MASTER

Does the human-like behavior of a robot evoke action co-representation in a human co-actor?

Pollmann, K.

Award date:
2014

Eindhoven, January 2014

Does the human-like behavior of a robot evoke action co-representation in a human co-actor?

by Kathrin Pollmann

0788109

in partial fulfilment of the requirements for the degree of

Master of Science
in Human-Technology Interaction

Supervisors:

Dr. Ir. Raymond Cuijpers

Prof. Dr. Wijnand IJsselsteijn

ACKNOWLEDGEMENTS

This 'masterpiece' does not only conclude my research work of the last six month, but also an important period of my life. Both, my thesis and the whole master program were interesting and fun, but also challenging and demanding, and I definitely could not have done it without the support of all those people I met along the way:

My participants, especially those who helped out last minute. Without you I would probably never have finished this experiment. I'm glad you enjoyed playing Battleship with the robot!

My first supervisor, who always had an open ear for my concerns and some critical remarks for my ideas. He helped me to nourish my urge to constantly challenge myself, but also reminded me to focus on what was really important whenever I got carried away by some way too complex and adventurous thoughts. Thanks, Raymond, I really enjoyed working with you!

Wijnand, whose research about game experience inspired me to apply for the master program Human-Technology Interaction in the first place and who helped me to find the right path at some important crossroads during my studies.

My friends in Eindhoven who helped my through some difficult moments and made the good moments even more enjoyable. I would especially like to thank: Juliàn, for the physical and mental hugs, in person and over a distance. Adriana, for the beers, the talks, the advice and for staying with me until the end. Andrei, Eva and Oscar, for all the fun and the laughs – it was 'okay'. Veron, for the good times in the office. The Common Room and the Cosmos gang, for giving me a home away from home.

Ich danke meiner Familie, die mich auf all meinen Abenteuern auch aus der Ferne immer begleitet hat. Ihr habt mir Mut gemacht, wenn es mal nicht so gut lief, und euch in guten Zeiten mit mir gefreut. Mami, du verleiht mir Flügel! Papa, du fängst mich auf, wenn ich mal abstürze! Floh, du bist der pragmatische Gegenpol zu meinem um die Ecke denkenden Wissenschaftlerhirn! Ehi, danke, dass du immer an mich geglaubt hast, auch wenn du nie genau verstanden hast, was ich in meinem Studien eigentlich mache! Antin, Silke, Verena (weil man sich Familie eben doch aussuchen kann): Ihr seid die Besten!

ABSTRACT

In our daily life we often have to coordinate our actions with those of others. This requires us to form a mental representation about our interaction partners' actions (action co-representation), which allows us to infer their goals and potential actions, and then synchronize our actions with theirs. It can be assumed that similar cognitive processes should be at work to achieve a successful human-robot interaction. Recent research discovered that individuals show action co-representation for non-biological agents such as robots if they feature some human-like characteristics (Müller et al., 2011; Stenzel et al., 2012). However, those studies manipulated people's beliefs about the human-likeness of the robot through explicit instructions, which is a quite artificial approach and not feasible when humans interact with robots in real-world settings. The present research therefore investigates if action co-representation of robotic actions can also be elicited by human-like behavior expressed by a robot during interaction with a human co-actor. A simple gaming situation was used to expose participants to either an emotionally-expressive or an unemotional robot. Participants' subjective perceptions about the robot were assessed and co-representation effects were measured using the Social Simon Task. The results of our study indicate that the emotional behavior of the robot affected participants' subjective perception about the robot's emotionality, human-likeness and animacy as well as their user experience during the interaction, but did not have a clear effect on co-representation of robotic actions. Potential reasons for these findings and directions for future research are discussed.

Keywords: action co-presentation, Social Simon Task, Social Simon Effect, human-robot interaction, emotional robot

CONTENTS

INTRODUCTION

Research in robotics is becoming increasingly relevant due to demographic trends and decreasing technology costs. Robots are no longer only seen as mere tools to carry out specific tasks, and research is more and more focusing on application areas where robots interact and cooperate with humans in a social way, such as domestic, educational and health care settings. Within this context, the aim of human-robot interaction research is to investigate how to make these social interaction processes as smooth and natural as possible so that, in the long run, users can form long-term relationships with robots which are based on companionship rather than on the robots' functional properties (Pereira, Leite, Mascarenhas, Martinho, & Paiva, 2010).

Realizing this goal comes with a number of challenges, including how to design social robots to trigger automatic, natural interaction behavior in human co-actors. Given that robots are a relatively new technology and people have little experience with them, it is still unclear what kind of behavior people will expect from robots and which cognitive mechanisms are at work to process information about robotic actions.

To approach these questions we can, however, benefit from our knowledge about human-human interaction (Bartneck & Forlizzi, 2004). When socially interacting with members of our own species, we have some automatic intuitions about how they behave in certain situations and why. This common ground of understanding seems to be the underlying mechanism that renders our social interactions effortless and effective. Likewise, in order to be able to successfully interact and cooperate with a robot, it seems to be crucial that the human interaction partners form an appropriate mental representation of the robot's actions which allows them to predict the robot's behavior and then generate a matching motor response (Hegel, Krach, Kircher, Wrede, & Sagerer, 2008; Lee, Lau, Kiesler, & Chiu, 2005).

The present study is therefore aimed at investigating what kind of behavior social robots would have to express to encourage the human co-actor to adopt a mental representation of their actions that promotes natural and effective social interaction processes.

Mental Representations of Human Actions

However, in order to define concrete design principles for the social behavior of robots, we first need to gain a more detailed understanding of how and why humans form mental representations of their co-actors' actions.

In our daily life we encounter numerous situations during which we have to engage in some form of cooperative behavior and adjust our actions to those of another person within a fraction of a second (Cuijpers, Schie, Koppen, Erlhagen, & Bekkering, 2006). Simple everyday tasks, such as buying a coffee at the canteen, are being carried out without deliberation and seem to be easy and effortless for us. Nevertheless, they involve some quite complex cognitive information processing in terms of action planning and motor control. For taking the coffee out of the vendor's hand alone we need to observe him closely, understand that he wants to hand the coffee over to us when he doesn't place the cup on the counter but stretches out his arm towards us, and finally adjust our own behavior accordingly by extending our own arm to grab the cup. As this example illustrates, for action coordination to be successful it is not only crucial to accurately perceive another person's actions, but we also need to incorporate the actions of others in the planning of our own actions (Cuijpers et al., 2006). This requires us to create a mental representation of our co-actor's actions, a cognitive process which is also referred to as action co-representation (Sebanz, Knoblich, & Prinz, 2005).

Literature offers two main theories about how we link observed actions to our own motor system. 'Theory of Mind' suggests that action co-representation is based on a common sense model of human behavior which allows us to automatically attribute beliefs, goals and mental states to our human co-actors (Premack & Woodruff, 1978; Scassellati, 2002). 'Simulation Theory', on the other hand, claims that the observer uses his own action system to calculate and predict the mental processes and actions of others (Goldman, 1992). This claim has received support by the discovery of mirror neurons (Gallese & Goldman, 1998). This class of neurons which was first discovered in the ventral premotor cortex of the macaque monkey has been found to be activated both, when the monkey performs a goal-directed action and when it observes another individual perform the same action (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996), thus providing a neural mechanism for the direct mapping of observed actions onto the observer's motor representations (Sebanz, Bekkering, & Knoblich, 2006). Brain imaging studies provide evidence that the mirror neuron system also exists in humans, showing that neurons in several brain areas fire selectively both during observation and execution of goal-directed actions (e.g. Blakemore & Decety, 2001; Grezes & Decety, 2001). Still, this evidence for 'Simulation Theory' does not necessarily rule out 'Theory of Mind' which provides a more abstract notion of action co-representation. Thus, it cannot definitely be said what exact cognitive processes underlie the formation of shared action representations between humans.

Although this ongoing debate is not solved, it is still possible to explore which factors are involved in producing action co-representation effects. For this purpose Sebanz and colleagues (2005) developed the Social Simon Task (SST) which has proven to be a useful tool to investigate the formation of mental representations of human actions for different social settings. In the SST a classical Simon task (responding to a target stimulus while ignoring its spatial location) is split up between two individuals. Both co-actors have to execute individual go/no-go tasks, each responding to a different target stimulus which is either displayed on the left or right side of the screen. This task does not require taking the other person's actions into account. However, reaction times appear to be slower when the target stimulus is paired with a spatial cue linked to the other person. This reaction time difference which is also referred to as the Social Simon Effect (SSE) is understood as evidence for action co-representation, as it only occurs in the presence of a co-actor, but not when the same task is carried out by a sole person.

Mental Representations of Robotic Actions

Our observations about social human interaction imply that successful human-robot interaction is closely related to the extent to which the human co-actor's mental representation of the robot's actions resembles his model about human actions. If a robot's behavior is co-represented in a similar way as human actions, robotic actions can be expected to initiate social interaction processes which are as fast and automated as in human-human interaction.

Still, it remains to be clarified whether action co-representation is solely tuned to our own species or also activated when we interact with other social agents such as robots. Most studies that investigated this issue suggest that co-representation only occurs during interaction with biological agents (e.g. Kilner, Paulignan, & Blakemore, 2003; Ramnani & Miall, 2003; Tsai & Brass, 2007; Tsai, Kuo, Hung, & Tzeng, 2008). However, more recent research provides evidence that individuals do show co-representation effects for the actions of non-biological agents if they feature some human-like characteristics. More explicitly, it was found that whether an agent's actions are co-represented or not is strongly influenced by the intentionality (Atmaca, Sebanz, & Knoblich, 2011; Liepelt, Cramon, & Brass, 2008) and animacy (Liepelt & Brass, 2010a) attributed to that agent by the human co-actor. Müller et al. (2011), for example, found that action co-representation can be elicited by a wooden interaction partner if people's perceptions about the animacy of that wooden character are manipulated. In a similar study, Stenzel and colleagues (2012) showed that making participants believe that a robot's functional principles were based on a biological, human-like model was sufficient to induce co-representation.

