WHEN DO WE COOPERATE WITH ROBOTS?

Investigations in Human-Robot Interaction and Trust

Debora Zanatto

Faculty of Science and Engineering

University of Plymouth

This dissertation is submitted for the degree of

Doctor of Philosophy

June 2019

Nothing much could happen
Nothing we can’t shake
Oh we’re absolute beginners
With nothing much at stake
As long as you’re still smiling
There’s nothing more I need
I absolutely love you
But we’re absolute beginners
But if my love is your love
We’re certain to succeed.

Absolute Beginners

David Bowie

Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Graduate Sub-Committee.

Work submitted for this research degree at Plymouth University has not formed part of any other degree either at Plymouth University or at another establishment.

This study was financed with the aid of a studentship from the Air Force Office of Scientific Research, Air Force Materiel Command, USAF under Award No. FA9550-15-1-0025 and carried outn collaboration with the Centre for Robotics and Neural Systems (CRNS) at the University of Plymouth under the supervision of Prof. Dr. Angelo Cangelosi and Dr. Jeremy Goslin.

Relevant scientific seminars and conferences were attended at which work was often presented; external institutions were visited for consultation purposes and several papers prepared for publication.

Publications and Conference Proceedings:

– Zanatto, D., Patacchiola, M., Goslin, J., and Cangelosi, A. (2019). Generalisation of Anthropomorphic Stereotype *International Journal of Social Robotics*, pages 1-10. DOI:10.1007/s12369-019-00549-4.

– Torre, I., Goslin, J., White, L., and Zanatto, D. (2018). Trust in artificial voices: A congruency effect of first impressions and behavioural experience. In *Proceedings of the Technology, Mind, and Society*, page 40. ACM. DOI:10.1145/3183654.3183691.

– Zanatto, D., Patacchiola, M., Goslin, J., and Cangelosi, A. (2016). Priming Anthropomorphism: Can the credibility of humanlike robots be transferred to non-humanlike robots? In *Proceedings of the 11th ACM/IEEE Inter-*

national Conference on Human-Robot Interaction, pages 543-544. IEEE Press. DOI:10.1109/HRI.2016.7451847.

Conference Talks, Workshop attended and Seminars:

- The 11th ACM/IEEE International Conference on Human-Robot Interaction. Poster presentation *Priming Anthropomorphism: Can the credibility of humanlike robots be transferred to non-humanlike robots?*
- Hybrid Metaheuristics, 10th International Workshop, HM 2016, University of Plymouth, UK. 8-10 June, 2016.
- UK Robotics Week Centre for Robotics and Neural Systems (CRNS), University of Plymouth, UK. Talk and demonstrations. *Priming Anthropomorphism*. Best Presentation award. July 2016.
- Italian Institute of Technology (IIT) Genova (Italy). Seminar *Does Human-likeness always help trust in HRI?* 18th February, 2018.

Parts of this thesis are in preparation for submission by the author:

- Torre, I., Zanatto, D., Patacchiola, M., and Goslin, J. (2019). The effect of anthropomorphism and experience on trust attributions to a humanoid robot. *In submission*.
- Zanatto, D., Patacchiola, M., Goslin, J., and Cangelosi, A. (2019). Investigating cooperation with robotic peers. PLOS one. *In review*.
- Zanatto, D., Patacchiola, M., Thill, S., Goslin, J., and Cangelosi, A. (2019). Humans that imitate robots. Strategical social learning in Human-Robot Interaction. Cognition*Submitted*.
- Zanatto, D., Patacchiola, M., Thill, S., Cangelosi, A., and Goslin, J. (2019). Leader conformity and deviation in Human-Robot Interaction. *In submission*.

Word count for the main body of this thesis: **36877**

Signed: ______________________________

Date: ______________________________

Debora Zanatto
June 2019

Acknowledgements

I would like to thank my supervisors, Professor Angelo Cangelosi and Dr Jeremy Goslin for their expertise, encouragement, support and guidance.

So many people made an impact on these 4 years that it would be impossible to list them all. But I would like to single out a few whose contributions were especially significant.

Thanks to Sanjia for pushing me to do the interview for this position when I was losing any hopes in having a future in academia. I met you a few months before, but you changed my life more than many others, and I will never stop to thanking you for this.

Thanks to my office colleagues for standing me in my moody days, in particular to Leszek for assuring I was having healthy lunches, and to Massimiliano for teaching me how not to break a NAO.

Thanks to Alex, for being a true and sincere friend since the beginning. I know that, no matter how far, you'll always be there for me.

Thanks to Daniel, for being one of the nicest and sweetest persons I've ever met. You're one of the few capable of getting some uncynical sympathy out of me.

Thanks to Quentin, for showing me that we have to believe, always, that life is better than this, and that one day we'll find our way. Even if for little time, you've been my shoulder.

Thanks to Samuele, for proofreading this thesis, but most importantly, for being genuinely real, pro and cons. For teaching me that we are all different and all similar, but above all, for understanding that under this hard shell there is just a little girl.

Thanks to Riccardo, for the lovely meals we shared, but more importantly, for his genuine 'northern' friendship that reminds me where I come from.

Also, thanks to Guido, Mirco and Stefano, for 30 years of friendship. You, your partners and your kids are my roots and getting back home and see you warms my heart every single time.

Further, I would like to thanks my family. There are not enough words to describe how thankful I am for having taught me to fight and not be scared to be who I am. You truly are the strongest persons I've met in my life, and I am proud of being like you.

Finally, my greatest thanks go to Simone. Thanks for supporting me through all these 4 years, for putting me back on track every time I was feeling lost, and for all the patience shown in the most foolish days. You have been the greatest challenge of my life, but the final reward was worth all the effort and pain.

Abstract

WHEN DO WE COOPERATE WITH ROBOTS?
Debora Zanatto

Robotic usage is entering the world into many diverse ways, from advanced surgical areas to assistive technologies for disabled persons. Robots are increasingly designed and developed to assist humans with everyday tasks. However, they are still perceived as tools to be manipulated and controlled by humans, rather than complete and autonomous helpers. One of the main reasons can be addressed to the development of their capabilities to appear credible and trustworthy. This dissertation explores the challenge of interactions with social robots, investigating which specific situations and environments lead to an increase in trust and cooperation between humans and robots. After discussing the multifaceted concept of anthropomorphism and its key role on cooperation through literature, three open issues are faced: the lack of a clear definition of anthropomorphic contribution to robots acceptance, the lack of defined anthropomorphic boundaries that should not be crossed to maintain a satisfying interaction in HRI and the absence of a real cooperative interaction with a robotic peer. In Chapter 2, the first issue is addressed, demonstrating that robots credibility can be affected by experience and anthropomorphic stereotype activation. Chapter 3, 4, 5 and 6 are focussed in resolving the remaining two issues in parallel. By using the Economic Investment Game in four different studies, the emergence of human cooperative attitudes towards robots is demonstrated. Finally, the limits of anthropomorphism are investigated through comparisons of social human-like behaviours with machine-like static nature. Results show that the type of payoff can selectively affect trust and cooperation in HRI: in case of low payoff participants' increase their tendency to look for the robots anthropomorphic cues, while a condition of high payoff is more suitable for machine-like agents.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

> Sometimes the work was hard; the implements had been designed for human beings and not for animals, and it was a great drawback that no animal was able to use any tool that involved standing on his hind legs. But the pigs were so clever that they could think of a way round every difficulty.
>
> *Animal Farm*
>
> *George Orwell*

In this chapter, a literature overview on factors affecting cooperation in Human-Robot Interaction is given, although it is also extended in the individual following chapters.

