiment made the actual social interaction more salient over the task-related aspects, as if a new dimension in their evaluation came into play. Again, as the robotic agents, both the prosocial and the selfish, employed similar non-verbal behaviours, it might have in turn decreased the difference between the perceptions of their social attributes.

In **H3**, we have predicted that *participants will be more prosocial in the embodied condition*. We expected that either the higher degree of embodiment would inhibit participants' selfishness, or the lower degree of embodiment would promote it. However, our results did not support this hypothesis.

On the one hand, we believe that there might be a relationship between this result and the one obtained for group trust. In other words, the fact that participants equally trusted their teams in both embodiment conditions, might justify the absence of significant differences in the cooperation rate. Again, the presence of an unconditional defector on the team, the selfish agent, might have played a significant role in the strategy of the participants. On the other hand, we would like to point out that our experimental scenario contained only four decision points, which might be insufficient to notice significant differences. Future investigations using similar collective dilemmas should contain more rounds to clarify this issue.

Finally, regarding the independent variable of the outcome, its main effect on group identification and competence, as well as the interaction effect with the agents' strategy are both inline with previous findings.

5.5 Concluding Remarks

This project explores the role of embodiment in collective dilemmas with artificial agents. In this type of task, opting for cooperation usually incurs some cost or loss for oneself and, therefore, defection might seem more enticing. Within a wider research vision of creating agents that enhance human cooperation [112], this project is focused on the effect that the embodiment of those agents can have. In particular, we compared robots that use several embodied modalities (i.e., gaze, posture, animations) with disembodied agents that socially interact only with textual comments. Such comparison is not only novel, but also relevant considering the widespread use of intelligent disembodied voice assistants (e.g., Cortana, Google's assistant, Siri). *What is the impact of adding embodied affordances to interactive social agents in cooperative tasks?*

To analyse this question, we conducted an empirical study where participants were asked to form a team with two autonomous agents. One of the agents embraced a prosocial strategy (i.e., by always choosing to cooperate), while the other adopted a selfish strategy (i.e., by always opting to defect). We also manipulated the level of embodiment in two extreme levels: a robot and a disembodied agent. We analysed how participants (1) evaluated the team, (2) perceived each agent, and (3) their degree of cooperation towards the team.

Firstly, regarding group measures, participants identified themselves more with the team of embodied agents than with the team of disembodied agents. However, no significant difference was found on the trust towards the team between the two embodiment conditions. Secondly, in terms of the agents' perceptions, the main effect of the strategy in the three social attributes revealed that participants clearly distinguished between the most and the least collaborative strategies, by positively rating the prosocial agent and by negatively rating the selfish one. Nevertheless, when comparing the social attributes across different levels of embodiment, the results suggest that the strategies of the embodied agents were partially masked by their social behaviours. Thirdly and finally, our results did not show a significant difference from our embodiment manipulation on participants' degree of cooperation).

The following considerations for human-robot teams can be derived from the obtained results. Firstly, a misalignment of valences between the actions of an agent and its social behaviours can change the perceptions of those agents. In our user study, the social behaviour of both robotic agents was similarly neutral in valence, which might have caused a negative impact on the perception of a prosocial agent and improved the perception of a selfish agent. This constitutes an important consideration for the design of social behaviour in embodied agents. As future work, it would be interesting to explore if an agent executing prosocial actions but displaying negative social behaviours, rather than neutral ones, would be rated more negatively than an agent executing selfish actions but displaying positive social behaviour.

A second consideration we can draw from our results is that the embodied affordances of the agents seem to lead people to consider additional aspects during the interaction. Those aspects seem to be related to the social interaction that is beyond task-related actions, and can be established through those embodied affordances, i.e. gaze or facial expressions. According to our results, the embodiment may drive both positive and negative facets of human-robot interaction. On the one hand, it showed a positive impact on the degree of identity with the team. Group identification constitutes an important process of group settings and is reported, in the social sciences, as being further associated with other relevant group measures, such as team performance [136]. On the other hand, it was also able to partially mitigate the strategies that each agent took in the collective dilemma. If we focus on the perceptions of the selfish agent, it raises a negative perspective on the impact of embodiment. The presence of embodied features seems to have decreased the emphasis on its non-cooperative actions and to have improved how humans perceived that agent. This concern is particularly relevant if we consider that, in the future, socially intelligent agents can portray the selfish intentions of their owners, e.g. an individual or a company. By having these considerations in mind, practitioners can design adequate affordances and behaviours of social agents according to the goals and nature of the task.

One limitation in the results we obtained is that they might not generalise to virtual agents as our embodied condition refers to a robot, i.e. physically-embodied agent. To address this issue, a follow-up study should accommodate another condition with virtually-embodied agents. A final limitation that is common in the analysis of group interactions is the nested design between the embodiment and the strategy. The perception of an agent with a certain strategy can differ according to the strategy of the other agent on the team. Considering in our design we only had opposite strategies, to address this issue in future studies, other conditions where both agents employ the same strategy would be necessary.

Overall, embodied interactions seem to introduce more aspects into the evaluation of agents, possibly related to a higher social presence and/or a richer social interaction. Although they might improve how humans relate with agents in team settings, if these social aspects can positively mask questionable or even immoral decisions, a dark side of embodiment may emerge.

5.6 Other follow-up works

In a follow-up work, we investigated the role of agents' transparency, by having agents that reveal and tell the strategies they adopt in the game, in a manner that makes their decisions transparent to the other team members. Using the same game, "For The Record", we designed a 3 x 2 between-subjects experimental study, manipulating 3 different strategies, and the transparency of the agents (with vs without). The results showed an interaction effect between the agents' strategy and transparency on trust, group identification and human-likeness. Our results suggest that transparency has a positive effect in terms of people's perception of trust, group identification and human likeness when the agents use a tit-for-tat or a more individualistic strategy. In fact, adding transparent behaviour to an unconditional cooperator negatively affects the measured dimensions. The full description of this follow-up work can be found in [146].

Chapter 6

A model of Group-based Emotions

The current chapter describes a project that explored *emotional cohesion*, which is the intensity of the members to express group feelings. We were particularly interested in the robot's emotional expression of group feelings. One of the motivations behind this idea was the fact that the autonomous robotic agents described in the previous chapters (Chapters 3 and 4) were already developed on top of an architecture for emotional agents.

According to the group identity theory, there is a particular group of emotions that reflects a high level of identification with the group, which are called the group-based emotions. This idea led us to formulate the following research question: can the expression of group-based emotions by a robotic teammate increase people's identification and trust towards the team?

The current chapter starts with a brief background on group-based emotions in Section 6.1. Then, we propose an implementation for the group identification process in Section 6.2 and, in Section 6.3, we provide a detailed description of how it was applied in our card game scenario. The chapter proceeds with a user study to evaluate the model in Section 6.4 and a detailed discussion of the results in Section 6.5. Section 6.6 concludes with some considerations and remarks.

6.1 Background on Group-based Emotions

Group-based emotions are believed to be a result of self-categorisation and appraisal theories of emotions [133]. When the social identity of the perceiver leads him to think of himself as a *group member* rather than just an individual, events affecting his in-group may elicit such type of group emotions. Naturally, these emotional reactions occur during intergroup interaction and, according to Smith [133], the salience of social identity is elicited by 3 factors: (1) the presence of out-group members, (2) the perception of similarities with the in-group members, and (3) the competition between groups. A quite common example of a group-based emotion is the feeling of pride or shame for our favourite sports team.

Recently, Goldenberg et al. have proposed a process model of group-based emotions [67] that extends the "modal model" of Gross & Thompson [68]. The extension of a general model of emotion to account for group-based emotions is supported by the notion that these emotions, although different in their appraisals, have the same basic structure as regular emotions.

In this model, the process of generating group-based emotions, starts with an attention allocation to a given situation or stimuli. Then, the relevance of the situation being attended is conditioned by the group identity that the individual associates himself at the moment and how strong is that association. Finally, the process ends with an emotional response that may range from the expression of emotions to organised actions.

During intergroup interactions, there is a close relation between group-based emotions and group identification. Group identification was known as being an antecedent of group-based emotions until Kessler and Hollbach [89] have presented how group-based emotions can influence group identification and, therefore, how they can have bidirectional causality. Their results point to the fact that in-group identification may increase or decrease according to the type of emotions and the target being the in-group or the out-group. In particular, happiness towards the in-group and anger towards the out-group increase in-group identification, the same way happiness towards the out-group and anger towards the in-group decrease in-group identification. The authors also reveal a positive correlation between the intensity of emotions increasing identification and the final identification level, which may lead to a positive feedback loop that explains the development of a collective action frame.

6.2 A model of group-based emotions for social robotic characters

Social robotic characters can generate group-based emotions in similar contexts as humans if they are equipped with mechanisms that properly approximate the human psychological process that leads to these emotions. Our model (see Figure 6.1) is aimed in this direction with its mechanisms being grounded on the recent psychological model of group-based emotions, proposed by Goldenberg et al. [67].

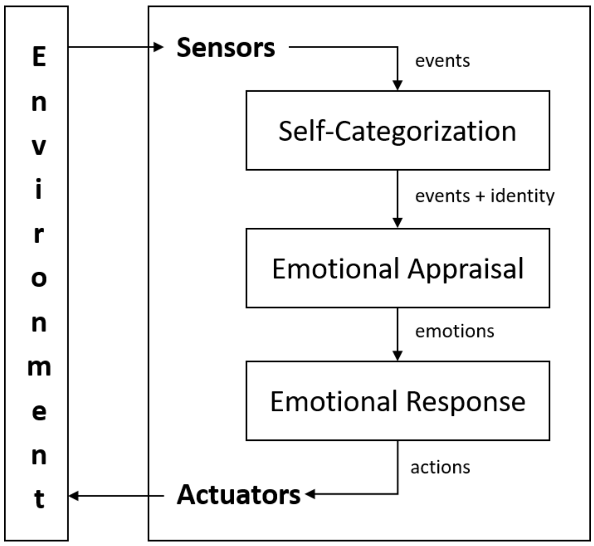


Figure 6.1: Diagram of the group-based emotions model.

The first component of the model is the *Self-Categorisation* component, which is responsible for managing the current context of the interaction as well as the social groups that are present in that context, if any, and their members. These elements constitute a social layer on top of the physical reality that is being perceived by the robot's sensors. Based on the Self-Categorisation Theory [77; 147], when the robot detects a presence of an out-group then its own group identity will become more salient.

The emotional appraisal is the second component of the model. This component is responsible for generating emotions in response to the events that occur within the current social context. An event can correspond to a performance of an action or to a change in a property of the environment. For each event perceived, the emotional component performs a series of value judgements about that event in relation to the robot. Then, a set of emotions are synthesised in accordance to those judgements. In emotional psychology, these judgements are referred to as appraisal variables with different theories of emotion proposing different sets of variables [103]. For instance, according to the OCC theory [110], when someone judges an event to be desirable for him or her, that person is likely to experience joy afterwards in proportion to the level of desirability attributed. The same theory also proposes that when someone performs an action that is considered blameworthy than that person is inclined to feel shame. However, if the blameworthy action is performed by another person, then a reproach emotion is felt instead by the observer.

While many researchers have already been able to integrate an emotional appraisal component in a social robotic architecture, the innovative aspect of our model lies in the notion that our appraisal component is capable of considering a social group as the actor of an event even if, in reality, all actions are being performed by individuals. This is the result of introducing the *Self-Categorisation* component before the appraisal takes place in order to determine whether the robot sees itself and others acting based on their individual or their group identity. In the latter case, actions performed by individuals that are sharing the same group identity in the current context are appraised as if they are actions performed by the robot itself. Consequently, in a context where the robot is performing a team-based activity and one of its partners performs a blameworthy action, the appraisal component will generate in the robot a group-based emotion of shame, rather than a reproach emotion towards its partner.

Finally, the last component of the model is the *Emotional Response* component, which is responsible for managing how the robot expresses the emotions that result from the appraisal process. This process must take into account the different possibilities that are afforded by the robot's embodiment. Assuming the robot has the ability to change its facial expression and body posture then these are matched to the current emotional state of the robot. In addition to non-verbal signals, the dialogue acts chosen by the robot are also influenced by its emotions. This is particularly important when trying to convey social emotions such as admiration or pride that can be hard to distinguish from more basic emotions as the ones proposed by Ekman [53], using only non-verbal modalities.

Algorithm 2 Group-based emotions generation process

```
while true do
    self ← Robot.Name
    e ← Sensors.PerceiveNewEvent()
    SG ← ContextManager.GetSalientSocialGroups()
    if SG ≠ ∅ then
        g ← IdentityManager.SelfCategorisation(SG, self)
        if e.ResponsibleAgent ∈ g then
            e.ResponsibleAgent ← g.Name
            self ← g.Name
    AV ← Appraisal.DetermineVariables(e)
    E ← Appraisal.GenerateEmotions(AV, self)
    se ← StrongestEmotion(E)
    for all c ∈ Actuators.GetEmotionChannels() do
        Express(se, c)
```

Algorithm 2 provides a more detailed view of how these three components work together in a continuous cycle to create group-based emotions. As shown, the cycle starts by defining the parameter *self* as equal to the robot's name, which should be a unique identifier of the robot. The next step is to check for the perception of a new event e by the robot's sensors. Then, the set of salient social groups

SG is determined, taking into account the last event perceived. This set will be empty in the case where the current context or activity has no salient groups. If SG is not an empty set then the group g is selected as the one that the robot identifies the most with. Afterwards, the algorithm checks if the robot/person who caused the event to occur (*e.ResponsibleAgent*) is a member of the same group g. When that is the case, the event's responsible agent is replaced by the name of group g as well as the parameter *self*. At this point, the *Self-Categorisation* component of the model ends and the *Appraisal* component begins. Based on the event perceived e, the set of appraisal variables AV is now calculated. Subsequently, those variables are used to generate a set of emotions E, from which the strongest emotion *se* is extracted. The strongest emotion is considered to be the one with the highest intensity value. Finally, the emotion *se* is expressed by all the available channels the robot has to express an emotion.

Note that the algorithm presented here is aimed to be general enough so that it can be applied in different domains with different robots and using different models of appraisal and self-categorisation. As such, the use case that is described in the following section is to be viewed as just one possible way in which the proposed model can be fully implemented.

6.3 Card game Scenario

In order to explore how group-based emotions influence human-robot teams, we used the described model to create social robots that are able to autonomously behave according to different levels of self-categorisation. We decided to choose a task with two adversarial teams to make the in-group and out-group distinction more prominent. The task chosen was *Sueca*, a card game played by exactly four players divided in two opposing teams. Both cooperation with the partner and competition with the opponents have a strong effect on the game result since its goal is to beat the score of the other team. The players should play according to their understanding of their partners' game state. Another motivation for choosing a card game as our task comes from the fact that several studies [106; 71; 100; 130] have demonstrated that card games are a successful activity for creating engagement in a human-robot interaction that is designed to be primarily social.

Sueca is a trick-taking game containing the element of chance, and is played with a standard deck. The players start with ten cards in their hands, and have to decide which one to play during each trick. The suit of the first card played in a trick is considered to be the leadsuit of that trick and all players must follow it. If a player does not have any card from the leadsuit, then he or she is allowed to use a card with the trump suit, which is more valuable than the other suits. This is considered as "cutting the trick". The team that wins the trick consequently collects the sum of all its points. This occurs when any of its players played the highest trump card or the highest from the leadsuit when there is no trump card on the table.