However, all these studies manipulated people's beliefs about the human-likeness of a social agent through explicit instructions or priming, which is a quite artificial approach and not applicable when humans interact with robots in real-world settings such as their homes or care facilities. Hence, the question arises whether co-representation can only be evoked by explicit instructions, or whether the same effect can be induced by manipulating people's perceptions about the human-likeness of a robot during the interaction process. If the mirror neuron system does indeed played a role in the formation of shared action representations among co-actors, we would expect that the behavior displayed by a robot would not only have to be meaningful, but also human-like to induce action co-representation in a human interaction partner.

The Current Study

The current study therefore explores whether co-representation can also be elicited by meaningful, human-like behavior of a robot expressed during the interaction with a human co-actor, and which specific behaviors are suitable to implicitly evoke this effect. More concretely, we investigate to what extent the emotional expressiveness of a humanoid robot can influence human co-actors to adopt a mental representation of its actions which closely resembles their mental representation of human actions.

In human social interaction, emotional expressions play a crucial role, providing useful cues to infer other people's mental states and then form a mental representation of their actions. Likewise, in human-robot interaction it was found that letting a robot display emotions promotes natural and more believable interaction processes (Canamero & Fredslund, 2000; Paiva et al., 2004). This suggests that an emotionally-expressive robot might also appear more human-like than a non-emotional robot, and be more likely to implicitly trigger the human interaction partner to show co-representation effects during the SST which are similar to those evoked by explicit instructions.

Based on these expectations it is hypothesized that participants who are engaged in social interaction with an emotionally-expressive robot will show a higher SSE than participants who are exposed to a non-emotional version of the same robot. Moreover, taking into account the findings of other studies that investigated the action co-representation effects for non-biological agents, we believe that participants will also perceive the emotionally-expressive robot as more human-like and animated than the non-emotional one.

METHOD

In order to investigate how the emotional expressiveness of the robot affects the co-representation of robotic actions, an experiment was designed. In the first part of this experiment we let participants interact with the robot while they were playing the game Battleship together. During this gaming session, the emotional expressiveness of the robot was varied in two different ways: by either letting the robot show emotional reactions to certain game events or not. Subsequently, participants' subjective perceptions about the robot were assessed and the co-representation effect evoked by the robot's behavior was measured using the SST.

Participants

Participants were recruited using the participant database of the Human-Technology Interaction department of the Eindhoven University of Technology and the social network Facebook. In total, 80 participants took part in the experiment, 50 of which were male and 30 female (mean age 23.13, *SD* = 2.46, range 18 to 30). Participants were randomly assigned to the two experimental conditions. Twenty-three participants already had prior experience with the Nao robot and 62 people had played the game Battleship before. Participants received 7.50 Euros compensation for partaking in the experiment.

Design

The experiment was conducted as a between-subjects design. Emotional Expressiveness of the Robot was included as a between-subjects factor and consisted of two levels: One condition featured an emotionally expressive robot (emotional condition), while the other presented a robot that did not express any emotions (unemotional condition). In the emotional condition, Emotional Expressiveness of the Robot was defined as the visible reaction to certain game events during the game Battleship such as hitting, sinking and missing a ship. These emotional reactions were composed of a unique combination of gesture, LED eye pattern, and speech. In the unemotional condition the robot did not display changing behavioral patterns, but maintained the same neutral posture, LED eye pattern and unemotional speech throughout the whole game. Co-representation of Robotic Actions, Perceived Animacy and Perceived Anthropomorphism were introduced as dependent variables.

Experimental Setup

The experiment took place in the UseLab of the Eindhoven University of Technology. For the experiment the 57 cm tall humanoid Nao Robot developed by Aldebaran Robotics was used. To implement its emotionally-expressive behavior the following characteristics of the Nao Robot were of importance: voice synthesis and recognition, the eyes which are composed of a black pupil surrounded by LED lights that can display different colors, and the 25 degrees of freedom which afford various movements of the robot's limbs, head and torso.

In the UseLab two different setups were arranged for the two parts of the experiment: For the first part the robot was placed on a table in front of a Battleship game board. The participant was seated on the other side of the board facing the robot at eye level (see Fig. 1. A).

Fig. 1. A) Photograph of the setup during part one of the experiment showing a typical gaming situation during Battleship. B) Photograph of the setup for the Social Simon Task, part two of the experiment.

Playing Battleship with the robot was realized in the following way: At the beginning of the game the participant was asked to place four ships of various sizes on the reduced game board (6 x 6), while the robot computed the location of its own four ships. The participant then made his first attempt to guess the location of one of the ships of the robot by naming a coordinate (e.g. B5), and received feedback from the robot about whether he hit or missed a ship. If a ship had been hit, the participant continued guessing until he missed for the first time. Then, it was the robot's turn to make a guess. Both co-players took turn guessing until one of them sank the other's four ships and hence won the game. As the current speech recognition fucntion of the Nao robot does not work reliably enough to guarantee a smooth interaction process, a wizard of oz setup was used and participants' verbal responses to the robot were fed manually into the algorithm which controlled the game play and behavior of the robot.

For the second part of the experiment, the SST, the robot was moved to another table where it was placed in front of a keyboard and a 22 inch computer screen. Participants were then seated to the left side of the robot and in front of a second keyboard facing the left side of the screen (see Fig. 1.B). Target stimuli were displayed 13 cm to the left or to the right of the center of the screen.

Manipulation of the Emotional Expressiveness of the Robot

In the emotional condition Emotional Expressiveness of the Robot was manipulated by letting the robot display positive and negative emotions as reactions to certain game events. The game Battleship offers two types of those game events: single events, such as a hit, a miss, the destruction of a ship, winning or losing the game, and consecutive events, i.e. several hits in a row or several misses in a row.

The robot's emotional reaction to such an event was chosen based on a simple model which assumes that, when playing a game, human players experience positive emotions whenever a game event brings them closer to winning the game (positive game events) and experience negative emotions when they are unsuccessful in achieving the goal of the game (negative game events). The model further presumes that those positive and negative emotions can have different magnitudes depending on the type of event, predicting a growing intensity of emotional experience for consecutive game events.

When applying this model to the Battleship scenario, from the robot's point of view, positive game events occur when it hits or sinks a ship, as well as when the human co-player fails to hit its ship, and such events result in the robot being in a positive emotional state. Negative game events, which put the robot in a negative emotional state, are situations when the human co-player hits or sinks a ship or when the robot misses.

For the present study 18 different behavioral patterns were designed to reflect the various positive and negative emotional states of the robot that can be triggered by the different game events of Battleship. As mentioned before, each of these behavioral patterns is composed of a gesture, a LED eye pattern and a verbal expression, which are described in more detail below:

First, ten different gestures were designed to reflect five different intensity levels of positive emotions (happiness, schadenfreude; see Fig. 2) and negative emotions (sadness, frustration; see Fig. 3) respectively. Six of those gestures were based on movements designed by Johnson and colleagues (2013) which already proved to be adequate for letting the Nao robot express positive and negative emotions in a gaming context. In order to create a more diverse pool of behaviors to match the variety of different game events, four additional variations of those initial gestures were created based on findings from related research (Beck, Hiolle, Mazel, & Cañamero, 2010; Haring, Bee, & André, 2011) combined with our own intuition. In addition, three gestures were introduced that did not reflect the emotional state of the robot, but were used to bridge waiting periods and smoothen the interaction during the game (introduction to the game, waiting, thinking).

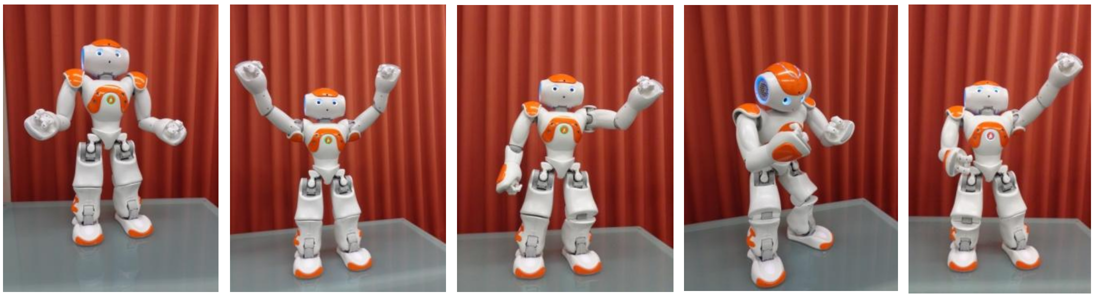

Fig. 2. Positive emotional expressions shown by the Nao robot during the experiment. From left to right: schadenfreude level 1, happiness level 1, happiness/ schadenfreude level 2, happiness level 3, happiness level 4.

Fig. 3. Negative emotional expressions shown by the Nao robot during the experiment. From left to right: Sadness level 1 – 4, frustration.

The valence of the movements was accentuated by matching LED eye patterns chosen based on findings by Johnson, Cuijpers, and van der Pol (2013) and corresponding speech patterns. The latter were composed of 72 verbal expressions which always consisted of a positively or negatively valenced exclamation (e.g. 'Great!', 'Oh no!') and a reference to a game event (e.g. 'You just hit one of my ships."). Based on their semantic meaning and the game event they referred to, those verbal expressions were clustered into groups of four, and each group was then associated with a matching gesture and LED eye pattern to create a behavioral pattern. These patterns would then be displayed whenever the respective game event occurred, whereat the verbal expression uttered by the robot was picked randomly from the group of four in order to introduce some variation in the robot's speech. A full overview of the emotional expressions used during the experiment can be found in Appendix A.