1.1 How do we cooperate with humans?

Although not unique to our species, human cooperation has the scale and scope far beyond other mammals and forms a central underpinning of our psychology, culture, and success as a species (Boyd & Richerson, 2009). Most of our decisions depend on social interactions and therefore are based on concomitant choices of others (Sanfey, 2007). Consequently, we are constantly checking others' behaviour and producing or changing our response to it. When proposing a suggestion or a solution to a problem, for example, people look for others' approval, thus indicating that a final decision needs to be found together (Sebanz, Bekkering, & Knoblich, 2006). Oskamp (1971) proposed that there can be no positive social relationship or outcome from cooperation unless both parties adopt a cooperative attitude. Therefore, to cooperate effectively, it is necessary to arrive at an understanding of the others' behaviour such that equal and fair allocation of effort and resources leads to a joint solution beneficial to both parties. Cooperation is conditional upon the expectation of reciprocation (Perugini, Gallucci, Presaghi, & Ercolani, 2003). That is, cooperation is enhanced only when there are an understanding and willingness to establish a common ground, engendering a high level of trust in the partner.

Within previous literature, these types of cooperative interactions have been exhaustively examined, most prominently using the experimental paradigm of Behavioural Game Theory. In game theoretic social dilemmas, one player's strategy can be considered cooperative or competitive to the extent it affects collective rather than selfish interests (Deutsch, 1958). Within this domain, the relative coherence of strategies adopted by cooperative partners can provide a useful measure of conflict resolution or avoidance. Once partners establish a set of mutual beliefs regarding the current state of the task, the respective roles and the capabilities and responsibilities of each partner, then these behaviours tend to be maintained over time (Pilisuk, Skolnick, & Overstreet, 1968). Conversely, if partners are focussed upon maximising their own rewards, then their adopted strategies would have reduced coherence. Partners are more likely to adopt a shifting pattern of active and reactive strategies as they seek dominance over an adaptive opponent. In most established scenarios the adoption of a fixed mutual non-cooperative strategy is less generous than that of a cooperative strategy (Van Lange, Joireman, Parks, & Van Dijk, 2013). Cooperative strategies can be encouraged through the adoption of control mechanisms that punish non-coherent strategies or behaviours, such that cooperation is more profitable than non-cooperation (Gächter & Herrmann, 2009), with partners shown to use personally costly punishment mechanisms to increase future cooperation (Fehr & Gächter, 2002).

A recent study by Wu, Paeng, Linder, Valdesolo, and Boerkoel (2016) suggests that these types of cooperative interactions may also extend to human-robot partnerships, with cooperative behaviours in a trust game with a robot found to be very similar to players' behaviours with another human.

1.2 Do we cooperate with robots?

With the development of new robotic technologies, robots have the potential of being applied in many fields as reliable helpers. For example, the growing interest in the robotic application in industrial environments has allowed to improve and speed production and manufacturing processes. By involving robots in tasks like acquisition and manipulation, a reduction in human workload, costs and errors can be assured. Furthermore, the potential benefits deriving from HRI can be extended also to other fields, like agriculture, military missions, surgery and assistive care.

1.2.1 Human-Robot Cooperation in agriculture

Agricultural robots are autonomous or semi-autonomous systems that can help in solving problems at several stages of the process. Used mainly in repetitive tasks, these robots are designed to reduce farmers' workload and reduce production costs. Adamides (2016) built an interface that permits a human to teleoperate a pesticide-spraying robot. Freitas, Zhang, Hamner, Bergerman, and Kantor (2012) tested a localisation system for a vehicle used in tree fruit production. Bechar and Edan (2003) evaluated different collaboration levels in an HRI target recognition system for melon harvesting. Berenstein and Edan (2017) designed a robotic sprayer platform which decreased the spayed material by 50%.

Although there is evidence that HRC can have a positive impact on agriculture, harvesting and handling fruits are still challenging and complex problems. One of the main reason can be found on the high variability of the agricultural objects (i.e., colour, texture, shape), as well as the environmental conditions (e.g., changing illumination directions, shading, and targets occlusion).

1.2.2 Human-Robot Cooperation in military missions and rescue robotics

Robotics has been recently applied in military environments and search and rescue missions. These fields mainly focus on increasing the efficiency of military missions, as well as supporting victims of disasters or citizens in danger. The benefits of these applications derive, in particular, from increasing the speed and accuracy of the emergency response. For example, a semi-autonomous control architecture has been developed for robots in Urban search and rescue (USAR) environments, allowing them to make decisions about which tasks need to be carried out at a given time (Liu, 2019). Robots can also be programmed to autonomously fly and navigate through the environment or recognise fires and explosions (Bruemmer, Dudenhoeffer, & Marble, 2002).

Nevertheless, these applications are still on a simulative level, mostly due to the nature of the events involved. Data collection in dangerous situations, in fact, can be difficult or even impossible. For this reason, most of the developments and tests are still performed in robot test arenas (Jacoff, Messina, Weiss, Tadokoro, & Nakagawa, 2003).

1.2.3 Human-Robot Cooperation in surgery

Surgical procedures are becoming increasingly complex and require more refined skills. Minimal invasive techniques, for example, can limit the complications and reduce patients' recovery time. However, these techniques require extensive training. A solution to the problem could be found on robotics applications for surgery. Among these, robotic-assisted minimally invasive surgery (RMIS) can offer a functional alternative to manual surgery (Palep, 2009). For example, The Intuitive Surgical da Vinci®SP 1098 platform is composed of four interactive robotic arms for a great variety of laparoscopic procedures (Maurice, Ramirez, & Kaouk, 2017).

Unfortunately, most of the robotic platforms are still under development, mainly due to the numerous challenges that robotics in surgery still have to face, in particular, the limited perception abilities of the robots in surgical sites and the large number of DOFs needed in such systems (for a review see Brodie and Vasdev, 2018).

1.2.4 Socially assistive robots

With the coming of increasing ageing population and the consequential serious problems to our society, the interest HRI has moved from a mere instrumental to an assistive and social role. Robots could offer support in various tasks in our daily life, in welfare applications and office automation.

In particular, there is increasing interest in using socially assistive robots (SAR) as caretakers for the elderly and people with disabilities (Martinez-Martin & del Pobil, 2018). Paro –a baby seal robot– and NeCoRo –a cat-like robot– are designed to improve elderly quality of life by reducing distress and facilitate daily living activities (Kidd, Taggart, & Turkle, 2006; Libin & Cohen-Mansfield, 2004; McGlynn, Snook, Kemple, Mitzner, & Rogers, 2014; Nakashima, Fukutome, & Ishii, 2010; Sabanovic, Bennett, Chang, & Huber, 2013). Aido and BUDDY can provide companionship and assistance in daily activities like reminding medications, appointments and upcoming events (Gleaton, Shirley, & Carolyn, 2018). Mobile robots like HOBBIT, which is capable of taking commands from the user and recognise gestures, can also be used as service devices in assisting mobility and emergency recognition (Bajones et al., 2018; Fischinger et al., 2016).

Furthermore, robots can play a significant role in education (Pandey & Gelin, 2016). The humanoid Robovie, capable of human-like expressions, can recognise individuals and encourage interactive learning by increasing children engagement in social activities (Kanda, Hirano, Eaton, & Ishiguro, 2004).

Additionally, social robots seem to be effective in treating patients with Autistic Disorders (Thill, Pop, Belpaeme, Ziemke, & Vanderborght, 2012). The NAO robot is one of the main players in this scenario, which has been shown to improve the social skills of children affected by Autism Spectrum Disorder (Brockevelt, Manner, Richter,

Fig. 1.1 A team of social robots.

& Elison, 2018; Lytridis, Vrochidou, Chatzistamatis, & Kaburlasos, 2018; Tapus et al., 2012).