With the aim of exploring mixed human-robot teams, our *Sueca* scenario consists of two social robotic players that are both paired with a human partner. These robotic players are fully autonomous and their behaviours were developed on top of the SERA ecosystem [119]. Players interact with

Table 6.1: Examples of speech acts performed by each robot according to the game state and the strongest appraised emotion.

	Robot that expresses individual-based emotions				Robot that expresses group-based emotions			
	Admiration	Reproach	Pride	Shame	Admiration	Reproach	Pride	Shame
Partner increased trick score	I am impressed with your move!	—	—	—	—	—	We are the best!	—
Partner decreased trick score	—	With that move, I cannot win.	—	—	—	—	—	We were not so good this time...
Robot increased trick score	—	—	I played incredibly well!	—	—	—	I am impressed with our performance!	—
Robot decreased trick score	—	—	—	I am so ashamed of my move...	—	—	—	Sorry partner, for this unfortunate move.

the robots over a touch table using physical cards. The game application is then responsible for recognising the cards and forward all game events to the artificial players. Each of these artificial players is composed by an emotional agent [47] and an AI [34] that are responsible for the emotional behavioural responses and the game computations, respectively. Then, a behavioural planner schedules the non-verbal behaviours to the animation engine and the verbal behaviours to the Text-To-Speech (TTS).

Although the two robots have the same embodiment and the same interaction affordances, one of them generates group-based emotions and the other generates individual-based emotions. This is accomplished in the following manner. Following Algorithm 2, each robot starts by identifying itself with its name, among the four possible players ($\{P1, ..., P4\}$). The robot then perceives game events associated to the plays made by players and corresponding changes in the game state such as "$Event(P3, IncreasePoints(Trick, 11))$". In the context of *Sueca* there are two salient social groups, which correspond to the two teams playing ($SG = \{T1, T2\}$).

The following step of *SelfCategorisation* differentiates our two robots. Although this process in humans can be highly complex, our implementation for this particular scenario follows a rather simple logic. Namely, for the robot that expresses group-based emotions this step returns the team to which the robot belongs. For the example above, assuming that the robot is $P1$ and also that $P1 \in T1$, $T1$ will be assigned to g. Then, verifying the responsible agent ($P3$) belongs to that same social group, $T1$ will be assigned to both $e.ResponsibleAgent$ and $self$. In other words, the robot attributes the responsibility of the perceived event to its social group, instead of its partner, and appraises the event on behalf of the group.

On the contrary, the robot that expresses individual-based emotions was implemented without the self-categorisation step, which will lead it not to identify itself as a member of a social group, regardless of the groups contained in SG. In other words, the robot without self-categorisation attributes the responsibility of the perceived event always to an individual, instead of a social group, and consequently appraises events as an individual.

The appraisal and emotional generation steps of the proposed model were implemented with FAtiMA [47], an existing emotional architecture that is based on the OCC theory [110]. In our scenario, two appraisal variables of the OCC theory were used, namely, *Desirability* and *Praiseworthiness*. Using the rule-based mechanism present in FAtiMA, the value of these two variables are determined in the following manner. All the plays made by an opponent that increase/decrease the points of the trick for the robot's team are considered to be desirable/undesirable by an amount that is linearly proportional to the number of points increased/decreased. In the case of the plays made by the robot or its partner, if they increase/decrease the points of the trick they are seen as praiseworthy/blameworthy, also in a linearly proportional manner.

Once the appraisal variables are determined, the step of generating emotions occurs. Based on the OCC model, a positive/negative value of desirability generates an emotion of joy/distress. As for praiseworthiness, if it is positive, an emotion of pride or admiration is generated based on the event's responsible agent. Pride in the case where it matches the robot's *self* parameter and admiration otherwise. Inversely, if praiseworthiness is negative then an emotion of shame or reproach is generated instead. Shame in the case where it matches the robot's *self* parameter and reproach otherwise. Finally, the emotional response of our robots is based on the current strongest emotion. The responses may be utterances of verbal and non-verbal behaviour or physical postures.

6.3.1 Emotional Response

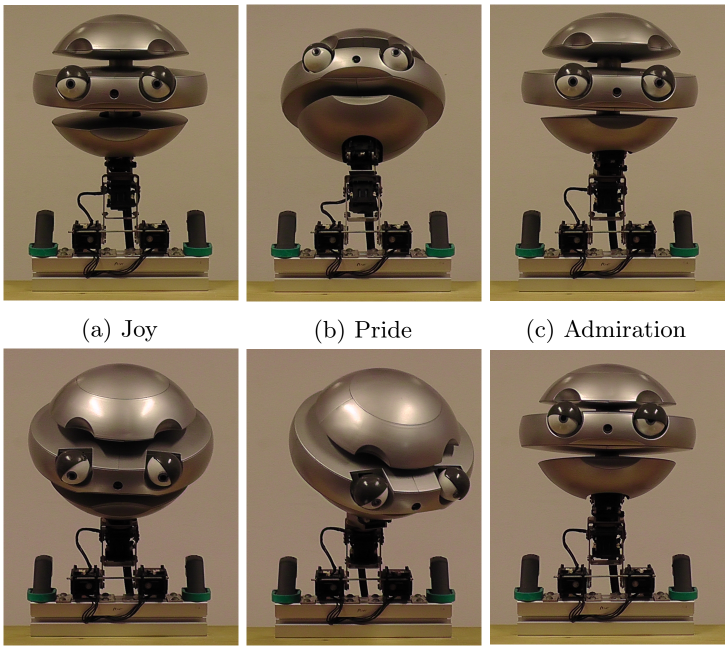
The behaviours and emotional responses of each robot may differ according to the properties of its embodiment. In our scenario, the EMYS robot [87] was used and its behaviours include utterances (i.e., dialogue acts, gazes, and animations) and physical postures.

Utterances

In total, two sets of utterances were created to convey the emotional state of the robots. One set was used by the robot with individual-based emotions and the other was employed by the robot with group-based emotions. In order for their dialogue acts to convey the nature of the emotion generated by the model, they may contain inclusive pronouns (e.g., "we", "us", "our") to express group-based emotions, or individual pronouns (e.g., "I", "me", "you") to suggest individual-based emotions. This distinction becomes more relevant to unambiguously distinguish emotions such as "Individual Pride" and "Group Pride". A similar language adaptation was employed by Brave et al. [21] to differentiate between self- vs. other oriented emotions.

The full list of utterances is available in [32] and contains about 100 utterances for each robot. Each robot has an extensive repertoire of sentences in order for the robots not to repeat themselves during the interaction, which lasts approximately 30 minutes with both robots interacting fully autonomously. Moreover, we tried to differentiate the wording of the sentences for both robots (while maintaining the same meaning), instead of solely changing the pronouns. Otherwise, it would seem that the robots were repeating each other and this could significantly break their believability as two independent social actors.

The selection of utterances takes into account the current game state and their strongest emotion. To be more precise, there are four particular game states where the robots have different emotions and, therefore, result in different behaviours, as evidenced by Table 6.1.



(a) Joy (b) Pride (c) Admiration

(d) Distress (e) Shame (f) Reproach

Figure 6.2: Postures embodied on the EMYS robot for each emotion.

Note that in the case where an opponent is playing, the robotic characters will only interact verbally in emotionally neutral situations (e.g. greeting, ask to shuffle). In emotionally charged situations, such as when an opponent cuts the trick, the robots will only express an emotion of joy or distress through non-verbal animations. Therefore, independently of the self-categorisation level of the robots, they react similarly towards the out-group.

We have validated all the utterances by asking three coders to classify the associated strongest emotion, among the possible 6 of joy, distress, pride, admiration, shame, or reproach. The average pairwise Cohen's kappa value was k=0.82, revealing good agreement between the coders and that the chosen sentences accurately reflect their intended emotions.

Postures

Another non-verbal behaviour of the robotic characters was the embodiment posture. While utterances and simple animations convey a reaction upon a particular game event, a posture is used to convey for a longer period of time the emotional state of the robots. During the turn of other players, the robots choose the strongest emotion to adapt their postures, see Figure 6.2.

6.4 User study

Using the Sueca scenario described in the previous section, we conducted a user study where two participants form a team with a robot to play the card game. We used the same embodiment – EMYS robot – although the robots expressed different types of emotions: group-based or individual-based emotions. Based on the previously discussed findings from intergroup interactions in human-human

Figure 6.3: Experimental setting for the user study.

[89; 4] as well as in human-robot [93; 71; 46; 149] scenarios, we expect to check the following hypothesis:

H1: Participants will have a stronger Group Identification with a robotic partner that expresses group-based emotions.

H2: Participants will have a more positive perception of a robotic partner that expresses group-based emotions.

H3: Participants will have a higher degree of Group Trust with a robotic partner that has group-based emotions.

6.4.1 Procedure

Each session of the experiment had two participants and took approximately 45 minutes. Participants read a consent form and were briefly introduced to the game activity. The game rules were described and two researchers played a sample game with the participants over a regular wooden table. Then, participants were randomly assigned to one robotic partner – which could express either group-based or individual-based emotions–, and moved to the touch table, where they played three consecutive games with the robots, see Figure 6.3. In the beginning one researcher explained how the game works over the touch table and the initial setup of assigning the robots' cards. At the same time, another researcher would set two cameras for video recording if participants had authorised. A researcher stayed in the experiment room until the end of the first trick. After this, both researchers left the room and let participants play the three games. Finally, they were given a questionnaire and the experiment ended with a cinema ticket being randomly awarded to one of the participants.

6.4.2 Measures

Independently of the robotic partner, all participants answered a final questionnaire that contained the following measures:

- **Group Identification** [94] with the Portuguese adaptation [117] to assess the in-Group Identi-

Table 6.2: Average number of utterances per emotion for the Robot with Group-based Emotions (RGbE) and the Robot with Individual-based Emotions (RIbE).

	Neutral	Admiration	Reproach	Pride	Shame	Joy	Distress	**TOTAL**
RGbE	6	0	0	15	4	4	5	34
RIbE	6	7	2	6	2	4	8	35

fication with their robotic partner;

- **Godspeed Questionnaire** [15], using the dimensions of Anthropomorphism, Animacy, Likeability, and Perceived Intelligence regarding their robotic partner;

- **Group Trust** [4] to assess the perceived trust by the participants regarding their team in the game;

- **Demographic questions**, i.e. gender, age, previous interaction with the EMYS robot, proficiency level in the *Sueca* card game.

Both Group Identification and Group Trust measures used 7-points Likert Scales, ranging from Strongly Disagree to Strongly Agree. The Godspeed Questionnaire was assessed in 5-points semantic differential scale.

6.4.3 Sample

We recruited a total of 48 university students (33 males and 15 females) with ages ranging from 19 to 33 years old ($M = 25.02 \pm 2.98$). 25% of the participants had already interacted with the EMYS robot and 77.1% reported at least a medium proficiency level of playing the *Sueca* card game.

6.4.4 Results

In the 24 collected sessions, the team of the robot with group-based emotions won 10 times, lost 11 times, and tied 3 times. Table 6.2 shows the number of utterances performed by each robot for each emotion and evidences a balanced average of total utterances per session. As expected, in both conditions there were more utterances for positive emotions in general than for negative ones. This can be attributed to the fact that players naturally avoided making bad plays as they were trying to win. Given that some of the following measures did not show normal distributions (as indicated by the Shapiro-Wilk test), we used the non-parametric Mann-Whitney U-test to compare the independent samples.

Group Identification

A reliability analysis was carried out on the satisfaction and solidarity dimensions comprising 4 and 3 items, respectively. Cronbach's alpha of 0.83 for the satisfaction dimension, and 0.81 for the solidarity dimension indicate a high level of internal consistency for the scale of Group Identification with this specific sample.

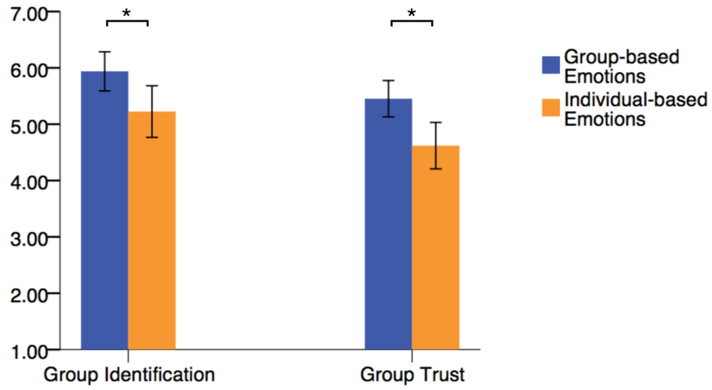


Figure 6.4: Group Identification and Group Trust averages attributed to each team with a robot that expresses either group-based or individual-based emotions. (*$p < 0.05$)

We compared the level of Group Identification perceived by the participants towards each robot. Participants had significantly higher levels ($U = 175.5, p = 0.02, r = 0.335$) of Group Identification towards the robotic partner with group-based emotions ($M = 5.94 \pm 0.17$) than towards the robotic partner with individual-based emotions ($M = 5.22 \pm 0.22$), see Figure 6.4.

Additionally, a Spearman's rank-order correlation was run to determine if there was a relationship between the number of points of the team and the Group Identification metric. The rationale was to check if having more points as a consequence of winning the game would also have a positive effect on this dimension. The correlation was non-significant ($r_s = 0.153$, $p = 0.30$), which suggests that these two factors are independent.

Perception of the robots

To check whether the expression of individual or group-based emotions was influencing the perception of the robots, we compared the levels of Anthropomorphism, Animacy, Likeability and Perceived Intelligence attributed to each robot. Results showed no significant differences in the Anthropomorphism ($U = 276, p = 0.80$), Animacy ($U = 275, p = 0.79$), and Perceived Intelligence ($U = 200, p = 0.07$) levels attributed to robotic partner with group-based emotions ($M_{ant} = 2.97 \pm 0.14; M_{ani} = 3.57 \pm 0.13; M_{pi} = 3.91 \pm 0.14$) and the robotic partner with individual-based emotions ($M_{ant} = 2.95 \pm 0.17; M_{ani} = 3.50 \pm 0.14; M_{pi} = 3.47 \pm 0.17$), see Figure 6.5. However, participants attributed significantly higher scores of Likeability ($U = 142.0, p = 0.002, r = 0.437$) to the robotic partner with group-based emotions ($M = 4.33 \pm 0.11$) than the robotic partner with individual-based emotions ($M = 3.49 \pm 0.21$), see Figure 6.5.

Additionally, a Spearman's rank-order correlation was run to determine the relationship between Group Identification and the four measured dimensions of the Godspeed Questionnaire. There was a strong, positive, and statistically significant correlations between Group Identification and Anthropomorphism ($r_s = 0.529$, $p < 0.01$), Animacy ($r_s = 0.318$, $p = 0.03$), Likeability ($r_s = 0.606$, $p < 0.01$), and Perceived Intelligence ($r_s = 0.595$, $p < 0.01$).