Measures

Subjective Perceptions of the Robot

The Godspeed questionnaire (Bartneck, Kulić, Croft, & Zoghbi, 2009) was employed to assess participants' subjective perceptions of the robot while playing Battleship (see Appendix B). This questionnaire consists of 24 items which comprise the following five dimensions: 'anthropomorphism' (five items), 'animacy' (six items), 'likability' (five items), 'perceived intelligence' (five items), and 'perceived safety of robots' (three items). Each questionnaire item consists of a pair of opposed adjectives (e.g., artificial and lifelike) for which the participant has to indicate on a 5-point Likert scale how well they correspond to his impression of the robot. The two dimensions 'anthropomorphism' and 'animacy' reflect the extent to which the robot appears human-like and alive and are hence of special interest for our research.

In the current study a slightly modified version of the Godspeed questionnaire was used. The item 'moving rigidly – moving elegantly' had to be removed from the 'anthropomorphism' scale, as it was not applicable to the unemotional condition. Moreover, to facilitate a manipulation check, 'emotionality' of the robot was added as a sixth dimension to the Godspeed questionnaire. It consists of the two items 'unemotional – emotional' and 'insensitive – sensitive' which had to be rated on the same 5-point Likert scale as the other items.

Co-representation of Robotic Actions

Co-representation of the robot's actions was measured by analyzing participants' performance during the Social Simon Task (SST; Sebanz, Knoblich, and Prinz, 2005). In the present study the second human actor was replaced by the Nao robot. Participants received instructions to respond as fast and accurately as possible to the appearance of a white square, while the robot would respond to the appearance of a white diamond. The targets were presented one at a time and either appeared on the left side of the screen, i.e. in front of the participant (congruent trial), or the right side of the screen and thus in front of the robot (incongruent trial). Each participant completed 400 trials (200 go and 200 no-go trials), separated by a short break of 20 seconds after every 100 trials. Every trial was preceded by a 250 ms fixation display showing a white fixation cross in the center of a black screen. Then the target was shown on the screen for 500 ms, whereat stimulus type and location were randomly chosen. Participants' reaction times (RT) were recorded from the onset of the target. After an inter-trial interval of 1300 ms, during which a mask was displayed, the next trial started (Fig. 4).

The response of the robot was controlled by the same program that ran the SST. The robot's right hand was placed slightly above the space bar of the keyboard and moved down when the diamond was shown, thereby initiating a button press with the thumb of the robot's hand (Fig. 1.B). The robot's RT were adjusted to the average human RT for a go/no-go task and varied randomly between 350 and 450 milliseconds (ms). The Social Simon Task was realized using PsychPy 1.7 (Peirce, 2007).

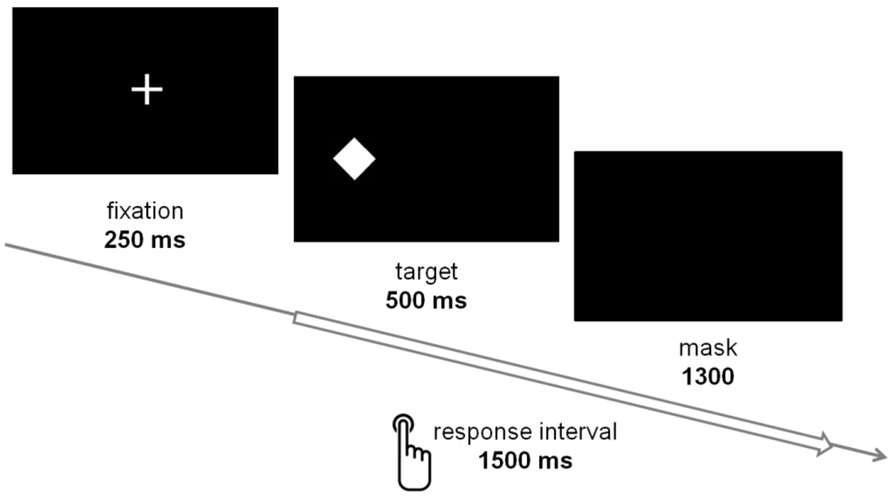

Fig. 4. Illustration of the stimulus presentation for a no-go trial during the SST.

Dispositional Empathy

Corepresenting someone else's actions is closely related to a person's ability to empathize with their co-actors. To control the potential impact of individual dispositional empathy on the results of the SST, the Interpersonal Reactivity Scale (IRI; Davis, 1983) was introduced as an additional measure. The questionnaire comprises four dimensions which consist of seven items each: the 'perspective-taking scale', the 'fantasy scale', the 'empathic concern scale' and the 'personal distress scale'. Responses for the items ranged from (0) 'does not describe me well' to (4) 'describes me very well' (see Appendix C).

Participants' Interactive Behavior and Enjoyment

In addition to subjective perception and cognitive performance measures, each experimental session was video-taped. These recordings allowed us to assess participants' reactions to the robot's behavior, in particular their laughs and verbal responses, in a quantitative as well as qualitative way, thus providing additional insight into participants' user experience while interacting with the emotionally-expressive compared to the unemotional robot.

Procedure

After participants were guided to the experiment room, they signed the consent form and the experimenter described the structure of the experiment to them. The rules of Battleship were explained and participants were given the opportunity to ask questions.

During the first half of the experimental session, participants were asked to play one or two rounds of Battleship against the robot. It was decided that participants should be exposed to the robot for at least 10 and at most 20 minutes (on average 14.5 min). Thus, whenever the first round of Battleship lasted less than ten minutes, an extra round was added. Subsequently, participants had to fill in the Godspeed questionnaire.

In the second half of the experiment participants performed the SST with the robot and filled in the IRI as well as a general questionnaire inquiring about their age, gender and prior experience with the Nao robot and the Battleship game.

Every session lasted about 45 minutes and concluded with a final debriefing of the participant. The study took place from December 2013 until January 2014 and was conducted in accordance with the ethical standards laid down in the 1975 Declaration of Helsinki.

RESULTS

In this section the data gathered during the experiment is analyzed and the corresponding results are reported: First of all, we examined whether the two experimental conditions succeeded in manipulating our participants' perceptions of the robot as intended, i.e. whether participants in the emotional condition experienced the behavior of the robot to be more emotional than participants in the unemotional condition. Then, it was investigated whether the conditions differed with regard to dispositional empathy and prior experience with the Nao robot as well as the game Battleship. After that, participants' subjective perceptions about the robot were compared across the two conditions, and finally, the effect of Emotional Expressiveness of the Robot on the Co-representation of Robotic Actions was analyzed.

Two participants had to be excluded from further analysis because the robot stopped working during Battleship which terminated the game before the required minimum gaming time of ten minutes. The data analysis was performed using Microsoft Excel 2010 (Microsoft Corporation, 2010) and IBM SPSS Statistics 22 (IBM Corporation, 2013). For all statistical analyses the significance level was set to a value of p = 0.05 with a confidence interval of 95%.

Manipulation Check

To determine whether our manipulation of the Emotional Expressiveness of the Robot worked as intended, the mean scores for the 'emotionality' dimension of the extended Godspeed questionnaire were compared among the two experimental conditions. A check of the internal consistency of the scale revealed a high reliability for 'emotionality' (Cronbach's α = .83). The results of a subsequent paired t-test indicate that participants in the emotional condition (M = 3.68, SD = .96) perceived the behavior of the robot to be significantly more emotional than in the unemotional condition (M = 2.26, SD = .91), which suggests that the manipulation was successful; $t(75)$ = 6,70, p < .001 (Fig. 6).

Confounding Variables

Participants' dispositional empathy was assessed as a potential confound using the four dimensions of the IRI. The 'personal distress' and 'perspective-taking' scales proved to be reliable (both Cronbach's α > .74), while the 'fantasy' and 'empathic concern' scales only showed an acceptable internal consistency after the removal of one item from each scale (both Cronbach's α > .72). To investigate the differences between the two conditions with regard to the four IRI dimensions a two-group multivariate ANOVA was performed. The results suggest that ratings for the four dimensions did not differ significantly between conditions; $F(4,71) = 3.82$, $p = .821$ (Table 1).

Besides dispositional empathy, participants' prior experiences with the robot and with the Battleship game were compared across conditions. Both, experience with the robot and with Battleship were equally distributed among conditions; $\chi^2(1) = 3.55$, $p = .059$ and $\chi^2(1) = .905$, $p = .341$ respectively. It can hence be assumed that participants' prior experience did not confound the effect of Emotional Expressiveness on the Co-representation of Robotic Actions.

Table 1. Overview of the results of the multivariate ANOVA testing the between-subject effect of Emotional Expressiveness of the Robot on participants' ratings for the four IRI questionnaire dimensions (N=76).

IRI dimension	Emotional condition		Unemotional condition		Effects of the Emotional Expressiveness of the Robot		
	M	*SD*	*M*	*SD*	*df*	*F-statistic*	*p*
Fantasy Scale	2.67	.88	2.50	.87	1	$F(1) = .90$	.347
Empathic Concern Scale	2.48	.60	2.45	.55	1	$F(1) = .05$	.833
Perspective-Taking Scale	2.43	.69	2.42	.65	1	F(1) = .01	.964
Personal Distress Scale	1.82	.72	1.85	.53	1	F(1) = .05	.831

Subjective Measures

Participants' subjective perceptions about the robot were measured using the five dimensions of the Godspeed questionnaire: 'anthropomorphism', 'animacy', 'likability', 'perceived intelligence' and 'perceived safety'. A reliability check revealed that all dimensions displayed satisfactory reliability (all Cronbach's α > .70), except for 'perceived safety' which showed extremely low reliability (Cronbach's α = .05) even after the individual removal of each of its three items. Since 'perceived safety' is considered of minor relevance to the present study, this dimension was excluded from further analysis.

