Groups of humans and robots: Understanding membership preferences and team formation

Filipa Correia*, Sofia Petisca†, Patrícia Alves-Oliveira†, Tiago Ribeiro*, Francisco S. Melo* and Ana Paiva*

*INESC-ID & Instituto Superior Técnico, Universidade de Lisboa
†INESC-ID & Instituto Universitário de Lisboa (ISCTE-IUL), CIS-IUL

Abstract—Although groups of robots are expected to interact with groups of humans in the near future, research related to teams of humans and robots still appears scarce. This paper contributes to the study of human-robot teams by investigating how humans choose robots to partner with in a multi-party game context. The novelty of our work concerns the successful design and development of two social robots that are able to *autonomously* interact with a group of two humans in the execution of a social and entertaining task. The development of these two characters was motivated by psychological research on learning goal theory, according to which we interpret and approach a given task differently depending on our learning goal (oriented more towards either relationship building or competition). Thus, we developed two robotic characters implemented in two robots: *Emys* (competitive robot) and *Glin* (relationship-driven robot). In our study, a group of four (two humans and two autonomous robots) engaged in a social and entertaining card game. Our study yields several important conclusions regarding groups of humans and robots. (1) When a partner is chosen without previous partnering experience, people tend to prefer robots with relationship-driven characteristics as their partners compared with competitive robots. (2) After some partnering experience has been gained, the choice becomes less clear and additional driving factors emerge: (2a) participants with higher levels of competitiveness (personal characteristics) tend to prefer Emys, whereas those with lower levels prefer Glin, and (2b) the choice of which robot to partner with also depends on team performance, with the winning team being the preferred choice.

I. Introduction

Our society is structured around groups, from small groups such as romantic pairs to larger groups such as committees, neighbourhoods, businesses and sports clubs. Until recently, these work groups have been composed solely of humans. However, with current technological developments, robots are beginning to execute tasks that have traditionally been performed by humans. Consequently, at some point, it is inevitable that humans will need to engage in groups of which robots will also be a part. However, for that to be possible, robots must be endowed with social competencies and advanced interaction capabilities. Significant work is currently being performed in which robots act as team members [4, 10, 12], extending the boundaries of embodied interactions and making them more meaningful and human-like. In the future, we may use robot companions to foster social relationships through deeper long-term interactions. According to Coradeschi and Saffiotti (2006), a paradigm shift is occurring in which autonomous and solitary robots will be replaced with pervasive robotic systems working in symbiosis (and in teams) with people and their environments [5]. In this paper, we study multi-robot and multi-person interactions, investigating people's preferences to robots with different goals.

In this context, several questions arise: What types of teams will emerge that will consist of humans and robots together? Will people actually be satisfied with partnering with a robot? If so, what type of robot will they prefer? The choice of whom to partner with on a team depends on many different factors, including the characteristics of the individuals and the tasks to be executed. Therefore, differences in the social competencies or personalities of the partner robots may influence the degree of willingness to have a robot as part of a human-robot team.

According to Hinds et al. (2000), when we, as humans, select a team member to work with, we tend to privilege homogeneous groups with high indicators of competence and with greater similarity and familiarity [17]. These preferences may be related to our attempts to make choices that will maximise our expectation of success. Whereas previous research on Human-Robot Interaction (HRI) has demonstrated that users prefer robots whose personalities match with theirs in terms of introversion/extroversion in a therapeutic task [31], this paper extends those results to the influence of user personality in both cooperative and competitive interactions.

At an individual level, people's goal orientations have a major effect on how they approach and respond to a task. Dweck (1986) extended the notion of goal orientation introduced by Eison (1979) [9] 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) [8]. For instance, teams consisting of individuals with a "learning-goal" orientation are reported to show high levels of mutual support behaviours and high qualities of interaction, team efficacy and commitment. By contrast, teams consisting of individuals with a "performance-goal" orientation are negatively correlated with team efficacy and commitment [25].

In this paper, we aim to investigate multi-robot and multi-person interactions in the context of a game to study team formation and preferences. Thus, we will describe an investigation of team formation with robots, considering robots with different goal orientations. To achieve this objective, we relied on these two notions of goal orientation [20] (learning- and performance-goal orientations) to develop two different robotic characters: (1) a more *relationship-driven* personality (named Glin), mapping to the learning-goal orientation, and

(2) a more *competitive* one (named Emys), mapping to the performance-goal orientation. These two robots autonomously played a card game with two humans. Thus, the autonomous robots were interacting in a group of four and partnered with either a human or the other robot[1].