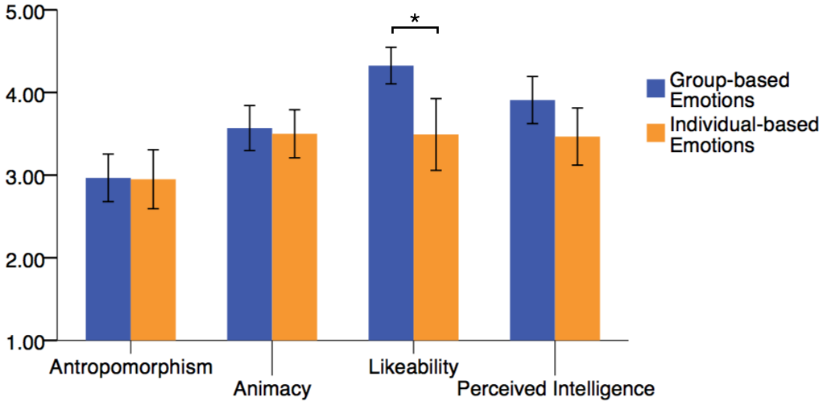


Figure 6.5: Averages of Godspeed's dimensions attributed to each robot. (*$p < 0.05$)

Group Trust

A reliability analysis was carried out on the Group Trust scale comprising 7 items, respectively. Cronbach's alpha of 0.74 indicates a high level of internal consistency for this scale with this specific sample.

To check whether the portrayal of individual or group-based emotions was influencing the group trust, we compared the Group Trust levels of each human-robot team. Results showed a statistically significant difference of Group Trust ($U = 148, p < 0.01, r = 0.417$). Partners of the robot with group-based emotions reported significantly higher levels of Group Trust ($M = 5.45 \pm 0.16$) than participants partnering the robot with individual-based emotions ($M = 4.62 \pm 0.20$), see Figure 6.4.

Additionally, a Spearman's rank-order correlation was run to determine the relationship between Group Trust and the number of points of the team. Again, there was no significant correlation between the Group Trust and the number of points ($r_s = 0.158$, $p = 0.28$).

6.5 Discussion

Overall, the following conclusions can be drawn from the obtained results. **H1** predicted a stronger Group Identification with a robotic partner that expresses group-based emotions. Indeed, the expression of group-based emotions lead to higher levels of Group Identification. This result seem to be coherent with previous findings from the social psychology, where group-based emotions are an antecedent of Group Identification [89].

Our results also partially support **H2**, which predicted a more positive perception of a robotic partner that expresses group-based emotions. Although participants perceived both robots similarly in terms of Anthropomorphism, Animacy, and Perceived Intelligence, they rated the robotic partner that expresses group-based emotions with significantly higher levels of Likeability. Our expectation was in-line with previous findings [93; 71], where the perceived group identity of a robot had a significant impact on how the robot is perceived. However, their manipulation of group identity referred to the extremes of belonging to the in-group or the out-group, stressing a stronger difference of Group Identification when compared to our setting. Therefore, even with a statistically significant difference of the Group Identification (H1), it was not prominent enough to elicit differences on the Anthropomorphism, Animacy, and Perceived Intelligence. Nevertheless, tendencies can be inferred in both Figure 6.5 and in the positive and significant correlations between the level of Group Identification and the measured dimensions of the Godspeed Questionnaire.

According to **H3**, we expected a higher degree of Group Trust with a robotic partner that has group-based emotions. This was confirmed as participants that were in the team of the robot that expresses group-based emotions attributed a higher level of trust towards the team. We believe this result is relevant for the emerging field of Human-Robot Teams, as trust constitutes one the most important constructs to support effective interaction and cooperation [70]. Furthermore, trust and social identity have been postulated as antecedents of positive team performance [4].

6.6 Concluding Remarks

As human-robot groups become more prevalent, it is important to explore new ways of improving their interactions and creating effective collaborations. Therefore, this work provides a first exploration of group-based emotions in social robotic partners and tries to understand their impact on human-robot teams. The contributions are two fold. First, a model for generating group-based emotions on social robots is proposed. This model enabled the development of two distinct robotic characters that express either individual or group-based emotions. Second, it also contributes with a user study, where two fully autonomous robots formed two human-robot teams to play a competitive card game. We compared participants' perceptions, Group Identification, and Group Trust towards their robotic partners. The results show that the expression of group-based emotions by a social robotic partner led participants to rate it as more likeable, to trust it more as a team member and also to identify a stronger social group. Overall, group-based emotions were able to emphasise important aspects of the human-robot collaboration, as trust and group identification, revealing a promising role on the design of social robotic team partners.

One limitation of our study is related with the *Sueca* scenario, as the generation of emotions is strongly related to the game itself. On the one hand, it is a very natural scenario for the interaction of autonomous robots with two people. On the other hand, it may result in an unbalanced number of positive emotions when compared to negative emotions in each game. As future work, it is important to explore scenarios where more negative events also occur. Additionally, the study reported here only analysed mixed human-robot teams of two. In the future, it would be important to analyse the effects of group-based emotions in larger teams as well. Finally, the fact that the robots used different wordings in their utterances could have generated a potential confound. For this reason, it would be interesting to analyse the effect of group-based emotions in a scenario with no use of verbal dialogue.

Our final argument is that social emotions such as shame or pride, which go beyond more basic emotions [53], should be given more attention in the creation of social robots, particularly in human-robot teams. The work presented here takes a step in this direction as it employs a more rich emotional model enabling social robots to autonomously produce more diverse emotional behaviour.

6.7 Other follow-up works

In a follow-up work, we explored people's emotional responses to robotic teammates according to theories of human-human group stereotypes [40]. Our goal was to analyse how different levels of warmth and competence displayed by autonomous robots can influence the users' perceptions, emotions and willingness to interact again in the future. To achieve this purpose, we used the same card game scenario in which two autonomous robots displayed different levels of warmth and competence (high and low). The results of our study suggested that participants are able to create stereotypes about robots based on their level of warmth and competence and that different perceived combinations of these traits result in different emotional responses. In addition we observed that warmth and competence can predict future intention to work and that the participants' perceptions of robots have a certain degree of endurance. The full description of this follow-up work can be found in [108].

Chapter 7

Gaze Behaviors in Multi-party Settings

Structural cohesion represents the roles, norms, or interactions among members of a group, and the topology of those structures is associated with the degree of cohesion of the group. Taking inspiration from this dimension of cohesion that is usually explored with network analysis, we speculated on the degree of connectivity that a robotic teammate should consider in its perceptive skills. For instance, should a robotic teammate perceive communicative acts only towards itself? Or should it also perceive communicative acts between other pairs of team members?

This project focused on the nonverbal gaze behaviour in a multi-party team setting. We created a silent coordination task by envisioning a shared workspace between humans and robots — e.g., an assembly line — in which coordinated behaviour is mainly achieved through non-verbal behaviours — e.g., due to a noisy environment. In such situation, the gaze behaviours of a robotic teammate during a silent coordination task hold two cognitive functions of gaze, *monitoring* and *communicative*. The former function aims at collecting information from interlocutors, and the latter focuses on the communication of interpersonal attitudes [1]. Therefore, we set out to investigate how a robot's gaze behaviour can display different levels of responsiveness, portraying these monitoring and communicative gaze functions. Moreover, *are those gaze behaviours of a robotic teammate able to enhance the coordination and fluency of its human-robot team*?

To address these research questions we developed an autonomous gaze system for a robotic teammate in multi-party settings, detailed in Section 7.1. The system perceives the gaze behaviours of human teammates and either responds to gazes towards the robot itself (Responding Gaze condition), or responds to any gaze by the teammates (Attentive Gaze condition). Section 7.2 describes as an evaluation from a third-person perspective to assess whether the gaze patterns that emerge immediately before a coordinated action in a silent human-human-robot team interaction could lead to a higher perception of coordination and fluidity.

7.1 Gaze in a Multi-party Silent Coordination

It is well established that gaze direction, gaze contact, and mutual gaze are modes of conveying and gathering information to and from others [88]. Research has also showed that blinking can have a similar role to back channeling and thus it serves to provide feedback in conversations [41]. For these reasons eye movements are of particular interest for coordinated collaboration [54].

As Kendon describes [88], the perceptual activity of an individual can function in two ways. A person A can control the degree to which she *monitors* person B, by looking or not looking at B. Those same actions can also be used by A to *communicate* internal states to B. The amounts of gaze contact, mutual gazes, sustained gaze, or gaze aversion, are emergent phenomena rooted on these two basic actions. In a coordination task (as is conversation), gaze direction functions as a *monitoring* and *communication* mechanism, as a floor-claiming device. In a collaborative setting it can aid shared attention and help to establish common ground [6].

In this project, we explore gaze in a less common social context of joint activity: a silent collaborative coordination task. This context brings out the function of gaze in social interaction as is linguistically independent. The robotic agent was programmed to autonomously handle human gaze perceptions, and trigger the appropriate gaze behaviors of the robot[1]. The set of behaviours of the robot was designed such that its perception mechanism is used to continuously acquire a model of the environment. Hence, human gaze behaviours are detected by the Head Pose Live program of the OpenFace library [11], which publishes the head pose and gaze direction at a rate of 25 fps. To classify the gaze target of each human, their current headpose and gaze direction are compared to a previous calibration of the interaction setup. The gaze targets of the robot are also set beforehand and they remained as static angles, assuming neither the humans nor the shared screen will change their position during the interaction. All modules of the robotic agent run on top of Thalamus middleware [118], and some belong to the SERA ecosystem [120], such as the animation engine and the behavior planner.

During social interaction the robot *monitors to where each human teammate is looking at*, then models its own gaze response using the following heuristics:

1. If the teammate X looks at the *Robot*, the robot will return the gaze so that "eyes would meet" (attempt to establish *mutual gaze*).

2. If the new gaze target of a teammate X is $Y \neq Robot$, it gazes at target Y (attempt to establish *joint attention*).

The rationale is simple. Through eye contact, the robot knows that it is affecting the human teammates in someway and can infer if they are making progress in their task [88]. The aforementioned behaviours (in combination with humans gaze behaviours) make emerge common phenomena, such as perception of changes in gaze direction, different needs to sample the environment for information acquisition (more gaze behaviours are positively linked to a higher need for information gathering [88]), looking away when mutual gaze is established and tentative to establish mutual gaze. Mutual gaze means that "one is being taken account off", which is important in coordination tasks. These resulting social behaviours are illustrated in Fig. 7.2 and they can signal different internal states that can impact teamwork coordination and fluidity.

Regarding the second rule, the robot's attention is intentionally-directed towards a person X's attention target, which not only does it reveal an understanding of X as a intentional agent [144] (i.e., theory of mind), but also reflects a shared intentionality between the robot and X. On the one hand, the robot following X's movements can be seen as a "mirror image", and it can help reaching synchronous behaviours. On the other hand, while blindly following X's attention in a multi-party interaction, the robot and X can share attention with one of the teammates, which might be perceived with additional intentions. For instance, [25] describes a state in which Y is "being looked at by mutiple [group members]" as a prestigious moment, and that the occurrence of such situations throughout a group interaction can later predict the leader of the group. As a result, we have further created an alternative version for the second rule mentioned above, which avoids establishing joint attention towards another teammate:

[1] https://github.com/GAIPS/the-mind

1. If the teammate X looks at the *Robot*, the robot will return the gaze so that "eyes would meet" (attempt to establish *mutual gaze*).

2. If the new gaze target of a teammate X is $Y \neq Robot$, it gazes back at target *Screen*.

7.2 Experimental Setup

The experimental study aims at exploring the impact of different gaze behaviours by a robotic teammate in a group interaction with humans. In particular, it addresses the research question of how *the perceived teamwork of a human-robot team is affected by the robot's responsiveness towards human gaze behaviours*. The evaluation was done through a third-person perspective, in which participants were asked to watch videos of the human-robot team interaction and to rate their perceptions of teamwork regarding each video. The following hypotheses reflect our expectations regarding the participants' subjective perceptions about the emergent behaviour of the team generated by the robot's responsiveness to the interaction.

H1 - A higher degree of gaze responsiveness by a robotic teammate will positively affect the perceived teamwork of its human-robot team.

H2 - A higher number of human teammates responding to gaze behaviors will positively affect the perceived teamwork of their human-robot team.

H3 - A human-robot team in which the two human teammates (2H) respond to gaze behaviors while the robot is staring at the screen will be perceived as having higher teamwork scores than a human-robot team in which one human and one robotic teammate (1H1R) respond to gaze behaviors while the second human is staring at his tablet.

The first hypothesis is based on the premise that the robot's gaze responsiveness is associated with the perception of understanding others' mental states. By reacting to human gaze behaviours, the robot would portray knowing where human teammates are gazing at and, to some extent, what their intentions are.

Our second and third hypotheses are based on two aspects. First, human teammates present lower reaction times, which demonstrate enhanced perceptive capabilities and a higher competence overall in this silent coordination task, compared to the robot. Second, we believe in-group bias towards humans will also favor theses differences in mixed human-robot teams.

7.2.1 Task and Video Scripts

The collective success in our silent coordination task is only achieved if the teammates perform their individual actions in the correct order. Nevertheless, each member does not know the exact actions that others have to do, and their coordinated behaviours should emerge only from non-verbal cues.

We examined a team of 2 humans and 1 robot performing the full task in a pilot. Fig. 7.1 depicts the setup of this task, in which there is a shared screen in the middle of the table, where the three

Figure 7.1: Silent coordination task in a triadic team composed of two humans and one robotic teammate.

teammates can observe the output of their actions and monitor the progress of the team. The humans' individual actions are executed in their own tablets, while the robot sends commands directly to the shared screen.

From the pilot, we collected examples of gaze behaviours among teammates immediately before coordinated actions took place. Those examples were used to create the video scripts of the 10 short scenes with gaze instructions for the human teammates (Fig. 7.2). The goal of these 10 scenes is to illustrate possible situations that may occur in this silent coordination task, precisely when teammates are required to actively achieve synchronization. The chosen fragments do not include touches from human teammates on their tablets in order to emphasise gaze behaviours. Note that the robot's behaviours are autonomous during video recordings.

We recorded 3 times each of the 10 scenes (one per condition) using two human actors to impersonate the gaze behaviours of the script. The robot's gaze behaviours varied between conditions. The gaze behaviours of human A were triggered by metronome beeps at a speed of 40 bpm (1.5s between beeps). Human B was asked to execute his gaze instructions slightly after human A. The procedure of our experiment, including the video recordings with the recruited actors, was approved by the Ethical Committee of our university. The two actors voluntarily impersonated humans A and B and signed the consent form regarding data management and usage.

7.2.2 Independent Variables

The study had two independent variables in a mixed design. First, we manipulated **gaze responsiveness of the robot** in a between-subjects design. The aforementioned 10 scenes were recorded 3 times, one per condition. In each condition the robot was reacting in one of the following three ways:

- Baseline (B) – The robot does not perceive and, therefore, does not respond to human gaze behaviours. It only blinks its eyelids.