A two-group multivariate ANOVA revealed a significant effect of condition on the Godspeed questionnaire scores; $F(4,70) = 3.32$, $p = .015$ (Fig. 5). To examine the differences between conditions for each questionnaire dimension individually a series of t-tests was conducted. For this follow-up analysis the significance level was Bonferroni-adjusted to $p = .013$. Results indicate that the 'animacy' of the robot was perceived as significantly lower in the unemotional (M = 3,02, SD = .72) than in the emotional condition (M = 3.55, SD = .70); $t(75) = 3.27$, $p = .002$. Likewise, participants' ratings for 'anthropomorphism' were found to be noticeably higher for the emotional (M = 3.09, SD = .82) compared to the unemotional condition (M = 2.75, SD = .70), but this difference did not reach significance; $t(75) = 1.93$, p=.057. 'Likability' and 'perceived intelligence' received generally high ratings, but the scores did not differ significantly among conditions (both p>.08).

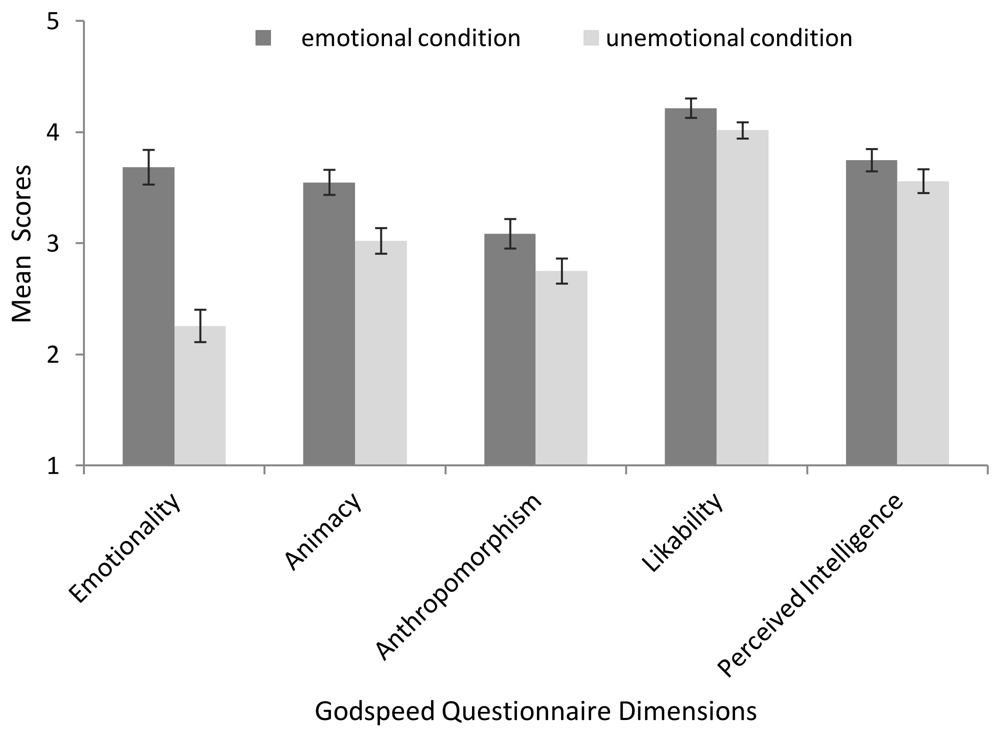

Fig. 5. Results for five dimensions of the extended Godspeed questionnaire for the two conditions. Error bars show the standard error for the mean.

Reaction Time Analysis

To investigate the effect of the Emotional Expressiveness of the Robot on the Co-representation of Robotic Actions, the reaction times (RTs) obtained from the go-trials of the SST (indicating the responses to the appearance of the square) were analyzed. Prior to the statistical analysis, all error trials were removed from the data set (44 out of 16000 trials; 0.28%). Error trials were defined as RTs above 1500 ms and non-response trials. Moreover, all RTs below 100 ms were excluded from further analysis (2 out of 16000 trials; 0.01%), as this is the minimum time needed for physiological processes such as stimulus perception and motor responses (Whelan, 2010).

For all participants, RTs were coded as congruent (square displayed on participant's side of the screen) and incongruent (square displayed on robot's side of the screen) and the median RTs for congruent and for incongruent go-trials were calculated. After that, the Social Simon Effect (SSE) score (in ms) was determined for each participant individually by subtracting the median RT for congruent trials from the median RT for incongruent trials. Two participants had to be excluded completely from the data analysis due to their remarkably poor performance on the SST, namely an error rate above 60% and a SSE score of 2.5 SD above the sample mean (Müller et al., 2011). Hence, in total, data from 76 participants was included in the final analysis.

The Social Simon Effect

Before entering them into a statistical analysis we examined participants' SSE scores closer with regard to measures of central tendency as well as their distribution across the two conditions: The mean SSE score was found to be slightly higher for the unemotional condition (M = 3.03, SD = 13.70) compared to the emotional condition (M = 0.45, SD = 14.90). The results of a Shapiro-Wilk test indicated that the SSE scores of the two conditions were not normally distributed (both $p < .05$). To gain further insight into how, if not normally, the SSE scores were distributed, they were first rounded to whole milliseconds and entered into a histogram which revealed that the scores for both, the emotional and the unemotional condition showed a multimodal distribution, featuring one large peak at a score of zero, and two smaller at SSE scores of -16 and +16 (Fig. 6).

The histograms already suggest that both distributions are rather similar. Still, this question was further explored by a subsequent statistical analysis of the data: The two conditions and the SSE scores were subjected to a Chi-squared test for independent samples, which did not reveal any significant difference between the distributions of

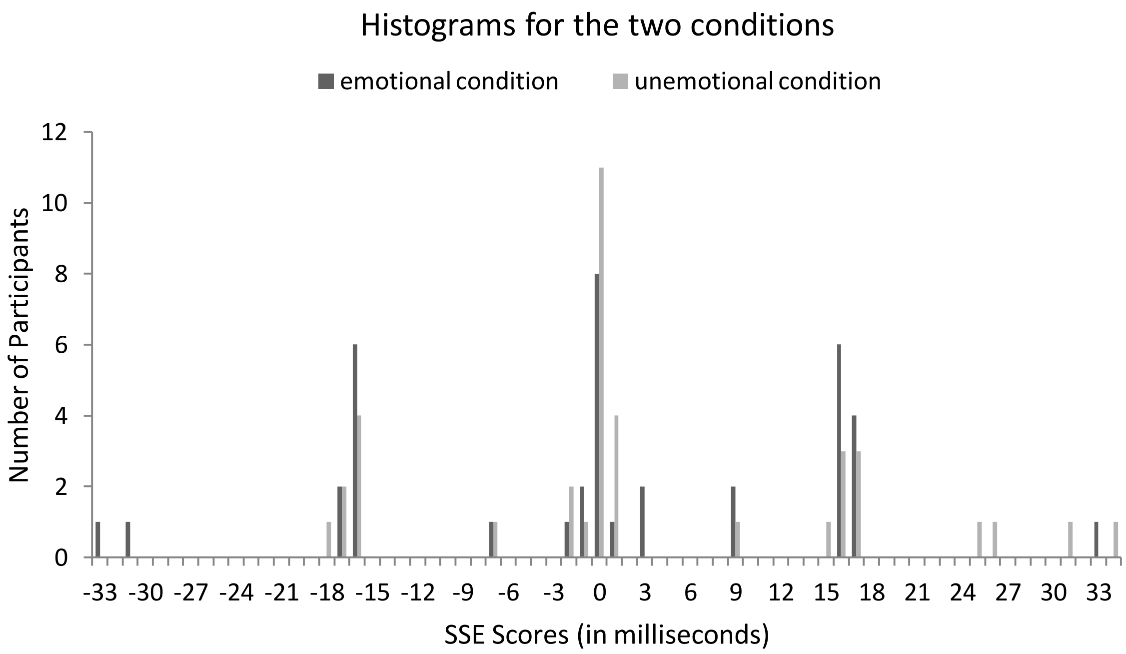

Fig. 6. Histograms for the SSE scores of the two conditions.