The great potential that robots could have in improving human quality of life has aroused the need for establishing mutual cooperation between humans and robots (Fong, Nourbakhsh, & Dautenhahn, 2003). Research has mainly focused on the subordinate and assistive role of the robot (e.g. Lenz et al., 2008; Huang, Cakmak, and Mutlu, 2015; Goetz and Kiesler, 2002; Severinson-Eklundh, Green, and Hüttenrauch, 2003; Hinds, Roberts, and Jones, 2004), while cooperation with a proper robotic peer has not been fully investigated. Moreover, most of the experiments conducted so far were based on self-reports, questionnaires and interviews (e.g. Nomura, Kanda, Suzuki, and Kato, 2008; Yagoda and Gillan, 2012; Charalambous, Fletcher, and Webb, 2016). Although these measurements allow collecting people's overt opinions, they exclude the possibility of measuring spontaneous and automatic behaviours. These type of measurements could permit to explore the implicit processes responsible for the establishment of functional interaction with robots.

1.3 It's all a matter of acceptance

As robots would become more present in human environments, it is increasingly important that they are capable of interacting socially. In human and animal societies, in fact, collaboration is an essential skill critical to learning, working and everyday social interactions. As social skills in humans are necessary for collaboration, robots also need to be equipped with these skills to be a productive and valuable collaborator in a wide variety of contexts. In this regards, social robots can change the conception of technology by promoting pro-social and collaborative actions and increasing team cohesiveness.

HRC focuses on applications where robots and humans work together to complete a task. This field faces particular challenges related to people's expectations and mental models about their team partners that can be not satisfied. In particular, the robots capabilities and skills perception become the main problem to solve; robots are still perceived as untrustworthy machines, intruding people's life and environment, thus cooperation is strongly affected (P. Hancock, Billings, & Schaefer, 2011). Consequently, one of the pivotal questions in social robotics is how to design a robot and which features (e.g., appearance, movements, voice, behaviours) can improve its interaction with humans. What makes us accept a robot and willing to work with it?

1.3.1 Theories and factors affecting robot acceptance

Despite the large development and interest in robotics, a theoretical model specific to robot acceptance is still missing (Beer, Prakash, Mitzner, & Rogers, 2011; Heerink, Krose, Evers, & Wielinga, 2009) (for a review of technology acceptance theories see Table 1.1).

The Technology Acceptance Model (TAM, Davis, Bagozzi, and Warshaw, 1989) proposes two main variables that affect acceptance: perceived usefulness (the feeling that using a particular system will improve the performance) and perceived ease of use (the perception of effort reduction in using the system). This model, however, has received several criticisms due to its restrictive vision of technology acceptance and the exclusion of several social variables that could affect people's interaction with technology (Bagozzi, 2007; Taylor & Todd, 1995).

The Unified Theory of Acceptance and Use of Technology Model (UTAUT, Venkatesh, Morris, Davis, and Davis, 2003) unified components from eight different technology acceptance models and defines acceptance on the basis of four main constructs: performance expectancy (expectations toward the system efficiency), effort expectancy (perceived ease of use), social influence (evaluation on the importance of using the system by others) and facilitating conditions (believes about the organisations supporting the use of the system). Although it has been considered the more robust technology acceptance model, criticisms focus on the number of predictive variables kept into account (Bagozzi, 2007).

The existence of numerous models for acceptance in HRI has led to the most diverse interpretations and investigations. De Graaf and Allouch (2013) identified variables like usefulness, adaptability, enjoyability, sociability, companionship and perceived behavioural control. Beer et al. (2011) proposed three specific variables as

Table 1.1 Technology acceptance theories

Technology Acceptance Theories	Components
Theory of Reasoned Action (TRA; Fishbein and Azjen, 1975)	1 Attitudes 2 Social Norms 3 Intentions
Technology Acceptance Model (TAM; Davis, Bagozzi and Warshaw,1989)	1 Perceived usefulness 2 Perceived ease of use
Theory of Planned Behavior (TPB; Ajzen, 1991)	1 Perceived Behavioural Control 2 Subjective Norm 3 Behavioural Attitude
Motivational Model (MM; Davis and Bagozzi, 1992)	1 Extrinsic Motivation 2 Intrinsic Motivation
Social Cognitive Theory (SCT; Compeau and Higgins, 1995)	1 Outcome Expectations – Performance 2 Outcome Expectations – Personal 3 Self-efficacy 4 Affect 5 Anxiety
Innovation Diffusion Theory (IDT; Rogers, 1995)	1 Relative Advantage 2 Ease of Use 3 Image 4 Visibility 5 Compatibility 6 Results Demonstrability 7 Voluntariness of Use
Unified Theory of Acceptance and Use of Technology (UTAUT; Venkatesh, Morris, Davis, and Davis, 2003)	1 Performance Expectancy 2 Effort Expectancy 3 Social Influence 4 Facilitating Conditions

potentially affecting robot acceptance: function, social capabilities and appearance. They also assumed that a discrepancy between the robot functionalities or social skills and the user's expectations would reduce its acceptance. On the other hand, the robot human-likeness level would have the same effect on user perception and judgments. Lastly, Gaudiello, Zibetti, Lefort, Chetouani, and Ivaldi (2016) identified six correlates of robot acceptance: representational, physical, behavioural, functional, social, and cultural. Among these dimensions, they underlined the role of functional and social acceptance as being particularly effective on HRI.

Among all these different theorisations and interpretations, three key factors in affecting robot acceptance have been found on performance capability, social competence and appearance. These main factors share a common and fundamental indicator of robot acceptance, which is the development of a sense of trust on the robot (Lewis, Sycara, & Walker, 2018; van den Brule, Dotsch, Bijlstra, Wigboldus, & Haselager, 2014) and has been increasingly used to measure the quality of HRI (Gaudiello et al., 2016; Schaefer, 2013).

1.4 It takes trust to accept a robot

Mayer, Davis, and Schoorman (1995) defined trust as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party". As the authors stated, trust involves interdependence, so that a person (called 'trustor') depends on another (called 'trustee') to accomplish his/her task. According to their view, the characteristics of the trustor will affect the trustor behaviour. These characteristics include:

- ability: skills and competencies that make the trustor believe that the trustee has the resources and the expertise to successfully complete the task;
- benevolence: the belief that the trustee wants to improve the trustor condition, thus the perception that the trustee is oriented toward the trustor;
- integrity: the expectation that the trustee will follow some principles that the trustor finds fundamental to the interaction.

Based on these three components, the level of trust can vary on a continuum that goes from low to high: if the trustor perceives high levels of ability, benevolence and integrity, the trustee would be categorised as trustworthy; on the opposite, lower levels would lead to lower trust.

The main goal of social HRI is to establish strong relationships in which people find robots suggestions credible, acceptable and usable. One of the necessary elements for building a social human-robot relationship is trust (P. Hancock et al., 2011; Kidd & Breazeal, 2005; Ullman & Malle, 2018). As Salem and Dautenhahn (2015) stated: "Especially with regard to critical decisions, trust plays an important role in human interactions and could therefore help to increase the robot's acceptance in its role as a collaborative partner (Hinds et al., 2004; J. J. Lee, Knox, Wormwood, Breazeal, & DeSteno, 2013). Since trust is strongly linked to persuasiveness in social interaction contexts (Touré-Tillery & McGill, 2015), it could also affect people's willingness to cooperate with the robot (Freedy, DeVisser, Weltman, & Coeyman, 2007), for example, by accepting information or following its suggestions". The less people trust a robot, in fact, the earlier they intervene during a task completion (De Visser, Parasuraman, Freedy, Freedy, & Weltman, 2006; Steinfeld et al., 2006).

Although a reliable robotic trustee could not embed all the features listed by Mayer et al. (1995), it still shares some common characteristics that are also related to robots acceptance. Starting from the previously listed factors affecting acceptance, a robotic trustee would be expected to:

- have the skills and expertise to positively accomplish the task (performance capability);
- be oriented toward the human trustor and follow his/her beliefs (social competence).