Two studies are reported in this paper. The first study **validated the characters' goal orientations** to be implemented in the robots for the second study. The second study was designed to investigate **which robotic character was preferred as a partner** in playing the card game, depending on the two goal orientations. To this end, teams of robots and humans were formed with the goal of playing a game in a social setting. Both robots played equally well, but one robot showed very competitive interactions, whereas the other was more interested in the quality of the interactions. We expected that overall, participants would tend to choose the more relational robot as their preferred partner, but we also hypothesised that the level of competitiveness would play a role in the choice, with higher scores being associated with choosing the competitive robot. The results showed that in general, the participants liked having the robots as partners, and upon first impressions, they preferred the more relational robot to the more competitive one. However, the results also showed that after repeated interactions and partnering with both robots, this choice became less clear and other driving factors emerged in the preference decision. For example, more competitively oriented people preferred the more competitively oriented robot. Furthermore, the findings also showed that team performance affects partner choice.

In general, this paper contributes to the study of teams of humans and autonomous robots; specifically, it addresses membership in mixed teams in the context of playing a game. Furthermore, this paper presents evidence that preferences regarding team formation with an autonomous robot depend not only on the robot's goal orientation (competitive vs. relationship-driven) but also on the characteristics of the people involved (such as personality). This finding has implications for the field of HRI, as it introduces factors that can impact preferences regarding the choice of a robotic partner and how these preferences vary over time.

II. RELATED WORK

Over the years, the field of HRI has been evolving from being mostly focused on "one-to-one" interactions to considering more complex scenarios in which (1) *individual users* interact with *multiple robots* or (2) *multiple users* interact with *individual robots*. However, a vision for the future concerns not only individual robots operating in a variety of human environments [13] but also multiple robots interacting with multiple people and with each other [27]. Indeed, Groom and Nass (2007) [15] have identified this trend by defining several benchmarks that enable the emergence of optimal social teams consisting of humans and robots and by emphasising how robots can complement and improve current human-human

teams. Furthermore, well-established and grounded social psychological theories postulate that people's behaviour changes when they are exposed to a certain group or individual (e.g., [2]), and therefore, the study of groups of humans and robots is undeniably a crucial area of HRI.

For social robots to be able to interact with multiple users, they need to be endowed with social competencies. In general, research findings suggest that humans often treat instruments of technology as social actors [26], applying social rules and expectations and exhibiting overlearned social behaviours such as politeness towards machines [24]. Several studies have been performed to analyse group effects related to individual robots, such as group membership and social categorisation (*e.g.*, [21]). In addition, studies on HRI have confirmed that social categorisation processes associated with groups also generalise to social robots. By manipulating group membership, Eyssel et al. (2012) showed that people anthropomorphise and prefer an in-group robot to a greater extent than an out-group robot [10]. Chang et al. (2012) studied the type of behaviour that humans chose to adopt (competitive or cooperative) depending on group size (a group of humans or an individual human player). The results showed that participants behave more competitively towards a robot when they are in a group than when they are interacting as individual players [4]. Additionally, a cross-cultural field study investigated participants' behaviour depending on the number of robots (single robot or group of robots) with which they were interacting and the type of behaviour (social or functional) the robot(s) exhibited. The results showed that people regarded single social robots more positively than social robots in a group. By contrast, people felt more positively towards a group of functional robots than a single functional robot. This research already suggests that the specific characteristics of robots (in this case, functional versus social behaviour) influence their group effects [12]. Moreover, personality appears to be an important variable influencing how people perceive and choose robots. Findings suggest that people tend to prefer robots whose personalities match their own in a therapeutic context [31], with similar findings related to pet-like robots [22]. Additionally, it has been found that people's personality traits are predictive of comfortable proximity distances when interacting with social robots [32]. More recently, [11] explored people's responses to groups of robots and compared the responses to different types of groups by varying the "diversity" of the groups. A Wizard-of-Oz (WoZ) approach was used to control the robots, and the human participants were directed to solve a task in the presence of 3 robots (under two conditions: high similarity and diversity). The results of this experiment showed that people perceive multiple robots that act and look the same as more threatening than a diverse group of robots.