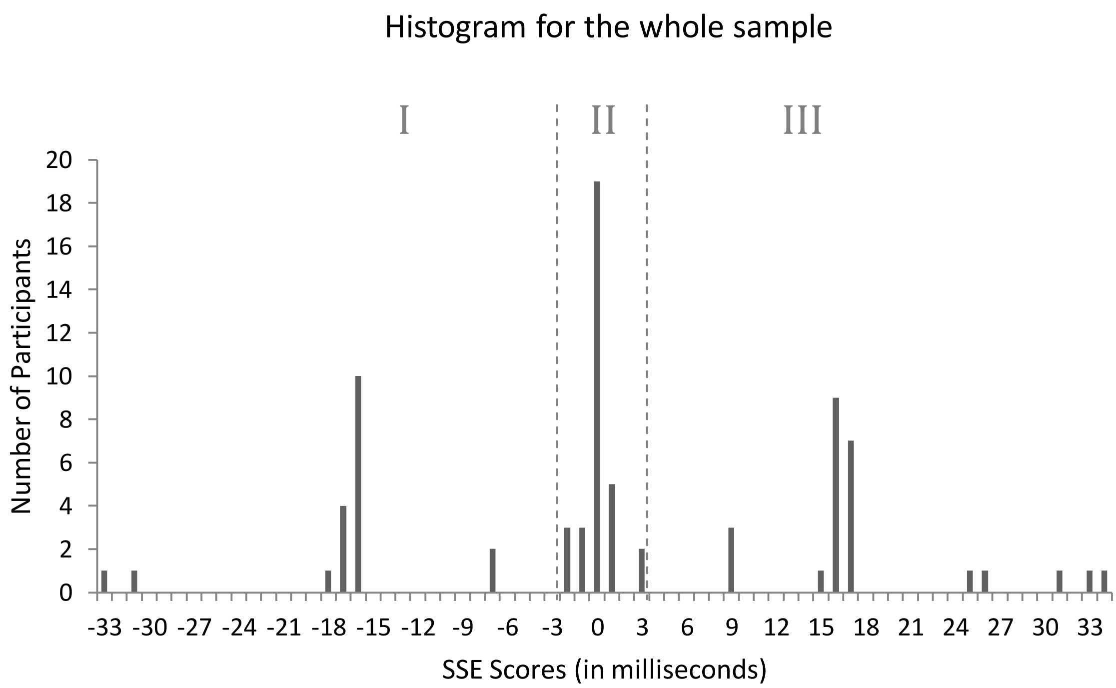

Fig. 7. Histograms for the SSE scores of the whole sample. Roman numerals depict the clustering of participants into subsets based on their SSE scores: expected SSE (III), no SSE (II) and a reversed SSE (I).

the emotional and unemotional condition; $\chi^2(13) = 9.39$, $p = .743$. Likewise, a follow-up Mann-Whitney-U-test showed that the SSE scores were equally distributed in both conditions; $U = 681.0$, $p = .670$. All subsequent analyses will therefore be based on the distribution of the whole sample (Fig. 7).

The histogram shows that the majority of the participants did not show any SSE (42%). Still, a subset of participants did show the expected SSE (33%), whereat, based on effect sizes found by Stenzel et al. (2012), 'showing the SSE' was defined by a score larger than three. Furthermore, another subset showed a reversed SSE (25%).

To be able to characterize potential cognitive mechanisms underlying the occurrence of positive and negative SSE scores, the co-representation effects of the three subsets were examined more closely. More concretely, it was investigated whether the difference between incongruent and congruent trials, may it be positive or negative, was caused by response facilitation (indicated by faster RTs for one trial type) or response inhibition (indicated by slower RTs for one trial type). For that purpose, the mean RTs of the three subsets were plotted for congruent and incongruent trials (Fig. 8). The resulting diagram revealed that participants who showed the expected SSE and those who showed the reversed SSE displayed an increase in RT for incongruent and congruent trials respectively. In both cases, they were never faster than the subset of participants who did not show any SSE, which indicates that the SSE and reversed SSE are products of response inhibition, but not

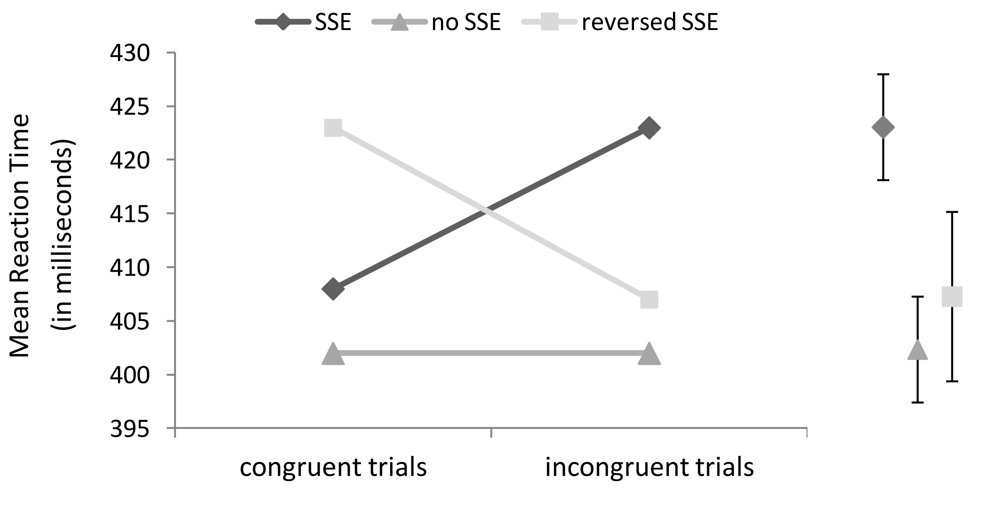

Fig. 8. Mean RTs for the congruent and incongruent trials of the three groups formed based on SSE scores. Error bars show the standard error for the mean. All groups displayed the same mean standard errors for both, congruent and incongruent trials.

response facilitation. Moreover, in both groups, mean RTs for trials with fast stimulus-response mapping did not diverge markedly from the average times produced by participants who did not show any SSE. A comparison of the mean effect sizes of the SSE (20 ms) and the reversed SSE (23 ms) indicated that they did not differ significantly; $t(42) = -.58$, $p = .566$.

Confounding Variables - Revisited

Our initial analysis of potential differences between the emotional and unemotional conditions with regard to our confounding variables showed that the scores for the IRI dimensions as well as prior experience with the Nao robot and the Battleship game were equally distributed across conditions. However, no effect of Emotional Expressiveness of the Robot on Co-representation of Robotic Actions was found, which limits the informative value of our initial analysis. To investigate potential direct effects of the confounding variables on participants' action co-representation, we entered them into a second analysis. No difference in SSE scores was found for participants with and without prior experience with the Nao and Battleship (both $p > .7$). The statistical analysis did, however, reveal a significant negative correlation between SSE scores and the ratings for the 'fantasy' scale of the IRI ($r(74) = -.388$, $p = .001$), indicating higher 'fantasy' scores for participants who showed a reversed SSE.

In addition to our initial confounding variables we also investigated the relationship between participants' performance on the SST their ratings for the extended Godspeed questionnaire. The statistical analysis did not, however, reveal any significant correlation between the four questionnaire dimensions and participants' SSE scores (all $p > .12$)

Error Rates

In addition to the SSE score analysis, participants' error rates were examined. There are two types of errors that can occur during the SST: Participants can either fail to respond to a go-trial within the given time limit of 1500 ms (omission) or they can respond to a no-go-trial (false alarm). To compare the error rates across the two conditions a repeated measures ANOVA was performed, including type of error as a within-subject factor and condition as a between subject factor. Results indicated that participants made significantly more false alarm errors than omission errors ($F(1,74)$ =

106.11, p < .001), but there was no effect of condition or the interaction between the two factors (both p>.6). In general, error rates were found to be rather low (Table 2).

Table 2. Overview of the mean omission and false alarm error rates for the two conditions (in %; N=76).

Error Type	Emotional condition		Unemotional condition	
	M	*SD*	*M*	*SD*
Omission	0.34	1.22	0.25	0.53
False Alarm	3.80	2.94	4.04	3.60

Participant-initiated Interactions and Enjoyment

Our video analysis suggests that the emotionally-expressive robot triggered more interactive behavior in participants than the non-emotional version. More than half of the participants in the emotional condition (54%) showed at least three verbal responses to the robot that were not related to giving game-related feedback and three of them kept chatting to the robot throughout the whole game. In the unemotional condition only five participants (12%) used verbal expressions that were not related to the feedback they were required to provide as part of the game.

Moreover, it was observed that, although everybody received exactly the same instructions including the specific phrasing for feedback of the robot ('hit', 'water', 'sunk'+ship-name), participants verbal responses deviated markedly from these pre-defined commands: In the emotional condition only 16% of the participants closely followed the instructions. The other 84% started off the game session using the words they were instructed to use, but then showed an increasing variation in phrasing their feedback (e.g. "It is not a hit.", "No, you are unlucky again."). In the unemotional condition, on the other hand, participants either exactly followed the instructions or started to adopt the exact same wording as the robot.

After the experiment, 25 participants of the emotional condition (64%) informed the experimenter on their own initiative that playing the game with the robot was fun. In the unemotional condition, comparable statements were only given by 15% of the participants. Participants' verbal feedback about their enjoyment was supported by the observation that 63% of the participants in the emotional condition laughed at least twice during the game, while laughs were only observed for two participants in the unemotional condition (5%).

DISCUSSION

Humans are remarkably skilled in coordinating their actions with those of others. This skill is based on their ability to automatically form a mental representation of their co-actors' actions which can then be used to predict their behavior and adjust one's own motor response accordingly. Based on these observations from human social interaction it can be assumed that similar cognitive processes could facilitate action coordination in human-robot interaction, thereby promoting more natural interaction processes between humans and robots. In this context, recent studies have evidenced that manipulating people's beliefs about the human-likeness of a non-biological, social agent through explicit instructions can promote action co-representation effects (Müller et al., 2011; Stenzel et al., 2012). The current study is based on these findings and aimed at extending existing research by investigating whether similar effects can be implicitly evoked by human-like behavior of a humanoid robot expressed during the interaction with a human co-actor. For this purpose participants were exposed to two different versions of the same robot which either showed emotional behavior or not. Co-representation of robotic actions was then assessed using the Social Simon Task (SST).