1.4.1 Factors affecting trust in HRI

In a meta-review on the factors affecting trust in HRI, P. A. Hancock et al. (2011) mentioned three main sources: human-related, environmental and robot-related factors. Human-related factors include personality traits, attitudes toward robots and engagement capabilities. In their review, the authors found little evidence for the effect of this source. Environmental factors resulted in having a moderate effect, depending in particular on the type of task, in-group membership and communication. Interestingly, robot-related factors were listed as the most important contributors and have been subsequently categorised as performance-based and attribute-based factors (Oleson, Billings, Kocsis, Chen, & Hancock, 2011).

Performance-based characteristics embed all the features that can affect performance, ability and functionality perception. Reduced reliability and the presence of numerous errors, for example, can decrease trust in a robotic system (Desai et al., 2012; Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Salem, Lakatos, Amirabdollahian, & Dautenhahn, 2015). On the other hand, if a robot shows consistency in performance, trust can be maintained. Moreover, robots that are functionally sophisticated are trusted more than robots with social abilities (Gaudiello et al., 2016).

Attribute-based characteristics comprise all the components related to the design and the socio-physical features of the robot. Some of these features include anthropomorphic appearance, closeness, predictability and social competence. Shiomi, Nakata, Kanbara, and Hagita (2017) reported higher preferences for a robot capable of hugging people, as well as an increase of motivation when a robot is capable of giving recommendations (Shiomi et al., 2010). Lohani, Stokes, McCoy, Bailey, and Rivers (2016) showed that matching social and emotional behaviours affects trust in a collaborative task. Martelaro, Nneji, Ju, and Hinds (2016) reported that their participants developed trust and feelings of companionship with a vulnerable robot. Furthermore, several studies indicated an increase in trust, motivation and cooperation with a highly predictable robot (Billings et al., 2012; Dragan, Bauman, Forlizzi, & Srinivasa, 2015; J. D. Lee & See, 2004). Interestingly, the mere presence of a robot compared to a virtual agent and the reduction of the distance between the human and the robot create positive feelings and increase trust (Bainbridge, Hart, Kim, & Scassellati, 2008; MacArthur, Stowers, & Hancock, 2017).

All the present findings suggest that humans' perception of the robot performance and its social competences are key factors in developing and maintaining trust in HRI. However, a precise description of the social characteristic that a robot should embed is missing and far from clear. In this regards, a contribution capable of clarifying the puzzling picture or social HRI is impelling.

Finally, among the attribute-based characteristics affecting trust in HRI, anthropomorphism has been rated as one of the greatest contributors (Duffy, 2003). Due to its multifaceted properties, anthropomorphism is treated in a separate paragraph.

1.5 Anthropomorphism in HRI

The tendency to anthropomorphise is fundamental to our psychology, a cross-cultural phenomenon with a history established in pre-historic art (Conard, 2003) and coined in ancient Greek philosophy (Xenophanes, as cited by Lesher, 1992). It has been proposed (Epley, Waytz, & Cacioppo, 2007) that a process of induction lies at the core of this highly prevalent phenomenon, one in which we use highly accessible knowledge on the self or human society as a base upon which we can explain or predict the unknown. Ascribing human intentions to unintentional agents, described as the *intentional stance* by Dennett (1989), eases our uncertainty on how we can reason about or describe these agents. This stance would also make it easier to form social connections with non-human agents, extending our fundamental desire for social relationships (Baumeister & Leary, 1995) outside of humankind. Thus, anthropomorphism would appear to offer us a simpler, more familiar world that is richer in social interaction, giving us greater confidence in our interactions with it.

Historically, research in this area has been dominated by studies on the accuracy and functionality of our anthropomorphic beliefs (Cheney & Seyfarth, 1992; Hauser, 2001), as well as their potential benefits (J. S. Kennedy, 1992; Nass, Isbister, & Lee, 2000; Tam, Lee, & Chao, 2013). Another major area of research has been the examination of the factors that influence the extent to which people anthropomorphise non-human agents. Guthrie and Guthrie (1993) noted the attribution of intentions and human-like mental states to literary characters were related to their morphological similarity to ourselves. This influence of morphology seems relatively straightforward; objects featuring human features (Baron-Cohen, 1997; Carey & Spelke, 1994; Dennett, 1989) or shape (Arnheim, 1969) are seen as more human-like than those that do not. Shape, in particular, is used as one of the earliest cues to object animacy in infancy (Graham & Poulin-Dubois, 1999). Additional features such as imitation, communication or indicators of intentionality can also influence the extent to which we anthropomorphise an agent (Dennet, 1997). However, similarity to human motion is probably the next most important cue after that of morphology (R. W. Mitchell & Hamm, 1997). Movement can give the impression of life (Tremoulet & Feldman, 2000) and even simple geometric figures can be anthropomorphised as long as the temporal relationship of their movements conforms to our own (Heider & Simmel, 1944). Previously mindless agents whose movements are slowed down or speeded up (Morewedge, Preston, & Wegner, 2007) to approximate human motion can gain mental states, while biologically meaningful motion can also lend intention to non-human agents (Dittrich & Lea, 1994).

Humans have the natural tendency to make anthropomorphic assumptions about the surrounding environment, animals, weathering, planets, or even geometrical figures (Burghardt, 2017; Eddy, Gallup Jr, & Povinelli, 1993; Guthrie & Guthrie, 1993; Milstein, 2011). Consequently, anthropomorphism has been increasingly used to sell products (Portal, Abratt, & Bendixen, 2018; Tuškej & Podnar, 2018) and design technology, particularly robots (Duffy, 2003; Duffy & Joue, 2004). On the attempt to explain the anthropomorphisation of non-human agents, Epley et al. (2007) proposed a Three-Factor Theory, in which seeing human in non-human depends on:

- Elicited agent knowledge: since people's knowledge about non-human agents is less wide compared to humans, applying an anthropomorphic interpretation to non-agents becomes a simpler and costless solution;
- Effectance motivation: people have an innate tendency and curiosity for understanding, control, and predict, which in turn, increase the tendency to anthropomorphise;

- Sociality motivation: the need for social connections and the tendency to avoid loneliness force people to treat non-human agents like human companions.

Following Epley's theory, studies have found that as a robot appears more human-like, people are more likely to appreciate it and work with it (Goetz, Kiesler, & Powers, 2003), follow its instructions, (Kiesler, Powers, Fussell, & Torrey, 2008) and even empathise with it(Riek, Rabinowitch, Chakrabarti, & Robinson, 2009a). Providing an anthropomorphic form to a robot might not be sufficient to facilitate people's interaction with it (Kahn Jr et al., 2007).

Research on verbal communication has focused on the role of accents and natural voices in increasing acceptance (Eyssel, De Ruiter, Kuchenbrandt, Bobinger, & Hegel, 2012; Markowitz, 2017; Tamagawa, Watson, Kuo, MacDonald, & Broadbent, 2011). Investigations on non-verbal communication (Breazeal, Kidd, Thomaz, Hoffman, & Berlin, 2005) instead, have focused on gestures and gaze, showing that gaze shifting during social interactions makes robots more enjoyable (Admoni & Scassellati, 2017). Moreover, robots performing joint attention are rated as more competent (Huang & Thomaz, 2010).

Even making mistakes is reported to increase the perception of the robot human-likeness (Mirnig et al., 2017). Ragni, Rudenko, Kuhnert, and Arras (2016) compares an erroneous robot with a perfect one during a reasoning task. The authors found that the erroneous robot triggered more positive emotions. Robots that exhibit typical human behaviours, like cheating, are also perceived as more human-like (Short, Hart, Vu, & Scassellati, 2010). In their study, Short and colleagues compared three robot behaviours during a "rock-paper-scissors" game between a robot and a human: a control condition with a normally behaving robot, a verbal-cheat condition with a robot lying about winning the game and an action-cheat condition in which the robot additionally changed its gesture to the winning gesture. The verbal-cheat was interpreted as a mechanical malfunctioning of the robot. On the opposite, the action-cheat condition was perceived as an intentional behaviour of the robot.