Regarding human-robot teams, other concerns arise when exploring how these partnerships can evolve in a symbiotic manner and contribute to improved human-robot collaboration. For example, the efficiency of work performed with a robot increases when the robot shares non-verbal cues with its teammate [3]. Furthermore, Shah and collaborators (2011) have

shown that team performance increases when the behaviour of the robot is based on human-human cooperation [29] or when it adapts to the user [18]. Another study involving groups of humans and robots in a team examined the role of backchanneling competencies in a robot [19]. The results of this study support the assumption that even in robots, backchanneling is important for team performance.

Despite the significant work that has been done in this new area of research, we believe that the work presented here makes a novel contribution to this new era, moving beyond "one-to-one" and "one-to-many" interactions to scenarios in which several robots and several humans are interacting with each other. Moreover, we also contribute to the field by demonstrating how some preferences regarding robotic partners are influenced by social characteristics of both the humans and the robots. Finally, we contribute to the field of HRI through the development of two autonomous robots that can interact between themselves and with two humans.

III. CREATING TWO CHARACTERS FOR TWO ROBOTIC GAME PLAYERS

To explore the topic of human-robot teams and the role of goal orientation in the formation of those teams, it was necessary to define a suitable scenario. We chose the card game SUECA, which is a four-player game played between two teams. Partners on the same team sit across from each other and must play symbiotically to succeed in winning the game. Especially because this is a hidden-information card game in which players do not know each other's cards, the relationship between each player and his or her partner constitutes a relevant part of the game. Traditionally, two partners who frequently play with each other do not like to switch to different teams, as they often develop communication signals or other complicit mechanisms and each partner understands how the other plays, thus making them a better team.

The fact that SUECA is a hidden-information game introduces an element of chance or luck into this game. This property can be seen as an inaccessible characteristic of the environment [28], which not only complicates the task for any autonomous agent but also enriches the scenario in comparison to scripted interactions or even WoZ-built scenarios. In our scenario, we consider a mixed environment (see Figure 1) in which humans play with physical cards and robots play with virtual cards. The human players hold their cards during the game and are responsible for shuffling them and distributing them to each player. As such, the robots must autonomously play the game (with virtual cards), which is in reality unpredictable for both the humans and the robots; consequently, this is a very naturalistic scenario. To assist with game play, the physical cards have fiducial markers that can be detected by a multi-touch surface, thereby perfectly blending the natural card game experience for the humans with the digital version required by the robots.

We faced several challenges in creating our robotic team players, in particular, (1) how to build two distinct robots that reflect different goal orientations through their social interactions 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.

We created two different characters, Emys and Glin. Emys was given a more 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 more learning-driven goal orientation; consequently, although it strives for its team to win the game, it is also focused on fostering team spirit and providing a good game experience.

To this end, the agent architecture relies on a set of utterances, voices, and playing capabilities, which are activated according to the game situation, the triggered emotional appraisal, and the different possibilities for action. These characteristics show similarities but also differences, allowing the two robots to be perceived as distinct characters.

A. Social Behaviours

Initial behaviours for the robots were created, inspired by the way in which people actually play this game (a series of recordings were made, annotated, and converted into utterances). Furthermore, we also needed to understand the interaction patterns, such as which game states trigger certain behaviours and certain emotions, in order to produce more natural and human-like utterances for the robots [6]. After building a neutral baseline of utterances, we then needed to adapt them according to the different characters we wanted to build. Thus, each robot was given a set of utterances pertaining to all relevant game states or events (treated as categories) and their corresponding triggering behaviours, including both verbal and non-verbal behaviours. Each list of utterances was used to author each character. The number of utterances per category and subcategory was the same for both Emys and Glin to ensure that neither would be more repetitive than the other. Moreover, both autonomous agents were designed to produce behaviours with similar frequencies to ensure that neither would exceed the other in its interaction rate. Importantly, Glin and Emys also have identical physical bodies (EMotive headY System (EMYS)).

For Emys, the utterances were built based on a competitive perspective, always in pursuit of the best score. 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 for the loss and will consequently blame the others, either the partner or the opponents. By contrast, Glin was built with different parameters, leading to a more relational perspective and the verbalisation of much 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 as a whole to win and therefore assuming more of a supportive role.