- Responding Gaze (RG) - The robot perceives human gaze behaviours towards itself and responds by gazing back at that team member. It also blinks as in the Baseline condition;

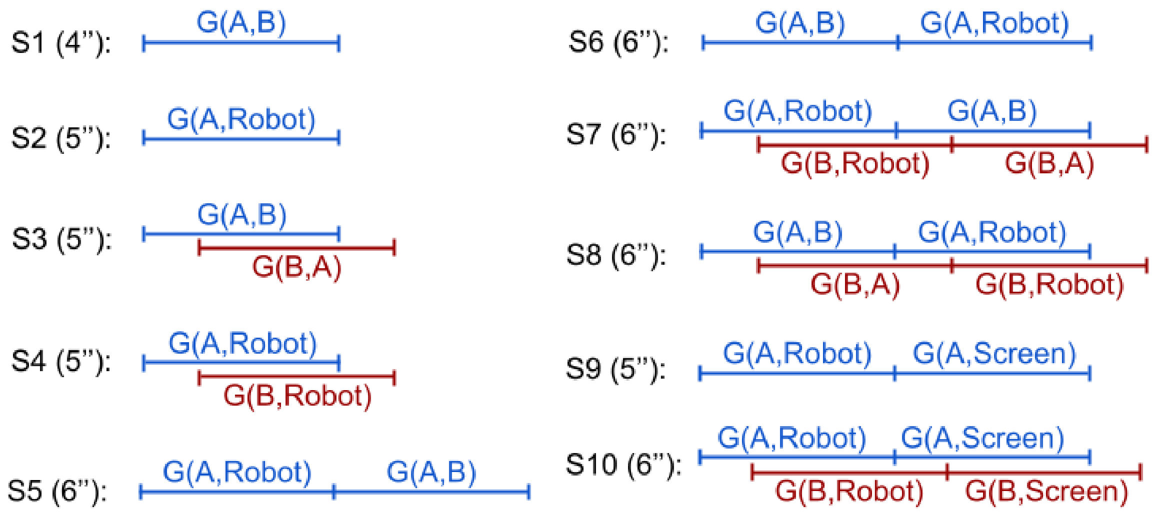


Figure 7.2: Gaze instructions for the human actors. $G(x, y)$ stands for x gazing at y. The human actors A and B start and finish all the scenes gazing at their own tablets. The robot starts all the scenes gazing at the shared screen in the middle of the table, and performs gaze behaviours according to the experimental condition.

- Attentive Gaze (AG) - The robot perceives human gaze behaviours towards any target and responds by either gazing back at the gazer, if the robot is the target, or gazing at the same team member that is currently being gazed by another. It also blinks as in the Baseline condition.

We also manipulated in a within-subjects design the **number of human teammates with gaze behaviours** in their script. According to video scripts on Tab. 7.2, half of the scenes only had gaze behaviours from one human teammate (scenes 1, 2, 5, 6, and 9), while the other half had gaze behaviours from the two human teammates (scenes 3, 4, 7, 8, and 10).

7.2.3 Dependent Variables

To assess the perception of teamwork we have used items of the subjective human-robot fluency metric proposed by Hoffman [73], namely the fluency, teammate traits, and cohesion measures. Additionally, we looked at the perceived synchronization and coordination of the team, based on questions made to listeners about singing ensembles [51].

Coordination was assessed with 3 items ($\alpha = 0.93$) and aims at capturing the perceived synchronization or, in other words, the sense of gazing at the right time. **Fluency** was assessed with 2 items ($\alpha = 0.72$) that consider the joint activity and its turn-taking aspects, and it specifically accounts for the contribution of the robot to the collaborative task. **Cohesion** was assessed with 5 items ($\alpha = 0.92$) and captures the work alliance and groupness between team members, by including understanding-, respect-, trust- and unity-related items. Finally, **teammate traits** were assessed with 3 items ($\alpha = 0.79$) focused on the positive traits that are required of a robotic teammate. The metric of teammate traits was only used to validate the first hypothesis.

7.2.4 Apparatus and Materials

The EMYS robotic head [87] operates autonomously in all conditions, according to the behaviour described in Sec. 7.1. Our setup used three machines, one to run the robotic agent that handles human gaze perceptions and triggers the gaze behaviors of the robot, and the other two to run two instances of Head Pose Live in order to maximize the performance of detecting the gaze behaviors of each human teammate separately. The cameras used to perceive human A and human B are a Papalook PA452 and an Intel RealSence Depth Camera D415, respectively. The setup also requires the round table and three tablets, two of them to simulate human input in the pretend task, and another one to serve as a shared screen for the team. Finally, the scenes were recorded on a third camera positioned outside the table and facing the robot. That perspective was chosen so that it was clear to notice where the robot was looking at. Additionally, although human teammates were captured slightly in profile, their gaze behaviours were also clearly perceived whilst other facial expressions were not salient from that perspective.

7.2.5 Procedure

Participants were collected on Prolific,[2] and were asked to answer a survey to assess their perceptions, from a third-person perspective, on the gaze behaviors of a human-robot team. Each participant was randomly assigned to one of the experimental condition (Sec. 7.2.2), and watched 5 videos of 10 possible scenes (Tab. 7.2), which were also randomly selected. The videos could be watched as many times as the participant wanted. For each video, the participants had to rate their subjective evaluation (using the items in Sec. 7.2.3). Participants were told the videos show small fragments of a collaborative task, in which the team is composed of 2 humans and 1 robot. Moreover, they were told: “the goal of the team is to successfully perform a coordination task. Coordination is achieved in silence and, therefore, the videos you will see have no sound”.

To control for the quality of sample, there were 3 attention checks after each video, with numbers from 1 to 5 as possible answers for the following questions: “How many robots are around the table?”; “How many people are around the table?”; and “How many members does the team have?”. The survey ended with some demographic questions and manipulation checks. Participants received 1 GBP upon completion,[3] and the average time to complete the survey was around 8.5 minutes.

7.2.6 Sample

We collected a sample of 180 participants stating that were fluent in English in their Prolific profiles. 17.2% were excluded as a result of failing at least 1 of the 15 attention checks, which left us with 149 participants to run the statistical analysis. Their age distribution ranges from 22 to 66 ($M_{Age} = 33.4 \pm 18.8$), and 43% identify with the female gender, 56% with male and 1% preferred not to answer.

There are 49 participants in Baseline condition, 46 in the Responding Gaze condition, and 54 in

[2] An online platform for participant recruitment [113].
[3] Based on 7.04 GBP cost for 1 hour of work.

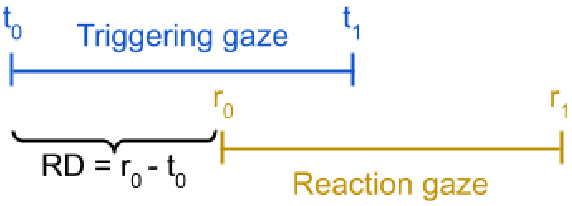


Figure 7.3: Illustration of the reaction delay (RD) that corresponds to the lag between t_0 of the triggering gaze and the r_0 of the reaction gaze.

the Attentive Gaze condition. As each participant saw 5 random videos of one condition, each video was seen, on average, 24.5 times in Baseline condition (min. 18 times), 23 times in Responding Gaze condition (min. 18 times), and 27 times in Attentive Gaze condition (min. 20 times).

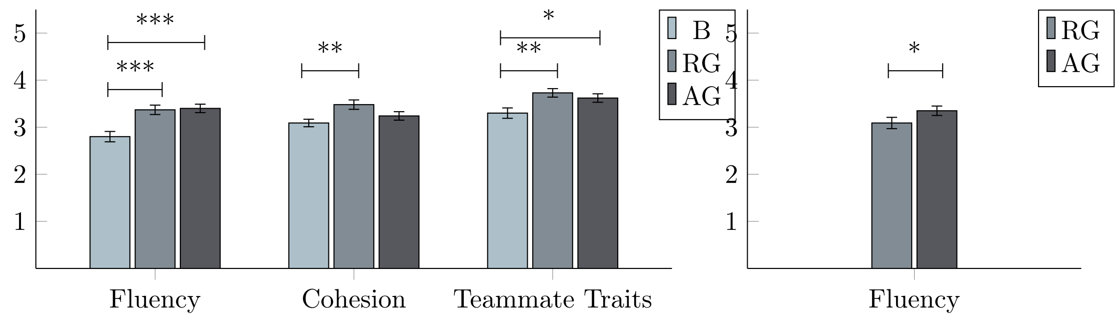
7.3 Results

According to the Shapiro-Wilk test, our dependent variables do not follow a normal distribution. As a result, we run non-parametric statistical tests. The comparisons between the three conditions of the robot's gaze responsiveness used the Kruskal-Wallis H test, and the followup pairwise comparisons used the Mann-Whitney U test. Comparisons among the two groups of the within-subjects variable (the number of human teammates with gaze behaviours) used the Wilcoxon Signed-Rank test. Finally, correlation analysis also used the non-parametric Spearman test.

7.3.1 Manipulation Checks

Manipulation checks (MC) served to verify whether the robot behaviour in all conditions was correctly perceived. To check the robot's behaviour in the baseline condition, we compared the agreement scores on the 1st item of MC, which was significantly different across the 3 conditions ($H(2) = 22.26, p < 0.001$). The pairwise comparisons revealed participants perceived the robot as less responsive in B than in RG ($U = 712, p = 0.001, r = 0.33$) and than in AG ($U = 655.5, p < 0.001, r = 0.44$). To check the perception of responsiveness to direct gazes towards the robot, we compared answers to the 2nd item and found significant differences on the agreement scores between the 3 conditions ($H(2) = 81.79, p < 0.001$). The pairwise comparisons revealed participants perceived the robot as more responsive to direct gazes in the RG compared to B ($U = 184.5, p < 0.001, r = 0.75$) and in AG compared to B ($U = 141, p < 0.001, r = 0.8$). To check the perception of responsiveness to any gaze behaviour, we compared answers to the 3rd item and we found the the agreement scores were significantly different between the 3 conditions ($H(2) = 64.12, p < 0.001$). The pairwise comparisons revealed participants perceived the robot as more responsive to any gaze in the AG compared to B ($U = 230, p < 0.001, r = 0.73$) and compared to RG ($U = 593, p < 0.001, r = 0.46$).

We also annotated the ground truth of gaze behaviours for each of 30 short videos. The goal was to compare if the reaction delays between the two actors, and between the robot and each actor were not different across conditions. A reaction delay (RD) corresponds to the lag between t_0 of the triggering



(a) (b)

Figure 7.4: (a) Differences of the conditions on the overall perceptions of fluency, cohesion and teammate traits for the 10 videos scenes. (b) Difference between the experimental conditions on the perception of fluency for the 6 video scenes where the gaze behaviour *G(A,B)* occurs. *$p < 0.05$ **$p < 0.01$ ***$p < 0.001$

gaze (X gazes at Y from t_0 to t_1) and r_0 of the reaction gaze (reaction to the driver behaviour from r_0 to r_1), where $r_0 > t_0$ (Fig. 7.3). We confirmed that the delay ($RD = r_0 - t_0$) between the two human actors was not significantly different across the 3 conditions ($H(2) = 0.14, p = 0.932$), as intended. The average reaction delay between human A and human B was $M_{delayA-B} = 0.52(s), SD = 0.19$. Moreover, the delays between human A and the robot, as well as between human B and the robot, in the RG condition were not significantly different from the same delays in the AG condition ($U = 32.5, p = 0.515$ and $U = 5, p = 0.286$, respectively). The average reaction delay between human A and the robot in both experimental conditions was $M_{delayA-R} = 1.19(s), SD = 0.14$, and between human B and the robot was $M_{delayB-R} = 1.84(s), SD = 0.20$.

7.3.2 Grouping the 10 scenes

We averaged the ratings for each video in each condition to compare the perceptions of the human-robot team in all the 10 scenes.

We did not find a significant difference on the perception of *coordination* between the three conditions ($H(2) = 2.28, p = 0.32$).We found, nonetheless, a significant effect of the robot's gaze responsiveness on the perception of *fluency* ($H(2) = 19.46, p < 0.01$), *teammate traits* ($H(2) = 8.87, p = 0.012$) and *cohesion* ($H(2) = 7.68, p = 0.021$), Fig. 7.4a.

The follow-up pairwise comparisons (full description in Tab. 7.1) allowed us to conclude that, when the robot was in the RG and AG conditions, the perception of its *teammate traits* and the team's *fluency* were higher than in the baseline condition. However, the *teammate traits* and the perceived *fluency* were not significantly different between RG and AG conditions. In other words, a **higher responsiveness** in the robot's gaze behaviour (either RG or AG) seems to **positively affect** its perceived *teammate traits* and the *fluency* of its human-robot team. A different pattern for the perception of *cohesion* was revealed in the pairwise comparisons. Participants rated the *cohesion* of human-robot team with **higher** values than the baseline condition only when the robot exhibited **Responding Gaze** behaviours (and no significant difference was found between B and AG conditions).

	Pairwise comparisions	Means and SD
Fluency	B < RG $U = 674.5, p = 0.001, r = 0.35$	$M_B = 2.80$, $SD = 0.81$; $M_{RG} = 3.37$, $SD = 0.70$; $M_{AG} = 3.40$, $SD = 0.63$
	B < AG $U = 699.5, p < 0.001, r = 0.41$	
	RG ∼ AG $U = 1132.5, p = 0.448$	
Teammate Traits	B < RG $U = 763.5, p = 0.007, r = 0.28$	$M_B = 3.30$, $SD = 0.78$; $M_{RG} = 3.73$, $SD = 0.58$; $M_{AG} = 3.62$, $SD = 0.64$
	B < AG $U = 968, p = 0.019, r = 0.23$	
	RG ∼ AG $U = 1133.5, p = 0.453$	
Cohesion	B < RG $U = 765.5, p = 0.007, r = 0.28$	$M_B = 3.09$, $SD = 0.55$; $M_{RG} = 3.48$, $SD = 0.68$; $M_{AG} = 3.24$, $SD = 0.65$
	B ∼ AG $U = 1120, p = 0.18$	
	RG ∼ AG $U = 1005.5, p = 0.102$	

Table 7.1: Statistical tests for the pairwise comparisons between conditions for the measures of fluency, teammate traits and cohesion, when considering the ratings of the 10 videos together. Means and standard error for each measure per condition.

7.3.3 Grouping scenes with G(A,B)

Based on the results of the previous analysis, which does not fully support **H1**, we grouped only the 6 scenes in which human A gazed at human B (scenes 1, 3, 5, 6, 7, and 8). The presence of at least one gaze behaviour between the two human teammates is what triggers an attentive gaze behaviour by the robot in AG condition, highlighting the difference between conditions RG and AG.

For this set, we compared the perceptions of *teamwork* between conditions RG and AG. We did not find a significant difference of the robot's gaze behaviour on the perceptions of *coordination* ($U = 11130, p = 0.438$), *teammate traits* ($U = 1157, p = 0.556$), nor on *cohesion* ($U = 1168.5, p = 0.611$). We found, however, a significant effect on the perceived *fluency* ($U = 942, p = 0.038, r = 0.21$, Fig. 7.4b). For the particular scenes where the attentive gaze behaviours were triggered, participants perceived the team as **significantly more fluent** in the AG condition ($M_{FluencyAG} = 3.35, SD = 0.74$) compared to the RG condition ($M_{FluencyRG} = 3.09, SD = 0.83$).

7.3.4 Grouping scenes by the number of human teammates

To analyse the impact of the number of human teammates with gaze behaviours, which was our within-subjects factor, we averaged participants' ratings into two variables depending on the number of human teammates with gaze behaviours in each of the 5 videos they watched.