Anthropomorphism also changes with experience, as previous research suggested (Fussell, Kiesler, Setlock, & Yew, 2008; Lemaignan, Fink, & Dillenbourg, 2014). In Fussell et al. (2008), participants rated a robot as possessing traits, moods and feelings more after they interacted with it than after they simply imagined one. Thus, interacting with a robot increases its anthropomorphism. In outlining a formal model of anthropomorphism, Lemaignan et al. (2014) added that the interaction context also influences the dynamics of anthropomorphism. Although humans can build an 'anthropomorphic model' of the robot during the interaction, the presence of unpredicted behaviours may lead to an increase of anthropomorphism.

1.5.1 Is it all anthropomorphism fault?

It could be concluded that anthropomorphism is a generalizable instrument that always facilitates interaction with robots. Nevertheless, it appears to be unnecessary or even counterproductive under certain conditions (Goudey & Bonnin, 2016). Too much human-likeness, in fact, can result in the negative Uncanny Valley effect (Masahiro, 1970), evoking feelings of disgust and eeriness. Androids, for example, have a striking resemblance with human appearance (Minato, Shimada, Ishiguro, & Itakura, 2004) but as soon as they start to move or speak, their machine qualities are immediately evident (Becker-Asano, Ogawa, & Nishio, 2010). Goudey and Bonnin (2016) suggested that anthropomorphism is associated with the levels of familiarity and practical experience with technology. Hence, people with less technological experience, or reduced exposure to robots, are less keen to accept a human-like agent over a machine type of robot.

Anthropomorphism seems to be also task-related. For example, people feel less embarrassed in interacting with a non anthropomorphic robot during medical check-ups (Bartneck, Bleeker, Bun, Fens, & Riet, 2010). Bethel and Murphy (2010) instead demonstrated that a non-anthropomorphic robot is rated more calming in Search and Rescue applications.

These studies indicate that anthropomorphism in HRI is not limited to the physical world. Perceiving robots as agents possessing a mind and the ability to execute intentional actions is equally affecting human's perception of the agent. Moreover, anthropomorphism is also a context-related instrument, that can endanger the interaction whenever used in the wrong way. Recent research demonstrates the need to define a more precise threshold of anthropomorphism. Moreover, these studies bring to light the need of investigating under which circumstances human-like features are positively affecting HRI and when, instead, a robot should be just a robot.

1.6 Aims, Scope and Research Questions

Cooperation with robots is a fundamental aspect of HRI. Increasing people's willingness to interact and cooperate with robots, would potentially increase their usage and guarantee a satisfying and high-quality performance. Nevertheless, cooperation outside of the industrial field has not been studied in depth, mostly because of the many issues HRI research has faced in terms of trust and acceptance. In order to improve cooperation with robots, it is fundamental to explore the boundaries of acceptance and trust in HRI. Although the studies that have investigated these two aspects have reported a wide list of elements, an objective guideline is missing.

This thesis focuses on investigating trust and cooperation in HRI in different social scenarios and the role of anthropomorphism in affecting this cooperation. Specifically, it explores how we extend the use of social collaborative cues to robots (both single, in groups, and the role of leadership), and what are the necessary requirements for that extension. This includes aspects of form, behaviour, gaze, and speech. For all these reasons, this thesis will cover the following questions:

- Can anthropomorphic stereotype activation affect robots credibility?
- How does trust in HRI evolve over time and under what specific circumstances anthropomorphism increases trust in HRI?
- Can we cooperate with robots as we cooperate with other humans? What is the role of anthropomorphism in affecting cooperation?
- Can humans choose to strategically imitate a robot and how would anthropomorphism influence social learning in HRI?
- Would humans conform to a robotic leader or would they prefer to follow a dissenting minority? Which features should a robotic leader embed to be trusted?

The following chapters attempt to answer these questions. In Chapter 2, the anthropomorphic level of robots is manipulated in order to establish the limit of acceptance. More specifically, participants are hypothesised to accept a robot suggestion through anthropomorphic stereotype generalisation. Throughout two experiments, participants' perceptions and judgements of machine-like robots are studied before and after being exposed to a human-like robot. The opposite scenario is also investigated in search of a dehumanisation effect for anthropomorphic robots. Results demonstrate that humans are capable of transferring the anthropomorphic stereotype from a human-like robot to a machine-like one, thus opening to the possibility that people could learn to accept a robot from experience and stereotype activation.

In Chapter 3, trust between humans and robots is investigated by manipulating the robot anthropomorphic features of voice and joint attention, as well as its reliability. The main core of this study is to establish under which conditions, both in terms of robot features and performance, anthropomorphism is a required element to build and increase trust in robots. In this chapter, the development of trust over time is investigated by using the same experimental method that is subsequently applied to all further studies. By using the Investment Game, participants' willingness to invest (and thus trust) in a more or less anthropomorphic robot is explored. Results from this experiment demonstrate that trust can evolve over time. More importantly, results show that anthropomorphism affects HRI in a selective way. While under specific circumstances (low payoff condition) it can increase the quality of the interaction, in others it can have a negative impact (like in case of high payoff).

Chapter 4 explores the emergence of cooperative attitudes in humans when interacting with a robot as a proper peer. Participants have to decide whether to cooperate with a more or less strategically selfish robot, in investing in another robotic agent. This experiment demonstrates that humans are capable of treating a robot as an 'au pair' companion and apply human social cooperative behaviours to it. Moreover, the selectiveness of anthropomorphism is shown to be consistent with the results in Chapter 3.

In Chapter 5, the cooperative scenario is extended to social learning. Here participants are hypothesised to strategically imitate a robot cooperative behaviour toward another robot. By exposing participants to a robotic model, it is shown that humans are disposed to copy a robot strategy whenever it is considered successful and reliable. Furthermore, anthropomorphism is also manipulated to increase the imitative tendency.

Stemming from the evidence that a robot can be a trustworthy source of imitation, in Chapter 6 the dyadic human-robot cooperation is extended to group cooperation and the role of robotic leadership in HRI is investigated. Specifically, the chapter is focused on the human's tendency to cooperate with a group of robots and to conform to the leader. Participants are asked to decide whether to conform or dissent from investing in a group of robots. The manipulation of the followers' consensus to the robotic leader, shows that participants conform to a leader that is receiving the overall consensus. On the opposite, they show dissenting behaviours when the leader loses the majority support. Lastly, two main leader features, namely height and prototypicality, are manipulated through two experiments to test whether these features can affect leader conformity. Results demonstrate that the physical attributes of a robotic leader increase conformity only when participants are exposed to a dissenting minority.

1.7 Contributions

This thesis is part of the collaborative THRIVE project, funded by the Air Force Office of Scientific Research (AFOSR). The aim of the project is to investigate the embodiment and socio-cognitive mechanisms in the development of trust between humans and robots through the method of developmental robotics. This is inspired by psychological theories of the developmental emergence of sociocultural skills and through empirical human-robot interaction experiments. This aim is achieved through the following research objectives:

- define a developmental robotics framework for the understanding and prediction of trust interaction between humans and robots;
- conduct HRI experiments validating the role of joint attention, joint action, imitation and group assimilation in the development of trust, coupled with embodiment and non-verbal cues;
- design a cognitive architecture for autonomous robots to support and sustain the building of trust interaction between humans and robots.

This thesis is also part of the collaborative CogNovo Project, funded by the EU Marie Skłodowska Curie initiative and the University of Plymouth, to foster research training in the emerging field of Cognitive Innovation. In particular, the study reported in Chapter 3 was conducted in collaboration with Dr Ilaria Torre, a Cognovo Research Fellow working on voice features for engagement and trust.

For all the experiments here presented, the codes for the behaviours of the robots have been produced by Massimiliano Patacchiola as part of the THRIVE Project. All the codes used can be found in the Github repository `https://github.com/mpatacchiola/naogui/tree/master/tzpgc_2016`.