Regarding non-verbal behaviours, each character produces expressive facial animations and postures according to its activated emotional state. The emotional agent architecture used [7] activates an emotional state in response each game

event, which will trigger different behaviours in the robot. To this end, the impact of each game event on the robot's goal of winning the game is calculated, and the emotional state is updated accordingly. As different emotions can be simultaneously activated by a single event (for example, sadness and hope), the strongest emotion among the activated emotional states is used to drive the robot's physical posture and to select the social animations used when talking. The following table exemplifies the differences between Emys and Glin's interactions for two possible game states.

TABLE I
EXAMPLES OF UTTERANCES FROM EMYS AND GLIN.

Game State	Emys	Glin
Game End: Loss	*"This cannot continue like this! You have to play better!"*	*"No worries, next time we will do better!"*
Playing	*"Watch and learn how this is played."*	*"I am so proud to be on your team!"*

When embodying different social characters that interact verbally, it is crucial to give each one a different voice to allow others to recognise it, especially when they are embodied in identical robots. Therefore, we used different male Portuguese voices from the same Text-To-Speech (TTS) engine to ensure that the two robots had similar voice characteristics in terms of lifelikeness, expressiveness, and quality.

B. Game Play Competency

Despite their different utterances and different TTS voices, they possess identical agent architectures and the same game play competencies. To achieve a natural scenario, the robots must be able to solve the card game autonomously. For this purpose, we provided them with the same algorithmic approach, as described in [6]. Considering that the game is a hidden-information card game and Monte Carlo methods have recently been proposed for solving games of this type, both robots use the Perfect Information Monte-Carlo (PIMC) algorithm, which was initially developed for the bridge domain[14]. The fact that both characters played the game using exactly the same search algorithm, parameters and heuristics is an important design consideration, as we wanted them both to play equally well when placed in the same situation.

However, the game of SUECA contains an element of chance, and players typically do not start the game on equal footing. Depending on the initial distribution of the cards as dealt by the players at the beginning of each game, one team might have a higher probability of winning the game than the other, and one player might have more opportunities to make a good play than the others. This characteristic complicates the task for the robot but simultaneously mirrors the real game experience humans have during card games, making this a very natural and competitive scenario.

C. Interaction in a Group

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

Given that these autonomous robots do not have the capability of understanding natural language, other mechanisms had to be 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 has already said the same thing or because there is no reason to say it 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 started, 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.

IV. STUDY 1: CHARACTER VALIDATION

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

A. 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 the EMYS, 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 took approximately 1 hour, and the instruments used were an EMYS robotic head [20], two video cameras to record the interactions, a multi-touch table, and a

Fig. 1.   Experimental setting for Study 1.

deck of physical cards with printed fiducial markers that could be recognised by the table.

B. 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 could 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. The two characters developed for the robots were validated individually. When finished, the participants were administered a set of questionnaires, filled out the consent form and received a thank-you gift (a movie ticket) at the end of the experiment. 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.

C. 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* [30], 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".
- *McGill Friendship Questionnaire* [23], using three of its dimensions, namely, Help, Motivation and Emotional Security, with scales ranging from "Totally Disagree" to "Totally Agree".
- *Relationship Assessment Scale* [16], adapted to the context and used to ascertain the level of quality of the relationship with the robot, ranging from "Few" to "A lot".
- *Godspeed Questionnaire* [1], 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.

D. Results

To understand whether the two characters were perceived differently, statistical analyses were performed. When a normal distribution was present, we performed 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 II.

For the *Competitiveness Index*, Emys was rated higher than Glin, with a statistically significant difference ($t(25) = -4.893$, $\rho < .001$). 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 ($t(28) = 2.312$, $\rho = .028$), Motivation ($t(28) = 3.686$, $\rho = .001$), and Emotional Security ($t(28) = 3.218$, $\rho = .003$), with Glin presenting higher scores than Emys. On the *Relationship Assessment Scale*, Glin was rated higher than Emys, with a statistically significant difference ($t(28) = 5.514$, $\rho < .001$).

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 greater helpful and motivating capacities 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 ($t(28) = 1.511$, $\rho = .142$). 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 ($U = 40.50$, $\rho = .002$).