We found a **significantly large effect** of the number of human teammates on the perception of coordination ($Z = -9.043, p < 0.001, r = 0.74$). The videos in which both human teammates had gaze behaviours were rated **higher** for *coordination* ($M_{Coordination2H} = 3.55, SD = 0.87$) than the videos in which only one human teammate had gaze behaviours ($M_{Coordination1H} = 2.70, SD = 0.93$). Similarly, we found two significant medium effects of the number of human teammates on the perception of

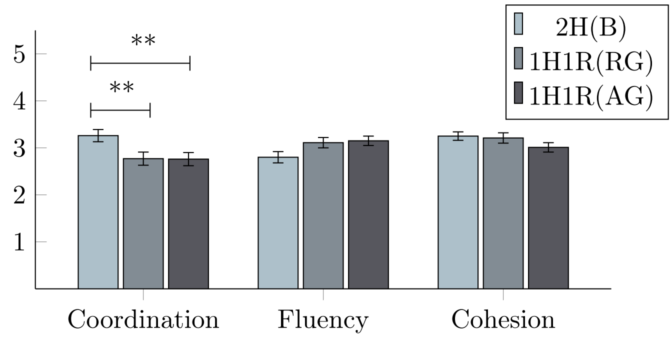


Figure 7.5: Differences between conditions on the perceptions of coordination, fluency and cohesion by considering scenes where only 2 teammates had gaze behaviours. In the baseline condition, the two human teammates had gaze behaviours while the robot was staring at the screen. In the two experimental conditions, one human teammate and the robot had gaze behaviours while the other human teammates was staring at his tablet.

fluency ($Z = -5.861, p < 0.001, r = 0.48$) and *cohesion* ($Z = -7.923, p < 0.001, r = 0.65$). Participants reported **higher** levels of *fluency* and *cohesion* in the scenarios where two human teammates had gaze behaviours ($M_{Fluency2H} = 3.37, SD = 0.87$ and $M_{Cohesion2H} = 3.50, SD = 0.70$), compared to the videos where only one human teammate had gaze behaviours ($M_{Fluency1H} = 2.98, SD = 0.84$ and $M_{Cohesion1H} = 3.01, SD = 0.74$).

7.3.5 1H1R vs. 2H scenes

The 30 videos created from the manipulation of our two independent variables include scenes where only one human and one robot (1H1R) had gaze behaviours (while the second human teammate was staring at his own tablet), and scenes where only the two human teammates (2H) had gaze behaviours (while the robot was staring at the shared screen). The 1H1R scenes were rated by participants in both RG and AG conditions, while the scenes with 2H were only shown to participants in B condition. As a result, we analysed the effect of the robot's gaze responsiveness on the average ratings of the scenes where exactly two teammates had gaze behaviours. For participants in the condition B, we consider scenes 3, 4, 7, 8, and 10 (2H) ratings, while for participants in the conditions RG and AG, we consider scenes 1, 2, 5, 6, and 9 ratings.

There was no significant difference between conditions on the perception of *fluency* ($H(2) = 5.632, p = 0.06$), nor on the perception of *cohesion* ($H(2) = 5.328, p = 0.07$). We found, however, a statistically significant difference of the three conditions on the perception of *coordination* ($H(2) = 11.082, p = 0.004$), as shown in Fig. 7.5. From pairwise comparisons, which are fully detailed in Tab. 7.2, we found that the human-robot team was perceived as more coordinated when the two humans were gazing at their teammates (2H) than when a human and a robot (1H1R) were gazing at their teammates.

	Pairwise comparisions	Means and SD
Coord.	2H(B) > 1H1R(RG) $U = 737, p = 0.004, r = 0.30$	$M_{2H(B)} = 3.26$, $SD = 0.88$;
	2H(B) > 1H1R(AG) $U = 890, p = 0.004, r = 0.28$	$M_{1H1R(RG)} = 2.77$, $SD = 0.93$;
	1H1R(RG) $\sim$ 1H1R(AG) $U = 1240, p = 0.989$	$M_{1H1R(AG)} = 2.76$, $SD = 1.00$

Table 7.2: Statistical tests for the pairwise comparisons between conditions for the perceived coordination, when considering 1H1R vs. 2H scenes. Means and standard error for each measure per condition.

7.3.6 Correlation analysis

As discussed in Sec. 7.3.1, we extracted the reaction delays for both humans and robot, which individually and overall may have an impact on each dependent variable ratings. As a result, we run an additional correlation analysis to assess any relationship between the delays and the dependent variables. Beyond the three previously described delays (human A-human B, human A-robot, and human B-robot), we calculated the average delay between all the pairs and named it average team delay. This is relevant because there are interactions with more gaze behaviours and thus more prone to interaction between the delays and dependent variables.

We found three weak positive correlations between the team delay and *fluency* ($r_s = 0.307, p < 0.001$), *teammate traits* ($r_s = 0.262, p = 0.001$), and *cohesion* ($r_s = 0.167, p = 0.041$). In other words, the higher the average reaction gaze delay between teammates, the higher the perception of fluency, robot's teammate traits and cohesion is for the human-robot team. All the remaining 13 correlations (with the delay between each pair) were not statistically significant. Although counter-intuitive, research has shown that such delays instead of amplifying perceptual-motor errors, may in fact facilitate synchronization in chaotic action sequences [150]. Future work will address whether this hypothesis is true for the perception of fluency and coordination in groups.

7.4 Discussion

In **H1**, we posited that *a higher degree of gaze responsiveness by a robotic teammate will positively affect the perceived teamwork of its human-robot team.* This hypothesis was partially supported as the two experimental conditions in which the robot had a higher degree of responsiveness (RG and AG) did have a positive effect on the perception of team fluency and robot's teammate traits compared to the baseline condition. Nevertheless, we did not find evidence that the attentive gaze behaviours in the AG condition have strengthened the perception of a cohesive human-robot team compared to having only responding gaze behaviours.

The autonomous gaze behaviours of the robot in RG and AG were different in 6 out of 10 scripted scenes. These 6 scenes contain gazes among the two human teammates—the ones that trigger the attentive gazes by the robot in AG condition. To double-check if the difference between RG and AG conditions was only noticeable in those particular scenes, we run an additional analysis (Sec. 7.3.3) and concluded that adding attentive gazes to the robotic teammate positively affected the perceived fluency of its human-robot team. Nonetheless, the attentive gaze behaviour seems to be more nuanced than expected as we could not find an effect on the perceived coordination, cohesion, nor on the teammate traits of the robot.

Our **H2** predicted that *a higher number of human teammates responding to gaze behaviors will positively affect the perceived teamwork of their human-robot team.* Our results support this hypothesis. A human-robot team in which the two humans display gaze behaviours is perceived as more coordinated, more fluent, and more united than a human-robot team in which only one human teammate displays gaze behaviours. A particularly interesting aspect of this result is the high effect sizes, compared to the other reported results. It seems this independent variable has highly affected the perception of teamwork of a human-robot team. However, a follow-up question arises on whether this result holds true for different team sizes and different proportions of humans and robots on the team.

In **H3**, we hypothesized that *a human-robot team in which the two human teammates (2H) respond to gaze behaviors will be perceived with higher teamwork scores than a human-robot team in which one human and one robotic teammates (1H1R) respond to gaze behaviors.* We found partial support for our hypothesis in the perception of coordination. Surprisingly, we did not find evidence that the gaze behaviours between 2H positively affected the perceived fluency and cohesion of the human-robot team compared to the gaze behaviours between 1H1R. An important consideration is that, while the items on coordination were previously used to assess human-human coordination, the subjective measures of fluency and cohesion were specifically created for human-robot collaborative scenarios. On one hand, the measures of fluency and cohesion might have failed to capture other relevant aspects of human-human interaction. On the other hand, the robotic teammate might have successfully reached a comparable gaze behaviour to the human teammate on those subjective scales. However, to verify such claim we would require a larger sample.

7.5 Implications

By looking at the results for the three hypothesis together, a more salient distinction between our dependent variables becomes apparent. Starting with the perception of fluency, the higher levels of responsiveness by the robot positively affected the perceived fluency of its human-robot team. More importantly, this measure showed a positive impact of the robot's attentive gaze behaviours compared to the robot that only had responding gaze behaviours. Additionally, we did not find evidence in our user study that the gaze behaviours between 2H can outperform the gaze behaviours between 1H1R in terms of the perceived fluency. It seems this measure might be more susceptible to consider the robot's contribution as any other teammate. Moreover, it reveals that attentive gaze behaviours can positively affect the sense of turn-taking and joint action in a team, as well as responding gaze behaviours at a lower level.

The responding gaze behaviours had a positive impact in the perception of cohesion when compared to the baseline condition in the overall score of the 10 scripted scenes. Compared to the other dependent measures, the perception of cohesion in this silent coordination task might be hindered by the inferred intention of the gaze behaviours. Although few claims can be drawn from our results, we speculate that the attentive gaze behaviours by the robot might have introduced other nuanced intentions into the perceived cohesion of a human-robot team, compared to a robot with only responding gaze behaviours.

Finally, coordination seems to have been highly affected by the gaze behaviours of the human teammates, regardless of the robot's responsiveness level. Not only did we not find support for the impact of the robot's responsiveness on perceived coordination, but also it seems that human teammates displaying gaze behaviours has a large effect on the same variable. We speculate that the robot's reaction delay, which was 2 times (and sometimes 3 times) longer than the reaction delay between humans, might impact the sense of "doing the right thing at the right time" that coordination metric were trying to capture. Mainly because most delays lead to a mismatch between the robot and human teammate gaze behaviour, causing a gazing mismatch. We decided to explore this idea in a follow-up study as discussed in the next section.

Overall, increasing the gaze responsiveness level of the robot revealed a promising avenue to design gaze behaviours within multi-party human-robot teams, both with responding gazes, that respond to direct gazes towards the robot, and with attentive gazes, that respond to gazes towards other teammates. An important consideration is that our results reflect a third-person perspective. Although a legitimate generalisation would require a similar assessment in the first-person perspective, recent findings support similar perceptions of gaze behaviours by first- and third-persons [114]. We also argue that within group interactions, humans often have to appraise interactions between the other group members.

7.6 Concluding Remarks

This project contributes to the existing literature on robotic gaze behaviours in multi-party team interactions—specifically, collaborative settings where a robot shares the workspace with human teammates and they either share the same goal or have interdependent goals. The autonomous gaze system we propose, as well as its experimental evaluation, shed some light into the role of gaze behaviours in this type of settings, by investigating *how the responsiveness level of the robot's gaze behaviours affects the perceived coordination and fluency of its team.*

Generally, the results of our user study revealed that higher responsiveness levels of the robot's gaze behaviours positively affect how the robot is perceived as a teammate, as well as the perceived fluency of its team. The perception of fluency was also sensitive to the difference between attentive gazes and responding gazes. We also found that perceptions of teamwork were highly affected by the number of human teammates with gaze behaviours. Finally, by looking at scenes with gaze behaviours by only 2 teammates (while the third one stared at an object), we only found differences in the perceived coordination among 2 humans compared to perceived coordination among 1 human and 1 robot.

We would like to highlight three possible future avenues for the current work. First, understanding the impact of the robot's reaction time and, ideally, decreasing the current mismatch between the triggering human gaze and the robot's reaction. Secondly, exploring the equivalent teamwork perceptions when the gaze behaviours are triggered by the robot, instead of by the human teammate. Lastly, the current setup and gaze strategies should be tested in a natural and less controlled user study.

Lastly, the work described in this chapter consists of an adaptation to the restrictions of running in-person experiments during the current pandemic of SARS-CoV-2. The initial idea for the scenario of this work was to use the card game "The Mind", a collaborative silent coordination game.

Chapter 8

Conclusions and General Discussion

This PhD thesis describes our work towards the creation of robotic teammates for multi-party settings. In this thesis, our main research problem was:

How can we endow a robotic teammate with social capabilities to improve the cohesive alliance in a multi-party setting with humans?

Our approach to address this problem was inspired by the multi-dimensionality of the cohesion construct, which states that there are five dimensions from which cohesion can flourish. Each of the five main projects in this thesis explores one of the dimensions of cohesion in multi-party settings between human(s) and robot(s), and analyses its impact on the shared sense of unity by the group members.

8.1 Summary of conclusions

- **Project I -** The membership preferences towards robotic teammates might be affected by the social traits of the robot, by people's personal characteristics, such as competitiveness, and also by the performance of the team. Hence, the goal orientation theory can be used to design the social behaviours of robotic teammates in multi-party competitive settings in order to support human-robot teamwork attraction or social cohesion.

- **Project II -** On the one hand, the portrayal of a low degree of collective cohesion by a robotic partner was negatively identified only when the performance of the team was compromised. Negative outcomes increased the people's awareness of what decisions robots took and also affected the reported identification with the team. On the other hand, a robotic teammate displaying a high degree of collective cohesion can be perceived more positively in terms of its social attributes, regardless of the game result. Generally, group-oriented decisions by social robots can foster collective cohesion.

- **Project III -** While comparing the portrayal of different levels of collective behaviour by the robotic teammates between the embodied and disembodied condition, the social aspects of embodied agents have positively masked selfish behaviours. It seems the affordances of a robotic teammate, when it is fully embodied, introduce other more facets to the interaction. Nonetheless, those embodied features were also able to positively affect people's reported identification with the multi-party group, which reveals the importance of embodied interactions to the collective cohesion of a human-agent team.

- **Project IV -** We developed a model of group-based emotions for social robots, which enables robotic teammates to autonomously express group feelings with their team. We also evaluated the model when compared to robotic teammates expressing individual-based emotions, and we showed its positive impact on the group identification of human teammates. Overall, robotic teammates can express emotional cohesion through group-based emotions and, in turn, foster collective cohesion.

- **Project V** - We proposed a model of gaze behaviours for robotic teammates in multi-party settings. We showed that the perceived fluency (by third-persons) of a team was higher when the robotic teammate employed attentive gazes, compared to when the robot reacted only with responding gazes. More broadly, the non-verbal communication structure assumed by a robotic teammate in a multi-party silent task can affect the perception of teamwork.

Considering the conclusions of all the projects, we can claim the cohesive alliance in multi-party settings can be established and supported when the social capabilities of robotic teammates consider a "shared sense of unity with the group". That sense of unity can be achieved in light of the dimensions of cohesion, by fostering attractions (social cohesion), identification with the group (collective cohesion), group feelings (emotional cohesion), and communication among teammates (structural cohesion).

8.2 Contributions

Across all the projects of this thesis, we identified three broad contributions:

C1 **Computational mechanisms to develop autonomous social behaviour for robotic teammates in multi-party settings** - We explored how can autonomous robots reason with a "shared sense of unity" at different levels: perception, cognition and expression (see Figure 8.1). For instance, Project V considered different perceptive skills on a robotic teammate in terms of gaze behaviours among team members, which translates into a detection or an awareness of different group patterns by the robot. Project IV introduced the social categorization step into the agent's appraisal process, which can be seen as a contribution at the cognitive level. This social categorization process allowed the robotic agent to autonomously express group emotions, rather than individual emotions, which is at the expressive level. Additionally, we contributed with heuristics to guide the robot's gaze behaviour, and the remaining projects explored the design of verbal utterances in multi-party settings.

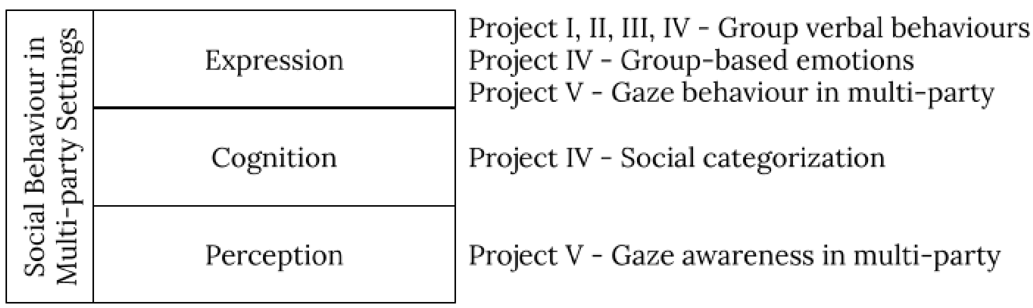


Figure 8.1: Reasoning with a "shared sense of unity".