The experiments in Chapter 2, 4, 5 and 6 have been planned, organised, conducted and supervised by Debora Zanatto as part of the THRIVE Project. The experiment in Chapter 3 has been jointly planned and conducted by Debora Zanatto and Ilaria Torre as part of the Cognovo Project (further details are reported in Chapter 3).

Chapter 2

Anthropomorphic stereotype activation

What, then, is truth? A mobile army of metaphors, metonyms, and anthropomorphisms – in short, a sum of human relations which have been enhanced, transposed, and embellished poetically and rhetorically, and which after long use seem firm, canonical, and obligatory to a people: truths are illusions about which one has forgotten that this is what they are; metaphors which are worn out and without sensuous power; coins which have lost their pictures and now matter only as metal, no longer as coins.

On Truth And Lie in the Extra-Moral Sense.

Friedrich Nietzsche

2.1 Can the credibility of human-like robots be transferred to non-human-like robots?

The following studies provide valuable insight into the range of factors that lead us to accept a particular nonhuman agent. In particular, the specificity of the anthropomorphic assignment is examined. If we apply the anthropomorphic status to a particular agent, do we also extend this human-like status to other exemplars of the same category of agents? Moreover, if there are variations in the distribution of anthropomorphic features across the category of agents, might this effect also lead us to extend human-like status to variants that may lack those features? As an example, the moon has visual features that approximate a human face which might lead us to anthropomorphise this particular astronomical body. Given that the moon is likely to be the first body of this type we would encounter, we may also generalise the anthropomorphic trait across hereto unseen astronomical bodies? If so, would this initial encounter increase the probability that we would also assign a human-like status to encounters with other exemplars of the astronomical body even if, like Mars, they are lacking the requisite anthropomorphic features? Conversely, if the first astronomic body we were to view was Mars, would we then be less likely to anthropomorphise the moon, as we had already categorised astronomical bodies as being non-human-like?

These potential effects would appear to be related to the concept of stereotype activation. Stereotypes allow to categorise and simplify the environment (Leyens, Yzerbyt, & Schadron, 1994) and to associate traits and behaviours to individuals belonging to a social category (Biernat, Manis, & Nelson, 1991). Stereotype activation is a useful scaffold when the amount of information is not adequate to form an impression (Wichman, 2012). People rely on their prior knowledge of a category to generalise and extend that stereotype to similar exemplars. This stereotypical generalisation, like the tendency to anthropomorphise non-human agents, derives from the innate need of filling the gap between the lack of information about the agent and the need to portray it and socially define it. Although little is known about the effect of stereotype activation on non-group members, recent research has shown that repeated exposure to stereotypes results in greater confidence on those stereotypes when making evaluations. Arendt, Steindl, and Vitouch (2015) found evidence of higher facial threat after reading criminal prototypical stories about dark-skinned strangers. Moreover, Kim-Prieto, Goldstein, Okazaki, and Kirschner (2010) demonstrated that exposure to American Indians icons increases people's willingness to endorse stereotypes about a different racial minority group, like Asian Americans.

It could be considered that a human-like robot is the epitome of anthropomorphism, an explicitly constructed instantiation of anthropomorphic form, behaviour and function designed to enhance our interaction with and understanding of a highly obtuse technological agent. Studies have shown that robots can be highly effective in solicitors of anthropomorphic projection (for a review see Złotowski, Proudfoot, Yogeeswaran, and Bartneck, 2015). We are inclined to project our own social schemas onto robots (Fussell et al., 2008), especially when they have human faces or bodies (DiSalvo, Gemperle, Forlizzi, & Kiesler, 2002) or are engaged in social roles (Fong et al., 2003). Thus, if a robot has human-like features it is perceived as anthropomorphic, it is expected to behave appropriately (Salem, Rohlfing, Kopp, & Joublin, 2011) and follow our social norms (Syrdal, Dautenhahn, Koay, Walters, & Ho, 2013). Anthropomorphic robots are also perceived as being more intelligent (Bartneck, Kulić, Croft, & Zoghbi, 2008; Krach et al., 2008), trustworthy (P. A. Hancock et al., 2011; Waytz, Heafner, & Epley, 2014), likeable (Nass et al., 2000) and attract more visual attention (Bae & Kim, 2011) and empathy (Riek, Rabinowitch, Chakrabarti, & Robinson, 2009b) than their less human-like variants. We even assign personality traits to the robot based upon idiosyncratic features of their height (Walters, Koay, Syrdal, Dautenhahn, & Te Boekhorst, 2009), gender features (Powers et al., 2005), face (Yee, Bailenson, & Rickertsen, 2007) or voice (Nass & Brave, 2005).

The social consequences of anthropomorphism have been widely studied in marketing by designing human-like features into a product or linguistic inference: cars are evaluated more highly if they have human facial similarities (Aggarwal & McGill, 2007), and people gamble more when slot machines are described in anthropomorphic terms (Kim & McGill, 2011; Riva, Sacchi, & Brambilla, 2015).

If the anthropomorphic status is extended across members of an agent category then these studies suggest that the status assigned to the first exemplar of the category should increase the probability that the same status is assigned to subsequently encountered exemplars. Therefore, if the first exemplar of a new type of agent is assigned human-like status, then this should prime anthropomorphism in a future exemplar of similar type and vice versa. The following studies, test this hypothesis by measuring the behavioural effects of anthropomorphism during interactions with an agent that has either been primed with a less or more anthropomorphic initial exemplar. Non-human agents are instantiated by two types of robot, each conforming to the same broad category of an agent, but with exemplars provoking different degrees of anthropomorphism based upon their form and behaviour.

2.1.1 Pilot study

In this first study, interactions between the anthropomorphic iCub and less-human-like Scitos G5 during a price judgement game used by Rau, Li, and Li (2009), have been compared (Figure 2.1). Participants were asked to accept price valuations on common objects provided by the robots. Their willingness to change their own price judgments to that provided by the robot was taken as the primary measure of the robot credibility. In the first experiment participants were only asked to interact with the less-anthropomorphic Scitos G5 robot, to be referred to as the "nonprimed Scitos G5". In the second experiment, participants first interacted with the anthropomorphic iCub robot, before repeating the task with the Scitos G5, to be referred to as the "primed Scitos G5". In addition, to increase the scope of anthropomorphism beyond the physical form of the robot, a behavioural condition was included in the experiment, in which robots could engage with participants during the game with either a social or fixed gaze. The credibility was expected to be higher for the anthropomorphic than the less anthropomorphic robot. Moreover, interactions using a social gaze should be more credible than those with a nonsocial fixed gaze. However, if we only apply our normal social stereotypes to anthropomorphic robots, then it might be surmised that the benefits of social gaze would only be seen in human-like robots. Finally, the credibility of the less-anthropomorphic robot should be significantly greater when interactions with this robot are preceded by interactions with an anthropomorphic robot.

2.1.2 Method

The participants were asked to select from one of two valuations placed on common objects presented by a robotic compatriot. After a selection was made, the robot would voice its agreement or disagreement with the selected price and, if the participants did not agree with it, give them the chance to change their decision so that it would conform to its own. The willingness of participants to change their judgement was taken as a measure of the robot credibility. This task was used to examine how the

Fig. 2.1 iCub (left) and Scitos G5 (right) robots

form and behaviour of the robot would affect their perceived credibility, and whether their initial experiences with a particular robot would prime their future interactions with a different type of robot. In the first experiment participants interacted only with Scitos G5. In the second experiment, each participant made a separate series of price judgement interactions first with iCub and then immediately repeated the game with Scitos G5. These robots could also be programmed to behave with either a social gaze or non-social gaze during their interactions.

Participants

Thirty participants (8 males, 15 females) between 18 and 30 years (mean age = 21.76 years, SD = 2.80 years) participated in the study.