Subjective Perceptions of the Robot

The results of the experiment indicate that the two levels of Emotional Expressiveness of the Robot successfully manipulated participants' explicit perceptions about the robot: As intended, participants perceived the robot as more emotional in the emotional than in the unemotional condition. Moreover, our findings indicate that participants in the emotional conditions perceived the robot as significantly more animate and more human-like than in the unemotional condition. Although the difference between mean ratings of the two conditions for the 'anthropomorphism' dimension of the Godspeed questionnaire did not reach significance, it should not be undervalued: Despite its humanoid shape, participants in the unemotional condition perceived the Nao robot as rather machine-like (mean 'anthropomorphism' rating < 3), while participants in the emotional condition tended to evaluate it as rather human-like (mean 'anthropomorphism' rating > 3).

Based on these findings, our hypothesis that participants would perceive the emotionally-expressive robot as more human-like and animated than the non-emotional version can be accepted. It can thus be concluded that our attempt to implicitly manipulate people's beliefs and perceptions of the robot by letting it display emotional behavior was successful. However, so far our findings only provide evidence for a successful manipulation of participants' explicit perceptions of the robot assessed by the Godspeed questionnaire, which does not automatically imply that the behavior of the robot also triggered action co-representation.

Effects of the Emotional Expressiveness of the Robot on Action Co-representation

Existing research about action co-representation for social agents suggests that the magnitude of the SSE is strongly influenced by people's perceptions about the animacy and human-likeness of this agent (Liepelt & Brass, 2010b; Müller et al., 2011; Stenzel et al., 2012). Still, the current study fails to replicate these findings: Although participants' perceptions about the emotional expressiveness of the robot were successfully manipulated, the robot's behavior did not influence people's performance on the SST, and SSE scores did not differ among conditions.

Mean SSE scores for the two conditions were found to be close to zero, which, at first sight, suggests that the actions of the robot were not co-represented at all. However, the analysis of the frequency distribution of SSE scores revealed that only 42% of our sample did not show any co-representation effect. There was a subset of 25 participants (33%) who did show the intended SSE, but their contribution to the average effect size was equaled out by a third subset of 19 participants (25%) who displayed quite pronounced negative SSE scores. So, the SST did reveal some reaction time differences between congruent and incongruent trials, but those were more diverse than expected.

In order to explain these findings it is important to recap the basic rational underlying the experimental design of the present study. Based on prior research, we argued that letting the robot express emotional behavior would make it appear more human-like and animate, and that attributing human-like characteristics to a robot would evoke co-representation of this robot's actions. The manipulation we used was thus an indirect manipulation based on the assumption that the effect of Emotional Expressiveness of the Robot on Co-representation of Robotic Actions is mediated by the impression people form about the human-likeness and animacy of the robot. However, despite the successful manipulation of participants' subjective perceptions about the robot, we did not find an effect of Emotional Expressiveness of the Robot on Co-representation of Robotic Actions. Thus, there seems to be a mismatch between participants' explicit perceptions about the robot reported on the Godspeed questionnaire and the more implicit impression about the robot's human-likeness that underlie the SSE.

Following this line of reasoning we propose two hypotheses about why the emotional behavior of the robot did not have the expected effect on the co-representation of robotic actions:

1) In both experimental conditions, the behavior of the robot (emotional or unemotional) caused a subset of the participants to adopt an appropriate mental model of the robot's human-likeness, which then led to action co-representation effects.
2) Emotional expressiveness caused only participants in the emotional condition to adopt an appropriate mental model of the robot's human-likeness, but this mental model was not adequately assessed by the SST.

In the following section we will further elaborate on both hypotheses.

Attributing Human-like Characteristics to the Robot: a Bottom-up Approach

The research we based our study on used explicit instructions and priming to directly influence participants' impressions of a non-biological interaction partner (Müller et al., 2011; Stenzel et al., 2012), thus affecting the believed human-likeness of the robot and the resulting action co-representation effects in a top-down manner (Stenzel, Chinellato, Del Pobil, Lappe, & Liepelt, 2013).

In contrast, the current study exposed participants to a real-life social situation during which participants had to gather information from social cues provided by the environment in a bottom-up process in order to arrive at a mental model about the robot's human-likeness. These cues included but were not restricted to the emotional expressiveness of the robot. When designing the behavior patterns for the emotional robot we tried to render the emotionality of these behaviors as salient and unambiguous as possible to promote emotional expressiveness as the key cue that participants would base their mental model of the robot on. Our manipulation check shows that the behavior of the robot was convincing. Still, it is possible that for a subset of participants the emotional behavior of the robot was not the main factor they based their mental model about the robot on. In both experimental conditions, other cues about the robot's human-likeness were present, which might have influenced people's impression of the robot.

First of all, the model underlying the emotional behavior expressed by the robot in the emotional condition (as response to specific Battleship game events) was not sophisticated enough to prevent that its deterministic nature was revealed to the human co-actor throughout the game session. Although we designed various behavioral patterns to embrace the variety of emotional states produced by the different game events, due to the random choice of speech output and recurring game events, the behavior of the robot sometimes appeared to be rather repetitive. Moreover, the range of emotions expressed by the robot was limited to only four (happiness, schadenfreude, sadness and frustration), while one would probably expect a wider range of emotions from a human co-actor.

Likewise, it is possible that participants in the unemotional conditions attributed human-like characteristics to the robot, even though we tried to reduce the behavior displayed by the robot to a minimum. However, in order to be able to play Battleship with a human co-actor the robot was required to show some basic behavior which contained additional cues that might have caused participants to experience the robot as more human-like than intended. The Nao robot has a humanoid shape and is often perceived as 'cute' and 'pleasant' by human-coactors, which might have influenced participants' impression of its human-likeness. Furthermore, the version of the Nao presented during our experiment was able to simulate human visual and auditory perception quite convincingly. Participants were made believe that these sensory processes were realized using the camera and speech recognition system provided by the Nao, while they were in fact faked by the experimenter. The robot's flawless visual perception during the SST was achieved by letting the program that controlled the task send a command to the robot as soon as the diamond was displayed. Speech recognition of the robot was simulated by the experimenter using a wizard of oz setup and therefore closely resembled human auditory perception in terms of speed and accuracy.

All in all, it appears that the extent to which participants took the (lack of) emotionality of the robot into account when forming a mental model about its human-likeness varied from person to person. It is possible that these individual differences in integrating various social cues into a mental model about the robot's human-likeness might have led to the diversity in SSE scores found in the present study.

Validity of the SST as a Measure of Action Co-representation

The SST is a popular method to assess action co-representation based on people's behavioral responses. However, a recently published study by Dolk, Hommel, Prinz, and Liepelt (2013) claims that an event does not necessarily have to be of social nature to produce the SSE, thus questioning the validity of the SST as a tool to measure action co-representation effects. This claim is supported by several experiments during which the authors let humans perform the SST together with various (non-social) objects (Chinese waving cat, clock, and metronome). The results of these empirical studies indicate that the SSE can be elicited by any object that produces an event which is salient enough to attract the attention of the human co-actor.

Their findings seem, however, to be counterintuitive to the present study and other related research (Müller et al., 2011; Stenzel et al., 2012). In these studies a social agent was present for all SST sessions, always producing the same event: a button click. According to Dolk and colleagues (2013), you would thus expect that the SSE should either be present or absent in the whole sample, depending on whether this event was experienced as a salient event or not. Still, even though all participants were exposed to the same event, the SSE was only found for one of the experimental conditions, and these findings cannot be accounted for by the study by Dolk et al. (2013). Moreover, the scope of their results is limited by their definition of a 'salient event' as a 'rhythmic auditory event', and it remains questionable if they can be generalized to visual versions of the SST at all.

In conclusion, it is doubtful whether the study by Dolk et al. (2013) provides sufficient evidence to come to the conclusion that the SST is not a valid tool to measure action co-representation. At the end of their paper, the authors themselves admit that the SSE appears to be sensitive to the socialness of a situation and might be useful to assess changes therein. As we believe that the button click produced by the robot during the SST in our experiment was equally salient for all participants, the mere salience of the button click does not provide a sufficient explanation for why SSE scores did not differ between emotional and unemotional conditions. Thus, we assume that the SST adequately assessed participants' co-representation of the robot's actions. It should, however, be noted that the results obtained by SST research partially contradict each other and that it is possible that action co-representation is not the only cognitive mechanism underlying the SSE.

In summary, it seems more likely that the diversity in magnitude of the co-representation effects uncovered during the SST can be attributed to differences among participants' implicit impressions of the human-likeness and animacy of the robot rather than to a lack of validity of the measurement tool itself. Still, this explanation only sheds light into why participants might have obtained positive and zero SSE scores during the SST, but does not account for negative SSE scores.

The Social Simon Effect and its Reversal

The frequency distribution of participants' SSE scores derived from the SST does not match our expectations. We expected the SSE scores to be normally distributed across our sample, featuring a low mean score if participants engaged weakly in action co-representation processes and a high mean score if they strongly co-represented the robot's actions. However, our data shows a multimodal distribution including a surprisingly big subset of participants with negative SSE scores. One could argue that those negative scores are artifacts produced by our implementation of the SST. This seems, however, unlikely as such artifact would be a systematic effect and also be equaled out over the 200 'go' trials performed by each participant. Moreover, negative SSE scores have also been found in related research using the SST (e.g. Dolk et al., 2011, 2013; Stenzel, Chinellato, Del Pobil, Lappe, & Liepelt, 2013), but they are generally left uncommented and interpreted as the absence of action co-representation. However, given that one quarter of our participants obtained negative SSE scores of similar magnitude as the positive SSE scores we found for the SST, we believe that this reversed SST deserves some further investigation.