C2 **Evaluation of people's behaviours and perceptions towards a mixed human-robot group** - We looked at how humans express a "shared sense of unity" in a mixed group with robots, both in terms of their objective behaviours and their subjective evaluations. We analysed objective behaviours such as the socioemotional support in Project I, and the prosocial decisions in the collective risk dilemmas of Projects II and III. In terms of subjective evaluations, we looked not only at how do people perceive the robot as a teammate, but also at how do they evaluate the team or the group as whole. In particular, some of the metrics we used in our user studies (e.g. group identification and group trust in Projects II, III, and IV) contribute with new insights to understand how group-level subjective measures can be assessed in HRI and how they relate with group processes [2]. This is particularly important as research on human-robot group interaction is still in its infancy and, therefore, the scientific community lacks measures and methodologies to assess group perceptions.

C3 **Understanding relevant group processes in mixed-groups of humans and robots -** Each of the five projects in this thesis sheds light on how cohesion can be fostered into a strong alliance between human(s) and robot(s) through different routes. Firstly, we showed social cohesion can emerge from the display of different character traits on the robotic teammate (Project I). Secondly, the portrayal of emotional cohesion (Project IV) or a higher degree of collective cohesion (Projects II and III) by a robotic teammate can also emphasise the cohesive alliance. Finally, we found evidence that structural cohesion can affect the perceived fluency in multi-party teamwork (Project V). We believe this constitutes a step forward into the understanding of how mixed-groups of humans and robots are established, and how they can be maintained.

These three contributions have a seamless relationship in the design, development and evaluation loop of robotic teammates for multi-party settings. In particular, there is mutual dependence between C1 and C2. In other words, the more we advance in the understanding of how humans behave and express their satisfaction with the human-robot team, the better we can develop computational mechanisms for autonomous robotic teammates. Such computational mechanisms might refer to perceptive and/or expressive social behaviours, as fully autonomous robotic teammates are required not only to perceive their human teammates, but also to express in a way humans conveniently understand. Finally, C3 can be seen as a meta-contribution that emerges from the interactive loop between C1 and C2. This means understanding the relevant group processes between human(s) and robot(s) unavoidably requires developing and evaluating such interactions.

Considering our research problem one last time, C1 answers to the part of "how can we endow", while C2 examines the part "to improve the cohesive alliance", and C3 clarifies "[which] social capabilities". Overall, we believe that reaching group intelligence in social robots can be strongly supported by the combination of these three contributions.

8.3 Applicability

The first and most straightforward application of this dissertation is **Entertainment Robotics**, which also includes **Smart Companions**. This is particularly relevant as there is a commercial trend for entertainment robots, for instance, Liku[1], Lovot[2], Photon[3], or even the former Jibo[4]. All the collaborative scenarios that were developed in the projects of this dissertation are games and, therefore, entertainment activities. We believe the design and development of entertainment robotics can take inspiration from the social behaviours of our robotic teammates. Furthermore, most entertainment activities are performed in group settings and are inherently social activities to engage and connect people together.

Another interesting application area for this thesis is **Collaborative Robotics**. We already have robots in industrial settings, but most of them barely interact or act near humans. For these robots to become teammates, to actively collaborate with humans, and to start entering any workspace, they require more advanced social capabilities. In particular, capabilities for perceiving, reasoning about, or acting around several humans, such as the ones we addressed in this dissertation.

Lastly, our thesis also holds relevant considerations for many real-world deployments under **Mixed-Motive Situations**. While we used games as research scenarios due to their controlled environment with well defined sets of actions, they can also produce very rich and complex social interactions. In particular, the presence of competitive teams (as in the *Sueca* card game), or the explicit mixed-motive goals (as in the *For The Record*) map decisions and dilemmas from the real-world. At the same time we might expect robots to become selfish and portray the interests of their owners (e.g. companies, government, individuals), we also hope our results encourage advances in **Prosocial Robotics** [111]. Our assumption is that robots capable of fostering the cohesive alliance with humans, in several daily-basis situations, might also promote people's pro-social behaviours.

8.4 Future work

We propose three research avenues to further advance the study of human-robot teamwork in multi-party settings:

- **Different group sizes and human-robot configurations:** Not only due to the existing in-group and our-group bias, but also to understand how group identification is established in mixed groups of humans and robots [134].

- **Different embodiments or anthropomorphic features:** An interesting avenue, especially after the work we presented in Chapter 5, is to clarify the impact of different anthropomorphic features or different levels of embodiement.

- **Autonomous perception of group patterns:** One of the assumptions in Group Dynamics theories is that groups evolve over time and that group members might have adaptive capabilities to cope with those changes. An initial step to make this vision possible for robotic group members is to advance their autonomous perceptive skills.

[1] http://likuwith.me
[2] https://lovot.life/
[3] https://photon.education
[4] https://jibo.com/

8.5 Summary of publications

The work presented in this dissertation resulted in 3 journal articles (JAAMAS, ISR and AuRo), 12 conference proceedings publications (RSS, AAMAS, HRI, RO-MAN, IROS, ASSETS, AIIDE), 7 workshop proceedings or refereed extended abstracts, and 2 book chapters. The detailed list is available in Appendix A.

We also provide a brief description of additional contributions performed during this thesis but falling somewhat outside the scope of this thesis, namely human-robot dyadic interactions, available in Appendix B.

Bibliography

[1] A. Abele. Functions of gaze in social interaction: Communication and monitoring. *Journal of Nonverbal Behavior*, 10(2):83–101, 1986.

[2] A. M. Abrams and A. M. Rosenthal-von der Pütten. I–c–e framework: Concepts for group dynamics research in human-robot interaction. *International Journal of Social Robotics*, pages 1–17, 2020.

[3] H. Admoni, B. Hayes, D. Feil-Seifer, D. Ullman, and B. Scassellati. Dancing with myself: The effect of majority group size on perceptions of majority and minority robot group members. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 35, 2013.

[4] 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*, pages 18–20, 2004.

[5] O. Amir, D. G. Rand, et al. Economic games on the internet: The effect of $1 stakes. *PloS one*, 7(2), 2012.

[6] M. Argyle and M. Cook. Gaze and mutual gaze. 1976.

[7] J. Avelino, F. Correia, J. Catarino, P. Ribeiro, P. Moreno, A. Bernardino, and A. Paiva. The power of a hand-shake in human-robot interactions. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1864–1869. IEEE, 2018.

[8] F. Babel, J. Kraus, P. Hock, H. Asenbauer, and M. Baumann. Investigating the validity of online robot evaluations: Comparison of findings from an one-sample online and laboratory study. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pages 116–120, 2021.

[9] W. A. Bainbridge, J. W. Hart, E. S. Kim, and B. Scassellati. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics*, 3(1):41–52, 2011.

[10] R. F. Bales. Interaction process analysis; a method for the study of small groups. 1950.

[11] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 59–66. IEEE, 2018.

[12] L. W. Barsalou, P. M. Niedenthal, A. K. Barbey, and J. A. Ruppert. Social embodiment. *Psychology of learning and motivation*, 43:43–92, 2003.

[13] C. Bartneck. Interacting with an embodied emotional character. In *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces*, pages 55–60. ACM, 2003.

[14] C. Bartneck, A. Duenser, E. Moltchanova, and K. Zawieska. Comparing the similarity of responses received from studies in amazon's mechanical turk to studies conducted online and with direct recruitment. *PloS one*, 10(4), 2015.

[15] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1(1):71–81, 2009.

[16] A. Bauer, D. Wollherr, and M. Buss. Human–robot collaboration: a survey. *International Journal of Humanoid Robotics*, 5(01):47–66, 2008.

[17] 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*, pages 391–398. IEEE Press, 2016.

[18] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, and F. Tanaka. Social robots for education: A review. *Science robotics*, 3(21):eaat5954, 2018.

[19] M. Bonani, R. Oliveira, F. Correia, A. Rodrigues, T. Guerreiro, and A. Paiva. What my eyes can't see, a robot can show me: Exploring the collaboration between blind people and robots. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 15–27, 2018.

[20] J. Brandstetter, P. Rácz, C. Beckner, E. B. Sandoval, J. Hay, and C. Bartneck. A peer pressure experiment: Recreation of the asch conformity experiment with robots. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1335–1340. IEEE, 2014.

[21] S. Brave, C. Nass, and K. Hutchinson. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International journal of human-computer studies*, 62(2):161–178, 2005.

[22] C. Breazeal. Social robots for health applications. In *2011 Annual international conference of the IEEE engineering in medicine and biology society*, pages 5368–5371. IEEE, 2011.

[23] M. Buro, J. R. Long, T. Furtak, and N. R. Sturtevant. Improving state evaluation, inference, and search in trick-based card games. In *IJCAI*, pages 1407–1413, 2009.

[24] S. E. Butner and M. Ghodoussi. Transforming a surgical robot for human telesurgery. *IEEE Transactions on Robotics and Automation*, 19(5):818–824, 2003.

[25] F. Capozzi, C. Beyan, A. Pierro, A. Koul, V. Murino, S. Livi, A. P. Bayliss, J. Ristic, and C. Becchio. Tracking the leader: Gaze behavior in group interactions. *Iscience*, 16:242–249, 2019.

[26] 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.

[27] M. L. Chang, R. A. Gutierrez, P. Khante, E. S. Short, and A. L. Thomaz. Effects of integrated intent recognition and communication on human-robot collaboration. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3381–3386. IEEE, 2018.

[28] 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*, pages 845–850. IEEE, 2012.

[29] W. F. Chaplin, J. B. Phillips, J. D. Brown, N. R. Clanton, and J. L. Stein. Handshaking, gender, personality, and first impressions. *Journal of personality and social psychology*, 79(1):110, 2000.

[30] H. I. Christensen. Intelligent home appliances. In *Robotics Research*, pages 319–327. Springer, 2003.

[31] M. Cicconet, M. Bretan, and G. Weinberg. Visual cues-based anticipation for percussionist-robot interaction. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '12, pages 117–118, New York, NY, USA, 2012. ACM.

[32] Correia. Group based emotions in teams of humans and robots - supplemental material, Jan 2018.

[33] F. Correia, P. Alves-Oliveira, N. Maia, T. Ribeiro, S. Petisca, F. S. Melo, and A. Paiva. Just follow the suit! trust in human-robot interactions during card game playing. In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pages 507–512. IEEE, 2016.

[34] F. Correia, P. Alves-Oliveira, T. Ribeiro, F. S. Melo, and A. Paiva. A social robot as a card game player. In *13th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 2017.

[35] F. Correia, S. Chandra, S. Mascarenhas, J. Charles-Nicolas, J. Gally, D. Lopes, F. P. Santos, F. C. Santos, F. S. Melo, and A. Paiva. Walk the talk! exploring (mis) alignment of words and deeds by robotic teammates in a public goods game. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–7. IEEE, 2019.

[36] F. Correia, C. Guerra, S. Mascarenhas, F. S. Melo, and A. Paiva. Exploring the impact of fault justification in human-robot trust. In *Proceedings of the 17th international conference on autonomous agents and multiagent systems*, pages 507–513, 2018.

[37] F. Correia, S. F. Mascarenhas, S. Gomes, P. Arriaga, I. Leite, R. Prada, F. S. Melo, and A. Paiva. Exploring prosociality in human-robot teams. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 143–151. IEEE, 2019.

[38] F. Correia, R. Oliveira, M. Bonani, A. Rodrigues, T. Guerreiro, and A. Paiva. Exploring collaborative interactions between robots and blind people. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 365–365. IEEE, 2019.

[39] F. Correia, S. Petisca, P. Alves-Oliveira, T. Ribeiro, F. S. Melo, and A. Paiva. I choose... you! membership preferences in human–robot teams. *Autonomous Robots*, pages 1–15, 2018.

[40] A. J. Cuddy, S. T. Fiske, and P. Glick. Warmth and competence as universal dimensions of social perception: The stereotype content model and the bias map. *Advances in experimental social psychology*, 40:61–149, 2008.

[41] F. Cummins. Gaze and blinking in dyadic conversation: A study in coordinated behaviour among individuals. *Language and Cognitive Processes*, 27(10):1525–1549, 2012.

[42] K. Dautenhahn, B. Ogden, and T. Quick. From embodied to socially embedded agents–implications for interaction-aware robots. *Cognitive Systems Research*, 3(3):397–428, 2002.

[43] E. J. De Visser, M. M. Peeters, M. F. Jung, S. Kohn, T. H. Shaw, R. Pak, and M. A. Neerincx. Towards a theory of longitudinal trust calibration in human–robot teams. *International journal of social robotics*, 12(2):459–478, 2020.

[44] C. Deligianis, C. J. Stanton, C. McGarty, and C. J. Stevens. The impact of intergroup bias on trust and approach behaviour towards a humanoid robot. *Journal of Human-Robot Interaction*, 6(3):4–20, 2017.

[45] E. Deng, B. Mutlu, M. J. Matarić, et al. Embodiment in socially interactive robots. *Foundations and Trends® in Robotics*, 7(4):251–356, 2019.

[46] M. Desai, M. Medvedev, M. Vázquez, S. McSheehy, S. Gadea-Omelchenko, C. Bruggeman, A. Steinfeld, and H. Yanco. Effects of changing reliability on trust of robot systems. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 73–80. ACM, 2012.

[47] J. Dias, S. Mascarenhas, and A. Paiva. Fatima modular: Towards an agent architecture with a generic appraisal framework. In *Emotion Modeling*, pages 44–56. Springer International Publishing, 2014.

[48] S. Dolcos, K. Sung, J. J. Argo, S. Flor-Henry, and F. Dolcos. The power of a handshake: neural correlates of evaluative judgments in observed social interactions. *Journal of Cognitive Neuroscience*, 24(12):2292–2305, 2012.

[49] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa. Effects of robot motion on human-robot collaboration. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 51–58. ACM, 2015.

[50] C. S. Dweck. Motivational processes affecting learning. *American psychologist*, 41(10):1040, 1986.

[51] S. D'Amario, H. Daffern, and F. Bailes. Perception of synchronization in singing ensembles. *Plos one*, 14(6):e0218162, 2019.

[52] J. A. Eison. *The development and validation of a scale to assess differing student orientations towards grades and learning.* PhD thesis, University of Tennessee, Knoxville., 1979.

[53] P. Ekman, W. V. Friesen, M. O'sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti, et al. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology*, 53(4):712, 1987.

[54] V. Evola and J. Skubisz. Coordinated collaboration and nonverbal social interactions: a formal and functional analysis of gaze, gestures, and other body movements in a contemporary dance improvisation performance. *Journal of nonverbal behavior*, 43(4):451–479, 2019.

[55] F. Eyssel and D. Kuchenbrandt. Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology*, 51(4):724–731, 2012.

[56] D. R. Forsyth. Group dynamics. 1990.

[57] M. Fraune, S. Kawakami, S. Šabanović, R. de Silva, and M. Okada. Three's company, or a crowd?: The effects of robot number and behavior on hri in japan and the usa. In *Proceedings of Robotics: Science and Systems*, Rome, Italy, July 2015.

[58] M. Fraune, S. Šabanović, and T. Kanda. Human group presence, group characteristics, and group norms affect human-robot interaction in naturalistic settings. *Frontiers in Robotics and AI*, 6:48, 2019.

[59] M. R. Fraune, S. Kawakami, S. Šabanović, P. R. S. De Silva, and M. Okada. Three's company, or a crowd?: The effects of robot number and behavior on hri in japan and the usa. In *Robotics: Science and Systems*, 2015.