TABLE II
STUDY 1 RESULTS: MEANS AND RANKS WITH STANDARD DEVIATIONS FOR THE QUESTIONNAIRE DIMENSIONS COMPARING THE EVALUATIONS OF THE EMYS AND GLIN CHARACTERS. *, $p \leq 0.05$

	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

In general, it seems that our implementations were perceived by the participants as we intended, and Glin was rated as more likeable than Emys.

V. Study 2: Choosing a Robotic Partner

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

A. 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 before interacted with a robot and had a moderate or high level of expertise in the game.

We measured the level of competitiveness of each participant using the Competitiveness Index [30]: 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 $M = 4.50$).
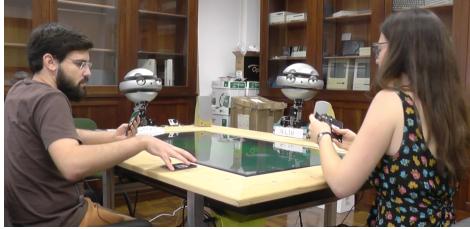
Fig. 2. Experimental setting for Study 2 when each robot was partnering with a human.

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 took 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).

B. 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.

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. After completion, the participants responded to the second part of the questionnaire.

2nd Session: Each participant partnered with one of the robots, which was randomly assigned to him or her, and the
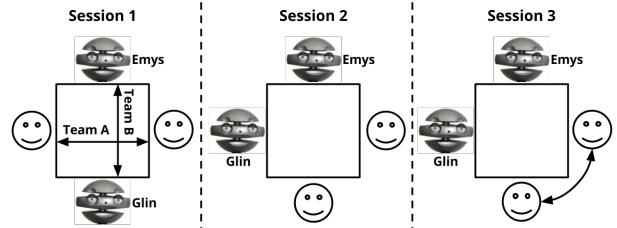

Fig. 3. Experimental setting for Study 2.

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.

C. 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 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 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.

D. Results (I) - Perception of the Robots

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

Regarding the *McGill Friendship Questionnaire*, there were statistically significant differences in the Help ($Z = -5.223$, $\rho < .001$), Motivation ($Z = -6.066$, $\rho < .001$) and Emotional Security ($Z = -5.837$, $\rho < .001$) dimensions, with Glin being rated higher than Emys. For the *Relationship Assessment Scale*, there also was a statistically significant difference ($Z = -4.392$, $\rho < .001$), 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 to have a greater capacity for being helpful and motivating and to provide more emotional security compared with Emys. Moreover, the participants perceived Glin to show 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 (which can play as both partners and opponents).

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 ($Z = -.733$, $\rho = .464$) or the second ($Z = -1.491$, $\rho = .136$). 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 ($Z = -3.451$, $\rho = .001$) and the second ($Z = -6.224$, $\rho < .001$).

TABLE III

STUDY 2 RESULTS: MEANS AND RANKS WITH STANDARD DEVIATIONS FOR THE QUESTIONNAIRE DIMENSIONS COMPARING THE ROBOTS EMYS AND GLIN. 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

E. 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 that 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 ($\chi^2(1) = 4.267$, $\rho = .039$), with more people preferring Glin (63.3%) compared with Emys (36.7%). In this stage 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 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 ($t(58) = 1.242$, $\rho = .219$). 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, as 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.

F. 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 ($\chi^2(1) = 1.667$, $\rho = .197$). 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 ($t(58) = 2.953$, $\rho = .005$), 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—the sum of the points that Glin's team earned in Session 2 + Session 3, independently of its human partners, comparing 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 ($\chi^2(1) = .571$, $\rho = .450$), we found a significant association with the partnering preference using Fisher's exact test ($\rho = .008$). 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 ($\rho = .002$). By contrast, for the participants whose choices did not change, no significant association was found according to Fisher's exact test ($\rho = .409$). 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 ($\rho < .001$), 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 taken into account.

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.

VI. General Conclusions

In this work, 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—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 started 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 two points in time: (1) before having partnered with either robot and (2) after they 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 had 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 Porter's study (2005), that teams whose members prioritise relational features are perceived more positively (e.g., reporting higher levels of supportive behaviour and higher-quality interaction) [25].

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 participants who changed their 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, when Glin was chosen, it was because it was much more relational, whereas Emys was chosen because it was more competitive.

These results have important implications for the creation of robotic teammates who can adapt to their human partners' specific characteristics. Consistent with recent findings [11] 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 be not only investigated but also engineered.