Apparatus

The objects for the price judgment game were chosen after a screening test. Twenty participants have been asked to judge the price of 80 commonly used objects, with 44 of these selected as stimuli for the experiment. Half of the stimuli were selected on the basis that participants would be uncertain how much they would cost, as demonstrated by a relatively high variance between the price judgements (SD = 4.61). The remaining half of the stimuli had low price-judgement variance (SD = 0.69), meaning that the prices of these objects should be well known to participants. Stimuli from these two categories were equally split for application with either the Scitos G5 or iCub robots. Statistical comparisons of the price variance for the objects used with the two robots showed no significant differences for high ($t(20) = 0.599$, $p = .555$) or low ($t(20) = 0.359$, $p = .722$) price variance objects. The same voice was used for both robots.

Procedure

The participants were seated at the same height as the robot, facing each other over a common table. In the preparatory phase of the experiment, a series of objects were placed on the table by the experimenter. For each object, the participants were asked to provide a description and two potential prices to the robot, which would then respond with a preference for one of the prices. After this preference was made, the participants were asked to provide final feedback on whether they thought the price selected by the robot was correct. In this phase, the description, the price alternative, and correct price for the objects were provided to the participants on a written sheet. This part of the experiment was designed to provide familiarisation with the main task, where the robot and participants roles were reversed.

In the main task, an experimenter would place the stimulus object on the table and the robot would provide a brief verbal description of it, followed by two prices. The participants then had to make a choice between the two prices offered by the robot, after which the robot would give positive or negative feedback on their selection. For the positive feedback, the robot used to say "I agree with you" while nodding its head. For the negative one, the robot used to say "I do not agree", while shaking its head. If the robot had given negative feedback, the participants were asked whether they wanted to change choice to agree with the robot, saying "Do you want to change your choice?". During this part of the experiment, the robot could perform two different gaze behaviours (social and nonsocial). In the social gaze behaviour, the robot looked first at the object on the table and before starting describing the object, moved its gaze to the participants. In the nonsocial gaze behaviour, the robot gaze was fixed at the stimulus.

All participants were presented to the same predefined script of robot responses to the objects, which made no account of their own choice in the game. Participants completed 22 trials per robot.

2.1.3 Results

Social interaction with the robot was quantified through analysis of trials in which the robot disagreed with the price choice of the participants. A *change rate* was calculated as the proportion of the trials where the participants changed their selection to agree with the robot, compared to those in which they stuck with their original decision. Change rates for iCub and the primed and nonprimed presentations of G5 are shown in Figure 2.2, presented for both the social and nonsocial gaze conditions.

A forward stepwise linear mixed-effects model was fitted to the data, with change rate as dependent variable, robot (nonprimed Scitos G5, primed Scitos G5 and iCub) and gaze behaviour (social and nonsocial) as independent variables, and with participants as a random factor. Bonferroni's corrections were applied, when required. The overall effect size of the model was $r^2 = 0.55$.

The model showed a significant main effect of the robot ($\chi^2_{(2)} = 7.79, p = .020$). Post hoc comparisons showed higher change rate for the primed Scitos G5 when compared with its nonprimed version ($p = .004$). Nonprimed Scitos G5 reported a lower change rate than iCub ($p = .009$), while no differences have been found between iCub and the primed Scitos G5 ($p > .050$). Gaze had no effect on the change rate ($p > .050$), but there was a significant interaction between robot and gaze ($\chi^2_{(3)} = 11.51, p = .009$). Post hoc testing revealed that the change rate was significantly higher ($p = .025$) in the social than the nonsocial gaze condition for the iCub robot. No other significant differences have been found ($p_s > .050$).

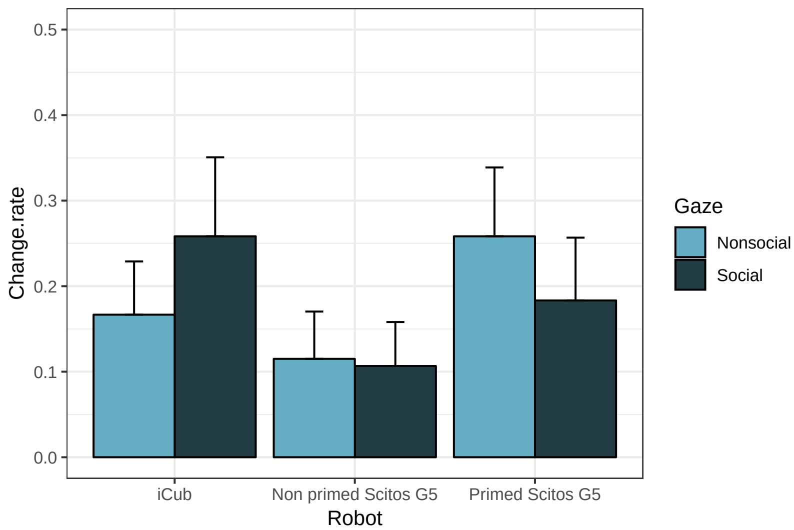

Fig. 2.2 Change rate for nonprimed Scitos G5, iCub and primed Scitos G5 for both gaze conditions (social/nonsocial)

2.1.4 Discussion

Similarly to previous findings (Fink, 2012), results from this pilot study indicate that people get more socially engaged with robots when they are more human-like. Participants were more likely to change their valuations to agree with the human-like iCub robot than with the less anthropomorphic Scitos G5 when that robot was presented in isolation. Moreover, the change rate increased when the iCub engaged in a more human-like social gaze behaviour, compared to the use of a fixed gaze. Conversely, this social gaze behaviour had no effect on interactions with the Scitos G5, where the change rate was not significantly different from the use of a fixed gaze. This would suggest that the credibility of anthropomorphic robots benefits from the automatic activation of social stereotypes that are normally absent during interactions with less-anthropomorphic variants.

However, once humans have experience of an anthropomorphic robot, they seem to extend this classification, and the associated social benefits, to subsequent interactions with a less human-like robot. Participants' change rate with the Scitos G5, in fact, was significantly higher if participants had first interacted with the iCub robot. Also, once the less human-like Scitos G5 had been presented after the anthropomorphic iCub, there was no longer difference in change rate between these two robots. It should be noted, however, that even after encountering iCub, Scitos G5 did not benefit from the use of social gaze. This could be due to the lack of congruence between gaze behaviour and robot physical form.

These findings have important consequences for the credibility of robots and their acceptance, e.g. in the increasing need and advocacy for the use of robots to assist the elderly in their homes (Burgoon et al., 2000). However, the functionality required in this environment often precludes the use of anthropomorphic physical forms, which hinders acceptance (Broadbent, Stafford, & MacDonald, 2009). This study suggests that if users are first exposed to anthropomorphic robots, they could be more accepting of their less human-like, but more functional, robotic relations.

2.2 Generalisation of Anthropomorphic Stereotype

On the bases of the previous results, a new study has been conducted to test the replicability and reliability of the pilot. Interactions with NAO, a humanoid robot widely used in social robotics (Kuchenbrandt, Eyssel, Bobinger, & Neufeld, 2011), have been compared with Baxter, an industrial development robot (Freire, Barreto, Veloso, & Varela, 2009). Participants socially engaged with these two robots using the previous price judgement task developed by Rau et al. (2009) and gaze and joint attention were manipulated like in the pilot study.

The aim of this experiment was to examine if attributions of anthropomorphism are specific to an encountered non-human agent, or whether they were also generalised across other members of that agent type. While half of the participants played the price judgement game with the NAO robot first and the Baxter robot second, this order was reversed for the remaining participants. According to the previous hypotheses, the credibility of the Baxter robot should be higher when it is presented second, after participants have played with the NAO robot, than when it is presented first. Conversely, the credibility of the NAO robot should be lower when it is presented second, following interactions with the Baxter, than when it is presented first. Similar comparisons were also made for the gaze behaviour of the robots, comparing the use of a more anthropomorphic joint attention to a condition where they simply stared fixedly ahead, expecting an increase of credibility in the first condition.