[60] M. R. Fraune, Y. Nishiwaki, S. Šabanović, E. R. Smith, and M. Okada. Threatening flocks and mindful snowflakes: How group entitativity affects perceptions of robots. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 205–213. ACM, 2017.

[61] M. R. Fraune and S. Šabanović. Negative attitudes toward minimalistic robots with intragroup communication styles. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 1116–1121. IEEE, 2014.

[62] M. R. Fraune, S. Šabanović, 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.

[63] M. R. Fraune, S. Sherrin, S. Šabanović, and E. R. Smith. Rabble of robots effects: Number and type of robots modulates attitudes, emotions, and stereotypes. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 109–116. ACM, 2015.

[64] M. R. Fraune, S. Sherrin, S. Šabanović, and E. R. Smith. Is human-robot interaction more competitive between groups than between individuals? In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 104–113. IEEE, 2019.

[65] S. Gillet, R. Cumbal, A. Pereira, J. Lopes, O. Engwall, and I. Leite. Robot gaze can mediate participation imbalance in groups with different skill levels. In *2021 ACM/IEEE International Conference on Human-Robot Interaction (TO APPEAR)*. IEEE, 2021.

[66] M. L. Ginsberg. Gib: Imperfect information in a computationally challenging game. *Journal of Artificial Intelligence Research*, 14:303–358, 2001.

[67] A. Goldenberg, E. Halperin, M. van Zomeren, and J. J. Gross. The process model of group-based emotion: Integrating intergroup emotion and emotion regulation perspectives. *Personality and Social Psychology Review*, 20(2):118–141, 2016.

[68] J. J. Gross and R. A. Thompson. Emotion regulation: Conceptual foundations. In J. J. Gross, editor, *Handbook of Emotion Regulation*, pages 3–24. Guilford Press, 2007.

[69] J. Guiochet, M. Machin, and H. Waeselynck. Safety-critical advanced robots: A survey. *Robotics and Autonomous Systems*, 94:43–52, 2017.

[70] 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*, 53(5):517–527, 2011.

[71] M. Häring, D. Kuchenbrandt, and E. André. Would you like to play with me?: how robots' group membership and task features influence human-robot interaction. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 9–16. ACM, 2014.

[72] S. S. Hendrick. A generic measure of relationship satisfaction. *Journal of Marriage and the Family*, 50(1):93–98, 1988.

[73] G. Hoffman. Evaluating fluency in human–robot collaboration. *IEEE Transactions on Human-Machine Systems*, 49(3):209–218, 2019.

[74] G. Hoffman and C. Breazeal. Effects of anticipatory action on human-robot teamwork efficiency, fluency, and perception of team. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 1–8. ACM, 2007.

[75] L. Hoffmann and N. C. Krämer. Investigating the effects of physical and virtual embodiment in task-oriented and conversational contexts. *International Journal of Human-Computer Studies*, 71(7-8):763–774, 2013.

[76] S. HOLM. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

[77] M. J. Hornsey. Social identity theory and self-categorization theory: A historical review. *Social and Personality Psychology Compass*, 2(1):204–222, 2008.

[78] C.-M. Huang and B. Mutlu. Anticipatory robot control for efficient human-robot collaboration. In *The eleventh ACM/IEEE international conference on human robot interaction*, pages 83–90. IEEE Press, 2016.

[79] B. Jensen, N. Tomatis, L. Mayor, A. Drygajlo, and R. Siegwart. Robots meet humans-interaction in public spaces. *IEEE Transactions on Industrial Electronics*, 52(6):1530–1546, 2005.

[80] M. F. Jung. Coupling interactions and performance: Predicting team performance from thin slices of conflict. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 23(3):18, 2016.

[81] M. F. Jung, D. DiFranzo, B. Stoll, S. Shen, A. Lawrence, and H. Claure. Robot assisted tower construction-a resource distribution task to study human-robot collaboration and interaction with groups of people. *arXiv preprint arXiv:1812.09548*, 2018.

[82] M. F. Jung, J. J. Lee, N. DePalma, S. O. Adalgeirsson, P. J. Hinds, and C. Breazeal. Engaging robots: easing complex human-robot teamwork using backchanneling. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 1555–1566. ACM, 2013.

[83] M. F. Jung, N. Martelaro, and P. J. Hinds. Using robots to moderate team conflict: the case of repairing violations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 229–236. ACM, 2015.

[84] M. F. Jung, S. Šabanović, F. Eyssel, and M. Fraune. Robots in groups and teams. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 401–407. ACM, 2017.

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

[86] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita. An affective guide robot in a shopping mall. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 173–180. ACM, 2009.

[87] J. Kędzierski, R. Muszyński, C. Zoll, A. Oleksy, and M. Frontkiewicz. Emys—emotive head of a social robot. *International Journal of Social Robotics*, 5(2):237–249, 2013.

[88] A. Kendon. *Conducting interaction: Patterns of behavior in focused encounters*, volume 7. CUP Archive, 1990.

[89] T. Kessler and S. Hollbach. Group-based emotions as determinants of ingroup identification. *Journal of Experimental Social Psychology*, 41(6):677–685, 2005.

[90] C. D. Kidd and C. Breazeal. Effect of a robot on user perceptions. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, volume 4, pages 3559–3564. IEEE, 2004.

[91] S. Kiesler, A. Powers, S. R. Fussell, and C. Torrey. Anthropomorphic interactions with a robot and robot–like agent. *Social Cognition*, 26(2):169–181, 2008.

[92] 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*, pages 593–618, 2007.

[93] D. Kuchenbrandt, F. Eyssel, S. Bobinger, and M. Neufeld. When a robot's group membership matters. *International Journal of Social Robotics*, 5(3):409–417, 2013.

[94] 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*, 95(1):144, 2008.

[95] K. M. Lee, Y. Jung, J. Kim, and S. R. Kim. Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *International journal of human-computer studies*, 64(10):962–973, 2006.

[96] I. Leite, M. McCoy, M. Lohani, D. Ullman, N. Salomons, C. Stokes, S. Rivers, and B. Scassellati. Emotional storytelling in the classroom: Individual versus group interaction between children and robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 75–82. ACM, 2015.

[97] I. Leite, M. McCoy, D. Ullman, N. Salomons, and B. Scassellati. Comparing models of disengagement in individual and group interactions. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 99–105. IEEE, 2015.

[98] S. Mascarenhas, M. Guimarães, R. Prada, J. Dias, P. A. Santos, K. Star, B. Hirsh, E. Spice, and R. Kommeren. A virtual agent toolkit for serious games developers. In *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 1–7. IEEE, 2018.

[99] W. Mason and S. Suri. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods*, 44(1):1–23, 2012.

[100] D. McColl, W.-Y. G. Louie, and G. Nejat. Brian 2.1: A socially assistive robot for the elderly and cognitively impaired. *IEEE Robotics & Automation Magazine*, 20(1):74–83, 2013.

[101] J. E. McGrath. Methodology matters: Doing research in the behavioral and social sciences. In *Readings in Human–Computer Interaction*, pages 152–169. Elsevier, 1995.

[102] M. J. Mendelson and F. E. Aboud. Measuring friendship quality in late adolescents and young adults: Mcgill friendship questionnaires. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 31(2):130, 1999.

[103] A. Moors, P. C. Ellsworth, K. R. Scherer, and N. H. Frijda. Appraisal theories of emotion: State of the art and future development. *Emotion Review*, 5(2):119–124, 2013.

[104] B. Mullen and C. Copper. The relation between group cohesiveness and performance: An integration. *Psychological bulletin*, 115(2):210, 1994.

[105] A. Nawroj, M. Toneva, H. Admoni, and B. Scassellati. An exploration of social grouping in robots: Effects of behavioral mimicry, appearance, and eye gaze. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 36, 2014.

[106] V. Ng-Thow-Hing, J. Lim, J. Wormer, R. K. Sarvadevabhatla, C. Rocha, K. Fujimura, and Y. Sakagami. The memory game: Creating a human-robot interactive scenario for asimo. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pages 779–786. IEEE, 2008.

[107] R. Oliveira, P. Arriaga, P. Alves-Oliveira, F. Correia, S. Petisca, and A. Paiva. Friends or foes?: Socioemotional support and gaze behaviors in mixed groups of humans and robots. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 279–288. ACM, 2018.

[108] R. Oliveira, P. Arriaga, F. Correia, and A. Paiva. The stereotype content model applied to human-robot interactions in groups. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 123–132. IEEE, 2019.

[109] R. Oliveira, P. Arriaga, F. Correia, and A. Paiva. Looking beyond collaboration: Socioemotional positive, negative and task-oriented behaviors in human–robot group interactions. *International Journal of Social Robotics*, 12(2):505–518, 2020.

[110] A. Ortony, G. L. Clore, and A. Collins. *The cognitive structure of emotions*. Cambridge university press, 1990.

[111] A. Paiva, F. Correia, R. Oliveira, F. P. Santos, and P. Arriaga. Empathy and prosociality in social agents. *Handbook on Socially Interactive Agents*, 2021.

[112] A. Paiva, F. P. Santos, and F. C. Santos. Engineering pro-sociality with autonomous agents. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[113] E. Peer, L. Brandimarte, S. Samat, and A. Acquisti. Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70:153–163, 2017.

[114] A. Pereira, C. Oertel, L. Fermoselle, J. Mendelson, and J. Gustafson. Effects of different interaction contexts when evaluating gaze models in hri. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 131–139, 2020.

[115] C. O. Porter. Goal orientation: effects on backing up behavior, performance, efficacy, and commitment in teams. *Journal of Applied Psychology*, 90(4):811, 2005.

[116] T. Quick, K. Dautenhahn, C. L. Nehaniv, and G. Roberts. On bots and bacteria: Ontology independent embodiment. In *European Conference on Artificial Life*, pages 339–343. Springer, 1999.

[117] M. R. Ramos and H. Alves. Adaptação de uma escala multidimensional de identificação para português. *Psicologia*, 25(2):23–38, 2011.

[118] T. Ribeiro, E. Di Tullio, L. J. Corrigan, A. Jones, F. Papadopoulos, R. Aylett, G. Castellano, and A. Paiva. Developing interactive embodied characters using the thalamus framework: a collaborative approach. In *International Conference on Intelligent Virtual Agents*, pages 364–373. Springer, 2014.

[119] T. Ribeiro, A. Pereira, E. Di Tullio, and A. Paiva. The sera ecosystem: Socially expressive robotics architecture for autonomous human-robot interaction. In *2016 AAAI Spring Symposium Series*, 2016.

[120] T. Ribeiro, A. Pereira, E. D. Tullio, and A. Paiva. The sera ecosystem: Socially expressive robotics architecture for autonomous human-robot interaction. 2016.

[121] N. Salomons, S. S. Sebo, M. Qin, and B. Scassellati. A minority of one against a majority of robots: Robots cause normative and informational conformity. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(2):1–22, 2021.

[122] N. Salomons, M. van der Linden, S. Strohkorb Sebo, and B. Scassellati. Humans conform to robots: Disambiguating trust, truth, and conformity. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 187–195. ACM, 2018.

[123] F. P. Santos, S. Mascarenhas, F. C. Santos, F. Correia, S. Gomes, and A. Paiva. Picky losers and carefree winners prevail in collective risk dilemmas with partner selection. *Autonomous Agents and Multi-Agent Systems*, 34(2):1–29, 2020.

[124] F. P. Santos, S. F. Mascarenhas, F. C. Santos, F. Correia, S. Gomes, and A. Paiva. Outcome-based Partner Selection in Collective Risk Dilemmas. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1556–1564. International Foundation for Autonomous Agents and Multiagent Systems, 2019.

[125] J. Schroeder, J. Risen, F. Gino, and M. I. Norton. Handshaking promotes cooperative dealmaking. *Available at SSRN 2443551*, 2014.

[126] S. Sebo, L. L. Dong, N. Chang, M. Lewkowicz, M. Schutzman, and B. Scassellati. The influence of robot verbal support on human team members: Encouraging outgroup contributions and suppressing ingroup supportive behavior. *Frontiers in Psychology*, 11:3584, 2020.

[127] S. Sebo, B. Stoll, B. Scassellati, and M. F. Jung. Robots in groups and teams: a literature review. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–36, 2020.

[128] E. M. Segura, M. Kriegel, R. Aylett, A. Deshmukh, and H. Cramer. How do you like me in this: User embodiment preferences for companion agents. In *International Conference on Intelligent Virtual Agents*, pages 112–125. Springer, 2012.

[129] C. E. Sembroski, M. R. Fraune, and S. Šabanović. He said, she said, it said: Effects of robot group membership and human authority on people's willingness to follow their instructions. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 56–61. IEEE, 2017.

[130] S. Shahid, E. Krahmer, and M. Swerts. Child–robot interaction across cultures: How does playing a game with a social robot compare to playing a game alone or with a friend? *Computers in Human Behavior*, 40:86–100, 2014.

[131] M. Shayganfar, C. Rich, and C. Sidner. Appraisal algorithms for relevance and controllability in human-robot collaboration. In *2019 IEEE International Conference on Humanized Computing and Communication (HCC)*, pages 31–37. IEEE, 2019.

[132] S. Shen, P. Slovak, and M. F. Jung. Stop. i see a conflict happening.: A robot mediator for young children's interpersonal conflict resolution. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 69–77. ACM, 2018.

[133] E. R. Smith. Social identity and social emotions: Toward new conceptualizations of prejudice. 1993.

[134] E. R. Smith, S. Šabanović, and M. R. Fraune. Human-robot interaction through the lens of social psychological theories of intergroup behavior. 2021.

[135] R. D. Smither and J. M. Houston. The nature of competitiveness: The development and validation of the competitiveness index. *Educational and Psychological Measurement*, 52(2):407–418, 1992.

[136] A. Somech, H. S. Desivilya, and H. Lidogoster. Team conflict management and team effectiveness: The effects of task interdependence and team identification. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 30(3):359–378, 2009.

[137] G. Stahl. *Group cognition: Computer support for building collaborative knowledge (acting with technology)*. The MIT Press, 2006.

[138] A. Steain, C. J. Stanton, and C. J. Stevens. The black sheep effect: The case of the deviant ingroup robot. *PloS one*, 14(10), 2019.

[139] G. L. Stewart, S. L. Dustin, M. R. Barrick, and T. C. Darnold. Exploring the handshake in employment interviews. *Journal of Applied Psychology*, 93(5):1139, 2008.

[140] 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*, pages 14–22. ACM, 2018.

[141] N. R. Sturtevant. An analysis of uct in multi-player games. In *International Conference on Computers and Games*, pages 37–49. Springer, 2008.

[142] H. Tennent, S. Shen, and M. Jung. Micbot: A peripheral robotic object to shape conversational dynamics and team performance. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 133–142. IEEE, 2019.

[143] K. A. Thomas and S. Clifford. Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior*, 77:184–197, 2017.

[144] M. Tomasello et al. Joint attention as social cognition. *Joint attention: Its origins and role in development*, 103130:103–130, 1995.

[145] M. L. Traeger, S. S. Sebo, M. Jung, B. Scassellati, and N. A. Christakis. Vulnerable robots positively shape human conversational dynamics in a human–robot team. *Proceedings of the National Academy of Sciences*, 117(12):6370–6375, 2020.

[146] S. Tulli, F. Correia, S. Mascarenhas, S. Gomes, F. S. Melo, and A. Paiva. Effects of agents' transparency on teamwork. In *International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems*, pages 22–37. Springer, 2019.