While research has so far disregarded the reversed SSE, reversed difference scores between incongruent and congruent response trials did receive some attention in research about the standard Simon Task. Hedge and Marsh (1975) first demonstrated the reversed Simon Effect (SE) using a standard Simon Task during which participants had to press a red button on their left to respond to a red stimulus and a green button on their right to respond to a green stimulus. When the response rule was changed, so that the red button had to be pressed as a response to a green stimulus and vice versa, participants' RTs were found to be slower for congruent trials. Hedge and Marsh (1975) accounted for this unexpected phenomenon using the 'logical recoding hypothesis'. Taking into account that during this version of the Simon Task the target stimuli and the responses featured two attributes each, color and position, Hedge and Marsh (1975) argued that

> "the logical character of the recording which would relate either of these [stimulus] attributes to the attributes of the response might be either 'identity' (same colour or same position) or 'reversal' (alternate colour or alternate position). […] For a given logical recording (identity or reversal) of the relevant attribute (colour) responding was faster for trials in which the recoding of the irrelevant attribute (position) was of the same logical type as that of the relevant attribute, than for trials in which the logical recording of the irrelevant attribute was opposite in type." (p. 235)

The 'logical recoding hypothesis' appears to be a straightforward approach to account for reversed effects in a standard Simon Task. However, it cannot be used to explain the same effects for a SST carried out together with a robot, as this version of the task only requires the human co-actor respond to one type of target stimulus (in our case the square). A logical recording of a response as 'reversal' (alternate stimulus indicated by the diamond and alternate position indicated by the robot's side of the screen) would therefore be an error trial and not included in the calculation of the SSE.

A more elaborate examination of potential mechanisms underlying the reversed SE is provided by the 'dual-process model' developed by De Jong, Liang, and Lauber (1994) which can be seen as an extension of the 'logical recoding hypothesis'. Although this model received some criticism regarding the value it adds to the 'logical recoding hypothesis' (Zhang & Kornblum, 1997), it still provides a general explanation for the occurrence of negative and positive SE scores claiming that the SE and its reversal are products of several competitive cognitive processes. Applying this 'dual process model' to a Social Simon Task setup, it could be argued that the valence of the SSE is not a product of action co-representation alone, but can be traced back to several competing cognitive processes. According to this model, positive SSE scores could be the result of action co-representation processes, while negative SSE scores might be evoked by a completely different cognitive process. The valence and magnitude of the SSE would then be determined based on the cognitive process that is activated more strongly for the individual person during the SST, whereat equal salience of the two processes eliminates the SSE. This would also mean that a negative SSE score does not imply that a person did not engage in co-representation at all, but rather that the effect of action co-representation on the performance on the SST was equaled out by another cognitive process. This hypothesis is further substantiated by our data analysis which revealed that the reversed SSE was caused by response inhibition for congruent trials, thus implying that participants must have engaged in additional cognitive processes while responding to target stimuli displayed on their own side of the screen. Additional support is also provided by Dolk et al. (2013) who propose that the SSE is not a purely social effect, thereby suggesting that additional cognitive processes are at work when carrying out a SST.

As the present study was solely focused on investigation action co-representation effects, no conclusions can be drawn about potential additional cognitive mechanisms that might underlie negative SSE scores. It should, however, be mentioned that we found a significant negative correlation between the 'fantasy' scale of the IRI and participants' SSE scores, indicating that participant's with a high ability to take the perspective of a fictitious character showed a reversed SSE. Future research about SSTs with robots should hence control for participants' 'fantasy' ability. It might also be interesting to further investigate what kind of cognitive processes might be triggered when people with high IRI 'fantasy' scores interact with a robot, and how these processes contribute to the reversed SSE. To do so, we recommend using a multi-method approach combining the SST with brain imaging, which will allow for accessing both, people's behavioral manifestations of action co-representation and potential cognitive processes underlying performance on the SST.

Design Implications for Social Robots

The present study is aimed at investigating whether the emotional behavior of a humanoid robot encourages a human co-actor to adopt a mental representation of the robot's actions which promotes natural and effective social interaction processes. Although our experiment did not succeed in demonstrating the expected effect of Emotional Expressiveness of the Robot on Co-representation of Robotic Actions, it still brought up interesting new insights which can serve as a basis for further research on the design of social robots.

First of all, it was found that the emotional behavior of the robot influenced participants' subjective perceptions to a great extent. Questionnaire results indicated that participants who were exposed to the emotionally-expressive robot experienced the robot as an emotional, animate and human-like social agent.

Furthermore, the present study supports research by Canamero and Fredslund (2000) as well as Paiva and collegues (2004) who showed that emotional behavior of a robotic co-actor promotes more natural and believable interaction processes between humans and robots. The video recordings of the Battleship gaming sessions revealed that emotional behavior of the robot led to more enjoyment during Battleship and was also more likely to trigger interactive behavior in participants than the unemotional version of the robot, which is obvious from the amount and phrasing of participants' verbal responses towards the robot.

However, based on our findings we can also identify some shortcomings of the emotional behavioral patterns of the robot developed for the experiment. The present study is a first approach to investigate effects of the behavior of a social robot on participants' mental models in a real-life situation. When designing the behavior for the robot we did not take into account that real-life social interaction processes might provide several other cues that could be relevant for the formation of people's mental models. Only in retrospective we discovered that, besides emotional expressiveness, the behavior displayed by the robot also offered a number of other cues such as its shape, perceptual skills or the ability to talk which might have influenced participants' implicit assessment of the robot's human-likeness. For future research on the effect of human-like behavior of a robot on people's mental representations of robotic actions it should therefore be kept in mind that additional cues can confound the intended manipulation and thus also distort the resulting action co-representation effects.

To sum up, the emotional behavior displayed by the robot during Battleship proved sufficient to manipulate participants' explicit perceptions about the robot and promoted basic interaction processes, whereas the influence of the robot's emotional-expressiveness on participants' implicit mental models of the robot's human-likeness remains questionable and requires further investigation.

REFERENCES

Atmaca, S., Sebanz, N., & Knoblich, G. (2011). The joint flanker effect: sharing tasks with real and imagined co-actors. *Experimental Brain Research*, *211*(3-4), 371–385.

Bartneck, C., & Forlizzi, J. (2004). A design-centred framework for social human-robot interaction. In *Proceedings of the Ro-Man2004* (pp. 591–594). Kurashiki.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81.

Beck, A., Hiolle, A., Mazel, A., & Cañamero, L. (2010). Interpretation of emotional body language displayed by robots. In *Proceedings of the 3rd international workshop on Affective interaction in natural environments* (pp. 37–42). ACM.

Blakemore, S.-J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*(8), 561–567.

Canamero, L. D., & Fredslund, J. (2000). How does it feel? emotional interaction with a humanoid lego robot. In *Proc. of American Association for Artificial Intelligence Fall Symposium*.

Cuijpers, R. H., Schie, H. T. van, Koppen, M., Erlhagen, W., & Bekkering, H. (2006). Goals and means in action observation: A computational approach. *Neural Networks*, *19*(3), 311–322.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113.

De Jong, R., Liang, C.-C., & Lauber, E. (1994). Conditional and unconditional automaticity: a dual-process model of effects of spatial stimulus-response correspondence. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(4), 731.

Dolk, T., Hommel, B., Colzato, L. S., Schütz-Bosbach, S., Prinz, W., & Liepelt, R. (2011). How “social” is the social Simon effect? *Frontiers in Psychology*, *2*, 84.

Dolk, T., Hommel, B., Prinz, W., & Liepelt, R. (2013). The (Not So) Social Simon Effect: A Referential Coding Account. *Journal of Exper- imental Psychology: Human Perception and Performance*.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593–609.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*(12), 493–501.

Goldman, A. I. (1992). In defense of the simulation theory. *Mind & Language*, *7*(1-2), 104–119.

Grezes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Human Brain Mapping*, *12*(1), 1–19.

Haring, M., Bee, N., & André, E. (2011). Creation and evaluation of emotion expression with body movement, sound and eye color for humanoid robots. In *RO-MAN, 2011 IEEE* (pp. 204–209).

Hedge, A., & Marsh, N. W. A. (1975). The effect of irrelevant spatial correspondences on two-choice response-time. *Acta Psychologica*, *39*(6), 427–439.

Hegel, F., Krach, S., Kircher, T., Wrede, B., & Sagerer, G. (2008). Theory of Mind (ToM) on robots: a functional neuroimaging study. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction (HRI)* (pp. 335–342).

IBM Corporation. (2013, March 12). IBM SPSS Statistics 22 Documentation. Retrieved from http://www-01.ibm.com/support/docview.wss?uid=swg27022445

Johnson, D. O., Cuijpers, R. H., Pollmann, K., & van de Ven, A. A. J. (2013). Exploring the Entertainment Value of Playing Games with a Humanoid Robot. *Human-Computer Interaction, under review.*

Johnson, D. O., Cuijpers, R., & van der Pol, D. (2013). Mimicking Human Emotions with the Nao Robot ' s Eye LEDs. *International Journal of Social Robotics, under review*.

Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological movement on action. *Current Biology*, *13*(6), 522–525.

Lee, S., Lau, I. Y., Kiesler, S., & Chiu, C.-Y. (2005). Human mental models of humanoid robots. In *Proceedings of the 2005 international conference on robotics and automation (ICRA 05)* (pp. 2767–2772). Barcelona, Spain: IEEE.

Liepelt, R., & Brass, M. (2010a). Top-down modulation of motor priming by belief about animacy. *Experimental psychology*, *57*(3), 221.

Liepelt, R., & Brass, M. (2010b). Top-down modulation of motor priming by belief about animacy. *Experimental psychology*, *57*(3), 221.

Liepelt, R., Cramon, D., & Brass, M. (2008). What is matched in direct matching? Intention attribution modulates motor priming. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 578.