References

[1] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana 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.

[2] Gary Bornstein and Ilan Yaniv. Individual and group behavior in the ultimatum game: Are groups more rational players? *Experimental Economics*, 1(1):101–108, 1998.

[3] Cynthia Breazeal, Cory D Kidd, Andrea Lockerd Thomaz, Guy Hoffman, and Matt Berlin. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 708–713. IEEE, 2005.

[4] Wan-Ling Chang, Jeremy P White, Joohyun Park, Anna Holm, and Selma 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.

[5] Silvia Coradeschi and Alessandro Saffiotti. Symbiotic robotic systems: Humans, robots, and smart environments. *IEEE Intelligent Systems*, 21(3):82–84, 2006.

[6] Filipa Correia, Patrícia Alves-Oliveira, Nuno Maia, Tiago Ribeiro, Sofia Petisca, Francisco S Melo, and Ana 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.

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

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

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

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

[11] Marlena Fraune, Yusaku Nishiwaki, Selma Šabanović, Eliot Smith, and Michio Okada. (to appear) threatening flocks and mindful snowflakes: How group entitativity affects perceptions of robots. In *International Conference on Human-Robot Interaction, HRI*. ACM Press, 2017.

[12] Marlena R Fraune, Satoru Kawakami, Selma Sabanovic, P Ravindra S De Silva, and Michio 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.

[13] Bill Gates. A robot in every home. *Scientific American*, 296(1):58–65, 2007.

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

[15] Victoria Groom and Clifford Nass. Can robots be teammates?: Benchmarks in human–robot teams. *Interaction Studies*, 8(3):483–500, 2007.

[16] Susan S Hendrick. A generic measure of relationship satisfaction. *Journal of Marriage and the Family*, pages 93–98, 1988.

[17] Pamela J Hinds, Kathleen M Carley, David Krackhardt, and Doug Wholey. Choosing work group members: Balancing similarity, competence, and familiarity. *Organizational behavior and human decision processes*, 81 (2):226–251, 2000.

[18] Guy Hoffman and Cynthia 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.

[19] Malte F Jung, Jin Joo Lee, Nick DePalma, Sigurdur O Adalgeirsson, Pamela J Hinds, and Cynthia 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.

[20] Jan Kedzierski, Robert Muszyński, Carsten Zoll, Adam Oleksy, and Mirela Frontkiewicz. Emys –emotive head of a social robot. *International Journal of Social Robotics*, 5(2):237–249, 2013.

[21] Dieta Kuchenbrandt, Friederike Eyssel, Simon Bobinger, and Maria Neufeld. When a robots group membership matters. *International Journal of Social Robotics*, 5(3): 409–417, 2013.

[22] Kwan Min Lee, Wei Peng, Seung-A Jin, and Chang Yan. Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *Journal of communication*, 56(4):754–772, 2006.

[23] Morton J Mendelson and Frances 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.

[24] Clifford Nass and Youngme Moon. Machines and mindlessness: Social responses to computers. *Journal of social issues*, 56(1):81–103, 2000.

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

[26] Byron Reeves and Clifford Nass. How people treat computers, television, and new media like real people and places. *CSLI Publications and Cambridge*, 1996.

[27] Daniela Rus, Bruce Donald, and Jim Jennings. Moving furniture with teams of autonomous robots. In *Intelligent Robots and Systems 95.'Human Robot Interaction*

and Cooperative Robots', Proceedings. 1995 IEEE/RSJ International Conference on*, volume 1, pages 235–242. IEEE, 1995.

[28] Stuart Russell, Peter Norvig, and Artificial Intelligence. A modern approach. *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs*, 25:27, 1995.

[29] Julie Shah, James Wiken, Brian Williams, and Cynthia Breazeal. Improved human-robot team performance using chaski, a human-inspired plan execution system. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 29–36. ACM, 2011.

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

[31] Adriana Tapus, Cristian Ţăpuş, and Maja J Matarić. Userrobot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Service Robotics*, 1(2):169–183, 2008.

[32] Michael L Walters, Kerstin Dautenhahn, René Te Boekhorst, Kheng Lee Koay, Christina Kaouri, Sarah Woods, Chrystopher Nehaniv, David Lee, and Iain Werry. The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 347–352. IEEE, 2005.