[147] J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher, and M. S. Wetherell. *Rediscovering the social group: A self-categorization theory.* Basil Blackwell, 1987.

[148] M. Vázquez, E. J. Carter, B. McDorman, J. Forlizzi, A. Steinfeld, and S. E. Hudson. Towards robot autonomy in group conversations: Understanding the effects of body orientation and gaze. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, pages 42–52. IEEE, 2017.

[149] N. Wang, D. V. Pynadath, and S. G. Hill. Trust calibration within a human-robot team: Comparing automatically generated explanations. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, pages 109–116. IEEE Press, 2016.

[150] A. Washburn, R. W. Kallen, M. Lamb, N. Stepp, K. Shockley, and M. J. Richardson. Feedback delays can enhance anticipatory synchronization in human-machine interaction. *PloS one*, 14(8):e0221275, 2019.

[151] T. Ziemke. What's that thing called embodiment? In *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, 2003.

Appendix A

List of publications and communications

Journal publications:

- Santos, F. P., Mascarenhas, S., Santos, F. C., Correia, F., Gomes, S., & Paiva, A. (2020). Picky losers and carefree winners prevail in collective risk dilemmas with partner selection. Autonomous Agents and Multi-Agent Systems, 34(2), 1-29.

- Oliveira, R., Arriaga, P., Correia, F., & Paiva, A. (2019). Looking Beyond Collaboration: Socioemotional Positive, Negative and Task-Oriented Behaviors in Human–Robot Group Interactions. International Journal of Social Robotics, 1-14.

- Correia, F., Petisca, S., Alves-Oliveira, P., Ribeiro, T., Melo, F. S. & Paiva, A. (2018, June). I Choose... YOU! Membership Preferences in Human-Robot Teams. In Autonomous Robots, Special Issue on Robotics: Science and Systems, 2018.

Conference proceedings publications:

- Correia, F., Gomes, S., Mascarenhas, S., Melo, F. S., & Paiva, A. (2020). The Dark Side of Embodiment-Teaming Up With Robots VS Disembodied Agents. Proceedings of Robotics: Science and Systems. Corvalis, Oregon, USA. [RSS, Rank: A* (CORE2018)]

- Correia, F., Chandra, S., Mascarenhas, S., Charles-Nicolas, J., Gally, J., Lopes, D., Santos, F.P., Santos, F.C., Melo, F.S. & Paiva, A. (2019, October). Walk the Talk! Exploring (Mis) Alignment of Words and Deeds by Robotic Teammates in a Public Goods Game. In 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) (pp. 1-7). IEEE. (Award Winner)

- Santos, F., Mascarenhas, S., Santos, F. C., Correia, F., Gomes, S., & Paiva, (2019, May). Outcome-based Partner Selection in Collective Risk Dilemmas. In Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (pp. 1556-1564). International Foundation for Autonomous Agents and Multiagent Systems. [AAMAS, Rank: A* (CORE2018)]

- Correia, F., Mascarenhas, S., Gomes, S., Arriaga, P., Leite, I., Melo, F. S., Prada, R. & Paiva, A. (2019, March). Exploring prosociality in human-robot teams. In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 143-151). IEEE. [HRI, Rank: A2 (Qualis)]

- Oliveira, R., Arriaga, P., Correia, F. & Paiva, A. (2019, March). The Stereotype Content Model Applied to Human-Robot Interactions in Groups. In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 123-132). IEEE. [HRI, Rank: A2 (Qualis)]

- Avelino, J., Correia, F., Catarino, J., Ribeiro, P., Moreno, P., Bernardino, A. & Paiva, A. (2018, October). The Power of a Hand-shake in Human-Robot Interactions. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 1864-1869). IEEE. [IROS, Rank: A1 (Qualis)]

- Bonani, M., Oliveira, R., Correia, F., Rodrigues, A., Guerreiro, T. & Paiva, A. (2018, October). What My Eyes Can't See, A Robot Can Show Me: Exploring the Collaboration Between Blind People and Robots. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (pp. 15-27). ACM. [ASSETS, Rank: A2 (Qualis)]

- Correia, F., Guerra, C., Mascarenhas, S., Melo, F. S. & Paiva, A. (2018, July). Exploring the impact of fault justification in human-robot trust. In Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (pp. 507-513). International Foundation for Autonomous Agents and Multiagent Systems. [AAMAS, Rank: A* (CORE2018)]

- Correia, F., Mascarenhas, S., Melo, F. S., Prada, R. & Paiva, A. (2018, February). Group-based emotions in teams of humans and robots. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (pp. 261-269). ACM. [HRI, Rank: A2 (Qualis)]

- Oliveira, R., Arriaga, P., Alves-Oliveira, P., Paiva, A., Petisca, S. & Correia, F. (2018, February). Friends or foes?: Socioemotional support and gaze behaviors in mixed groups of humans and robots. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (pp. 279-288). ACM. [HRI, Rank: A2 (Qualis)]

- Correia, F., Alves-Oliveira, P., Ribeiro, T., Melo, F. S. & Paiva, A. (2017, October). A Social Robot as a Card Game Player. In 13th Artificial Intelligence and Interactive Digital Entertainment Conference. [AIIDE, Rank: B2 (Qualis)]

- Correia, F., Petisca, S., Alves-Oliveira, P., Ribeiro, T., Melo, F. S., & Paiva, A. (2017, July). Groups of humans and robots: Understanding membership preferences and team formation. In Robotics: Science and Systems. [RSS, Rank: A* (CORE2017)]

Workshop proceedings and refereed extended abstracts:

- Tulli, S., Correia, F., Mascarenhas, S., Gomes, S., Melo, F. S., & Paiva, A. (2019, May). Effects of Agents' Transparency on Teamwork. In International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems (pp. 22-37). Springer, Cham.

- Correia, F., Mascarenhas, S., Gomes, S., Tulli, S., Santos, F. P., Santos, F. C., Prada, R., Melo, F. S., & Paiva, A. (2019, May). For The Record - A Public Goods Game For Exploring Human-Robot Collaboration. In Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (pp. 2351-2353). International Foundation for Autonomous Agents and Multiagent Systems - Demos Track, 2019 [AAMAS, Rank: A* (CORE2018)]

- Correia, F., Melo, F. S., & Paiva, A. (2019, March). Group Intelligence on Social Robots. In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 703-705). IEEE.[HRI, Rank: A2 (Qualis)]

- Correia, F., Oliveira, R., Bonani, M., Rodrigues, A., Guerreiro, T., & Paiva, A. (2019, March). Exploring Collaborative Interactions Between Robots and Blind People. In 2019 14th ACM/IEEE

International Conference on Human-Robot Interaction (HRI) (pp. 365-365). IEEE. [HRI, Rank: A2 (Qualis)]

- Oliveira, Raquel, Arriaga, P., Correia, F. & Paiva, A. (2018, March). Making Robot's Attitudes Predictable: A Stereotype Content Model for Human-Robot Interaction in Groups. In Human-Robot Interaction Workshop on Workshop Explainable Robotic Systems. [HRI, Rank: A2 (Qualis)]

- Correia, F., Arriaga, P., Petisca, S., Alves-Oliveira, P., Oliveira. R., Mascarenhas, S., Leite, I., Melo, F. S. & Paiva, A. (2017, August). Groups of Humans and Robots: the AMIGOS Project. In Robot and Human Interactive Communication Workshop on Groups in Human-Robot Interaction. [RO-MAN, Rank: B1 (Qualis)]

- Correia, F., Alves-Oliveira, P., Petisca, S., & Paiva, A. (2017, July). Social and Entertainment Robots for Older Adults. In Proc. Robotics: Science and Systems Workshop on Human-Centered Robotics: Interaction, Physiological Integration and Autonomy. [RSS, Rank: A* (CORE2017)]

Book chapters:

- Paiva, A., Correia, F., Oliveira, R., Santos, F. P. & Arriaga, P. (2021, TO APPEAR). Empathy and Prosociality in Social Agents. In B. Lugrin, C. Pelachaud, D. Traum (Eds), ACM Handbook of Socially Interactive Agents, Chapter 11. ACM, 2021.

- Paiva, A., Mascarenhas, S., Petisca, S., Correia, F. & Alves-Oliveira, P. (2018, April). Towards more humane machines: Creating Emotional Social Robots. In S. G. Silva (Eds), New Interdisciplinary Landscapes in Morality and Emotion, Chapter 9. Routledge, 2018.

Communications (International Conferences):

- The Dark Side of Embodiment-Teaming Up With Robots VS Disembodied Agents (video presentation), **RSS'20**, Robotics: Science and Systems, Corvalis, USA (held virtually)

- Walk the Talk! Exploring (Mis) Alignment of Words and Deeds by Robotic Teammates in a Public Goods Game (oral presentation), **ROMAN'19**, 28th IEEE International Conference on Robot and Human Interactive Communication, New Delhi, India

- Exploring prosociality in human-robot teams (oral presentation), **HRI'19**, 14th ACM/IEEE International Conference on Human-Robot Interaction, Daegu, South Korea

- Exploring the impact of fault justification in human-robot trust (oral presentation), **AAMAS'18**, 17th International Conference on Autonomous Agents and MultiAgent Systems, Stockholm, Sweden

- Group-based emotions in teams of humans and robots (oral presentation), **HRI'18**, 13th ACM/IEEE International Conference on Human-Robot Interaction, Chicago, USA

- A Social Robot as a Card Game Player (oral presentation), **AIIDE'17**, 13th Artificial Intelligence and Interactive Digital Entertainment Conference, Snowbird, USA
- Groups of humans and robots: Understanding membership preferences and team formation (oral presentation), **RSS'17**, Robotics: Science and Systems, Cambridge, USA

Appendix B

Other Work

In this appendix, we provide a brief description of additional contributions performed during this thesis but falling somewhat outside the scope of this thesis, namely human-robot dyadic interactions.

B.1 Fault Justification in Human-Robot Trust

Robots, like any other machines, are susceptible to fail or present some degree of error. We are all familiar with a robot that suddenly halts, starts repeating itself, says something out of context, and many other situations. Depending on the nature of the task and the purpose of the robot, the impact of failures can range from amusing to highly dangerous. However, even in low-risk situations, such as a conversational, entertainment or companionship, failures may have a significant adverse effect on trust, user engagement and even willingness to interact with the robot in the future. From a performance standpoint, if robots are able to understand and recover from their failures automatically, they will be more efficient and reliable. But as robots become more social and interact with humans in various forms, the expectations on how robots handle such failures may go beyond their capacity for autonomous recovery. For instance, in collaborative tasks where robots are interacting with humans, the robot's behaviour should also address the social implications of their failures upon others. If we expect others to justify and explain their failures to us, it is likely that we will expect social robots to do so as well.

This work sets out to understand how a robot can recover from a failure in order to mitigate its possible negative social effects. In particular, if a robot justifies the failure, will it mitigate the effects of it? By addressing these questions, this work contributes to the design of social agents that can autonomously overcome error situations in a more appropriate manner. We conducted a user study where an autonomous social robot collaborates with participants in a shared task, a puzzle game. At a certain point, the robot has a technical failure during the task and, depending on the experimental condition, adopts a different social recovery strategy. We assessed the impact of the failure and the recovery strategies on the reported trust towards the robot. The motivation behind analysing trust is the fact that it is one of the most critical and essential elements for an effective collaboration between humans and agents [70]. Moreover, according to Hancock et al., trust is strongly influenced by the agent's performance and other attributes, such as transparency.

Overall, the obtained results indicate that the recovery strategy of justifying the failure was able to mitigate the negative impact of the failure, but only when the consequence of the failure was less severe (when the failure did not compromise the task). That is, in scenarios where the failure is not too severe, a strategy of justifying a failure to the users can mitigate the overall trust in the robot. The implications of these findings are particularly relevant to the current growing interest in collaboration between humans and agents, or in tasks where agents act as peers or constitute a team with humans.

The full description of this work can be found in [36].

B.2 The Power of a Hand-shake in HRI

Handshaking is the default greeting ritual between humans in western civilizations, and frequently the first form of interaction between people. It is a powerful non-verbal behaviour that can influence how individuals perceive social interaction partners and even their interest in future interactions [48]. In fact, studies have shown that people make personality judgments based on handshakes [29], and that the way one performs a handshake has a strong impact on the perceived employment suitability in recruitment tasks [139]. Other studies have also claimed that handshakes influence negotiation outcomes and promote cooperative behaviour [125].

In our view, social robots should be able to perform and understand human norms and social rituals if they are to be acknowledged as influential parts of society. Applications of robot assistants include those of guides, negotiators, and coaches, roles where trust is critical. Furthermore, current applications for social robots go towards human-robot collaboration as it allows the exploitation of the complementary skills that humans and robots have through an optimal division of tasks. Interestingly, non-verbal cues seem to have an important role in human-robot teamwork [90], not only do people expect these social cues to convey the mental model of the robot, but also the robot should understand the same cues in humans.

As a result, we have conducted a user study that attempts to measure the impact of handshakes by the Vizzy robot in a task-based scenario. Moreover, we analysed the helping pro-social behaviour, which is not mandatory for the success of the person on task. We believe this is the first attempt at studying the effects of a robot handshake on the subjective evaluation of the robot, and on humans' pro-social behaviours and predispositions.

The results revealed that participants in the Handshake condition evaluated the robot as more warm, animated and likeable and were more willing to help it in the future compared to participants in the No Handshake condition. Overall, this work contributes to the HRI community by reporting some of the effects a handshake might have and emphasizes the urge to explore further questions related with this powerful non-verbal behaviour.

The full description of this work can be found in [7].

B.3 Collaboration Between Blind People and Robots

Blind people face challenges in their daily lives in tasks that are taken as granted if you are sighted. Examples are varied and include finding objects, correctly placing items, and identifying different colours, text, or other visual patterns. These difficulties render several common activities hard to accomplish without the help of others. The inclusion of visually impaired people in a society that fights for equal rights is severely hindered by this dependence and it manifests itself in an household setting but also in school and in the work environment.

In this work, we first explore how blind people perceive robots nowadays and what are their expectations and fears regarding the increasing dependence on these devices. To do so, we performed 4 focus groups with a total of 20 visually impaired people. Results showed that the participants take a practical stance to the inclusion of robots in their day to day albeit presenting common place concerns regarding safety, over reliance on robots, and mistrust in their abilities.

In a second study, we invited 12 visually impaired people to perform an assembly task collaboratively with a robot, in two conditions that varied in the degree of collaboration: in one condition, the robot would only issue voice instructions to the participants and react to requests for past instructions; in the second, the robot would physically indicate pieces and their target position and orientation, provide feedback on the current step of the assembly, and correct the participant, when needed. Results showed that the engagement in a physical form of collaboration by the robot enabled and improved the success on the task at hand when compared to the voice-only condition. Furthermore, this physical form of interaction between the two parties, human and robot, was welcomed by the participants that reported the robot to be more competent, useful and warm when it interacted with them through physical collaboration than through a voice-only collaboration.

Overall, we contribute to the area of human-robot interaction and accessibility by bringing new insights on the collaboration between people and robots. Particularly, 1) we provide knowledge on the acceptance, perceived usefulness, foreseen scenarios, and concerns regarding the inclusion of robots in blind people daily lives; and 2) we inspect the collaboration between a robot (Baxter) and blind participants in an assembly task, paying particular attention to the effect of two different forms of collaboration in participant's success on the task and on the engagement with the robot. This research yields valuable insights to researchers exploring the role of assistive robots in supporting blind people achieving a higher degree of independence and autonomy.

The full description of this work can be found in [19; 38].