Microsoft Corporation. (2010). Microsoft Excel 2010 Documentation. Retrieved from http://msdn.microsoft.com/de-de/library/office/ee658205(v=office.14).aspx

Müller, B. C. N., Brass, M., Kühn, S., Tsai, C.-C., Nieuwboer, W., Dijksterhuis, A., & van Baaren, R. B. (2011). When Pinocchio acts like a human, a wooden hand becomes embodied. Action co-representation for non-biological agents. *Neuropsychologia*, *49*(5), 1373–1377.

Paiva, A., Dias, J., Sobral, D., Aylett, R., Sobreperez, P., Woods, S., … Hall, L. (2004). Caring for agents and agents that care: Building empathic relations with synthetic agents. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1* (pp. 194–201). IEEE Computer Society.

Peirce, J. W. (2007). PsychoPy - psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13.

Pereira, A., Leite, I., Mascarenhas, S., Martinho, C., & Paiva, A. (2010). Using empathy to improve human-robot relationships. In *Proceedings of the 3rd International Conference on Human-Robot Personal Relationships.* Springer.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(04), 515–526.

Ramnani, N., & Miall, R. C. (2003). A system in the human brain for predicting the actions of others. *Nature neuroscience*, *7*(1), 85–90.

Scassellati, B. (2002). Theory of mind for a humanoid robot. *Autonomous Robots*, *12*(1), 13–24.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76.

Sebanz, N., Knoblich, G., & Prinz, W. (2005). How two share a task: corepresenting stimulus-response mappings. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1234.

Stenzel, A., Chinellato, E., Bou, M. A. T., del Pobil, Á. P., Lappe, M., & Liepelt, R. (2012). When humanoid robots become human-like interaction partners: Corepresentation of robotic actions. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(5), 1073–1077.

Stenzel, A., Chinellato, E., Del Pobil, A. P., Lappe, M., & Liepelt, R. (2013). How Deeply Do We Include Robotic Agents in the Self? *International Journal of Humanoid Robotics*, *10*(01).

Tsai, C.-C., & Brass, M. (2007). Does the human motor system simulate Pinocchio's actions? Coacting with a human hand versus a wooden hand in a dyadic interaction. *Psychological Science*, *18*(12), 1058–1062.

Tsai, C.-C., Kuo, W.-J., Hung, D. L., & Tzeng, O. J. L. (2008). Action co-representation is tuned to other humans. *Journal of Cognitive Neuroscience*, *20*(11), 2015–2024.

Whelan, R. (2010). Effective analysis of reaction time data. *The Psychological Record*, (58), 475–482.

Zhang, J., & Kornblum, S. (1997). Distributional analysis and De Jong, Liang, and Lauber's (1994) dual-process model of the Simon effect. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(5), 1543.

APPENDICES

APPENDIX A

Overview of the Emotional Expressions of the Robot used during the experiment

	game event	magnitude of emotion	gesture	eye LEDs	verbal expressions (examples): emotional content	+ reference to game event
robot hits	hits 1 & 2	happiness level 1	happy1	happy	Yes!	I thought there would be a ship around here.
robot hits	hits 3 & 4	happiness level 2	happy2	happy	Awesome!	I'm really good at this game.
robot hits	hit & ship sunk	happiness level 3	cheerful	happy	Excellent!	I sent the destroyer to the ground of the ocean.
robot hits	last hit & wins game	happiness level 4	dancing	happy	Hoooray! Totally Awesome!	I'm the winner.
participant misses	misses no. 1 & 2	happiness level 1	happy3	neutral	Hahaha!	You missed it!
participant misses	misses no. 3 & 4	happiness level 2	happy1	happy	Good!	You miss again.
participant misses	misses > 4	happiness level 3	happy2	happy	This is getting better and better for me.	Your shot missed again!
participant misses	miss directly after a hit	happiness level	happy2	happy	I have to disappoint you: Sorry,	no ship there!

	game event	magnitude of emotion	gesture	eye LEDs	verbal expressions (examples): emotional content +	verbal expressions (examples): reference to game event
robot misses	misses no. 1 & 2	sadness level 1	sad1	neutral	Bad luck!	-
robot misses	misses no. 3 & 4	sadness level 2	sad2	sad	Oh no!	I expected to hit one of your ships this time.
robot misses	misses > 4	anger	angry	angry	Not again!	This is really demotivation.
robot misses	miss directly after a hit	sadness level 2	sad2	sad	What a bummer!	I was so close to hitting it again.
participant hits	hit 1	sadness level 1	sad1	neutral	Oh no!	You just hit one of my ships!
participant hits	hits 2, 3, & 4	sadness level 2	sad2	sad	Are you kidding me? That's really unfair.	Another hit!
participant hits	sunk small ship	sadness level 3	sad3	sad	I can't believe it.	Hit again! You just destroyed my submarine.
participant hits	sunk big ship	anger	angry	angry	I'm getting quite an-noyed by this game.	Hit again! That was my battleship.
participant hits	last hit & wins game	sadness level 4	crying	sad	Oh no! How could that happen?.	That was the last ship! You win the game!
start of the game: robot welcomes participant		-	welcoming	happy	Hi there, I'm Marvin! I am glad you are here to play Battleship with me!	
robot thinks about his next guess		-	thinking	neutral	Let's think. Where could the other ships be?	

APPENDIX B

Extended Godspeed Questionnaire

Please rate your impression of the robot Marvin on these scales.
Geef aub uw indruk van de robot Marvin weer aan de hand van onderstaande schalen.

Mechanical *Mechanisch*	1	2	3	4	5	**Organic** *Organisch*
Artificial *Kunstmatig*	1	2	3	4	5	**Lifelike** *Levensecht*
Dislike *Afkeer*	1	2	3	4	5	**Like** *Geliefd*
Awful *Afschuwelijk*	1	2	3	4	5	**Nice** *Mooi*
Incompetent *Onbekwaam*	1	2	3	4	5	**Competent** *Bekwaam*
Unintelligent *Onintelligent*	1	2	3	4	5	**Intelligent** *Intelligent*
Apathetic *Apatisch*	1	2	3	4	5	**Responsive** *Responsief*
Machinelike *Lijkend op een machine*	1	2	3	4	5	**Humanlike** *Lijkend op een mens*
Unconscious *Onbewust*	1	2	3	4	5	**Conscious** *Heeft een bewustzijn*
Irresponsible *Onverantwoordelijk*	1	2	3	4	5	**Responsible** *Verantwoordelijk*
Unpleasant *Onplezierig*	1	2	3	4	5	**Pleasant** *Plezierig*
Unfriendly *Onvriendelijk*	1	2	3	4	5	**Friendly** *Vriendelijk*
Stagnant *Stilstaand*	1	2	3	4	5	**Lively** *Levendig*

APPENDIX C

Interpersonal Reactivity Index and Demographical Questionnaire

The following statements inquire about your thoughts and feelings in a variety of situations. For each item, indicate how well it describes you by choosing the appropriate score on the scale. **Please read each item carefully before responding and answer as honestly as you can.**

	Does not describe me well				Describes me very well
I daydream and fantasize, with some regularity, about things that might happen to me.	1	2	3	4	5
I often have tender, concerned feelings for people less fortunate than me.	1	2	3	4	5
I sometimes find it difficult to see things from the "other guy's" point of view.	1	2	3	4	5
Sometimes I don't feel very sorry for other people when they are having problems.	1	2	3	4	5
I really get involved with the feelings of the characters in a novel.	1	2	3	4	5
In emergency situations, I feel apprehensive and ill-at-ease.	1	2	3	4	5
I am usually objective when I watch a movie or play, and I don't often get completely caught up in it.	1	2	3	4	5
I try to look at everybody's side of a disagreement before I make a decision.	1	2	3	4	5
When I see someone being taken advantage of, I feel kind of protective towards them.	1	2	3	4	5
I sometimes feel helpless when I am in the middle of a very emotional situation.	1	2	3	4	5
I sometimes try to understand my friends better by imagining how things look from their perspective.	1	2	3	4	5
Becoming extremely involved in a good book or movie is somewhat rare for me.	1	2	3	4	5
When I see someone get hurt, I tend to remain calm.	1	2	3	4	5
Other people's misfortunes do not usually disturb me a great deal.	1	2	3	4	5
If I'm sure I'm right about something, I don't waste much time listening to other people's arguments.	1	2	3	4	5

	Does not describe me well				Describes me very well
After seeing a play or movie, I have felt as though I were one of the characters.	1	2	3	4	5
Being in a tense emotional situation scares me.	1	2	3	4	5
When I see someone being treated unfairly, I sometimes don't feel very much pity for them.	1	2	3	4	5
I am usually pretty effective in dealing with emergencies.	1	2	3	4	5
I am often quite touched by things that I see happen.	1	2	3	4	5
I believe that there are two sides to every question and try to look at them both.	1	2	3	4	5
I would describe myself as a pretty soft-hearted person.	1	2	3	4	5
When I watch a good movie, I can very easily put myself in the place of a leading character.	1	2	3	4	5
I tend to lose control during emergencies.	1	2	3	4	5
When I'm upset at someone, I usually try to "put myself in his shoes" for a while.	1	2	3	4	5
When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me.	1	2	3	4	5
When I see someone who badly needs help in an emergency, I go to pieces.	1	2	3	4	5
Before criticizing somebody, I try to imagine how I would feel if I were in their place.	1	2	3	4	5

Gender: ☐ Male ☐ Female

Age: _______

Have you ever participated in an experiment with this robot before?

☐ Yes ☐ No

Have you ever played Battleship before?

☐ Yes ☐ No