2.2.1 Method

Each participant made a separate series of price judgement interactions with two different types of robot, the human-like NAO or non-anthropomorphic Baxter robot. These robots could also be programmed to behave with either a social gaze or non-social gaze during their interactions.

Participants

Thirty participants (5 males, 25 females) between 18 and 30 years (mean age = 23.20 years, SD = 2.10 years), recruited from the School of Psychology of the University of Plymouth, participated at the study. The sample size was based on Rau et al. (2009) from which the task has been selected. Moreover, G Power 3.1 Faul, Erdfelder, Buchner, and Lang (2009) was used to determine that a sample size of twenty-five participants would provide 80% statistical power for detecting a medium-sized effect equivalent to what we observed in the previous study ($r^2 = 0.54$), assuming a two-tailed t-test and an alpha level of 0.05. Familiarity or any previous experience with robots was the main exclusion criteria, which was assessed during the recruiting phase. Participants were naïve as to the purpose of the investigation and gave informed written consent to participate in the study.

Apparatus

The NAO (Figure 2.3) is a small (57cm in height) humanoid robot widely used to study robot-human relations and is specifically designed to be both expressive and approachable and has been shown to stimulate social skills with Autistic children (Shamsuddin et al., 2012). Conversely, Baxter (Figure 2.3), although having two arms and a simple animated face, is an industrial robot designed for the study of HRC in tasks such as assembling and handling operations, as well as acting as monitoring for workers safety.

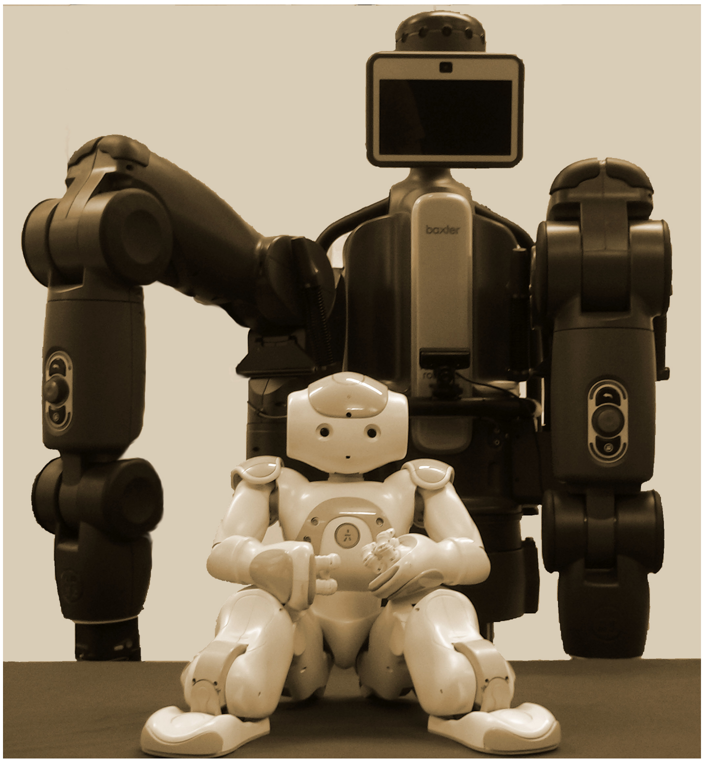

Fig. 2.3 NAO robot (foreground) and Baxter robot (background)

The same objects of the previous experiment have been used. The same voice was used for both robots. All the participants interacted with both Baxter and NAO robots, with the order of presentation counterbalanced between participants.

Questionnaires were used as secondary measures to the main experiment task. Three short scales measured Likeability, Trust and Credibility. The Likeability Questionnaire was based on Reysen (2005); the Trust scale was an adaptation of the Receptivity/Trust subscale of the Relational Communication Questionnaire and of the selection of Trust items in the IPIP International Personality Item Pool (Goldberg et al., 2006). The Credibility scale was based on McCroskey and Young (1981) Source Credibility Scale. Finally, Bartneck, Kulic, Croft and Zoghbi (2008) questionnaire was used to measure a range of HRI factors (Anthropomorphism, Animacy, Likeability, Perceived Intelligence and Perceived Safety).

Procedure

The procedure was identical to the pilot study. All participants were presented to the same predefined script of robot responses to the objects, which made no account of their own choice in the game. All participants completed 22 trials, after which they were asked to complete the battery of questionnaires.

The experiment was counterbalanced in a 2 (gaze behaviour: social or nonsocial) within-subject design and a 2 (order: first or second) by 2 (robot: NAO or Baxter) between-subject design.

2.2.2 Results

2.2.3 Game results

A forward stepwise linear mixed-effects model was fitted to the data, with change-rate as the dependent variable, robot (NAO and Baxter), order (first or second robot presented) and gaze behaviour (social and nonsocial) as independent variables, and with participants as a random factor. Bonferroni's corrections were applied, when required. The overall effect size of the model was $r^2 = 0.46$.

The model showed no significant main effects ($p_s > .05$), but there was a significant interaction between robot and order ($\chi^2_{(3)} = 10.93, p = .012$). Post hoc comparisons(Figure 2.4) revealed a higher change rate for Baxter when participants interacted with it after the NAO robot than when they interacted with it first ($p = .010$). Conversely, for the NAO robot, change rate was lower when participants interacted

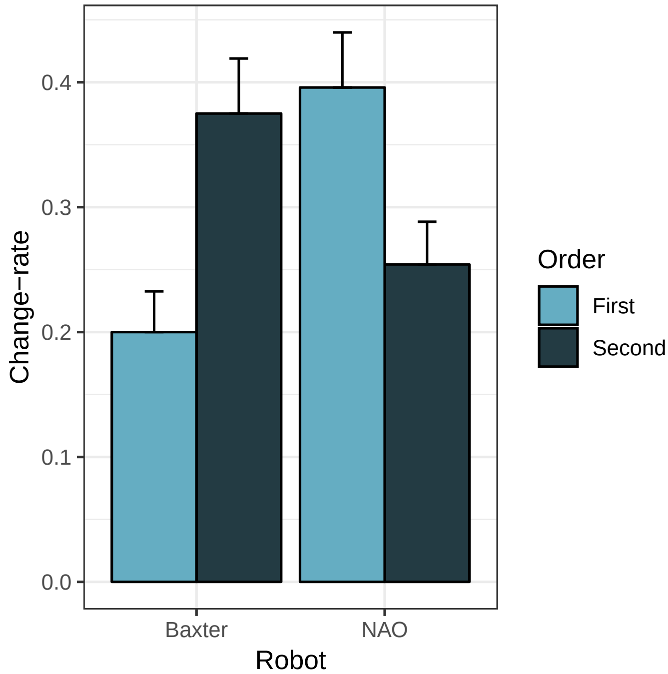

Fig. 2.4 Change rate for Baxter and NAO robots in their first and second presentation.

with it after the Baxter robot (p = .047). A comparison between robots in their first presentation also revealed that the NAO had a higher change rate than Baxter (p = .003).

A significant interaction between robot and gaze behaviour ($\chi^2_{(2)} = 14.77, p < .001$) revealed a higher change rate for the NAO robot when showing social gaze (p = .036). For Baxter robot, gaze behaviour had no effect on change-rate ($p > .05$).

The three-way interaction between robot, order and gaze was also found to be significant ($\chi^2_{(2)} = 11.83, p = .003$). Post-hoc comparisons, assessed using t-tests (Figure 2.5) , showed a significant effect of gaze only with the NAO robot, and then only when the NAO was the first robot encountered by the participants (p = .001). In this particular case, social gaze increased the change rate. None of the other pairwise post hoc comparisons was significant ($p_s > .05$).

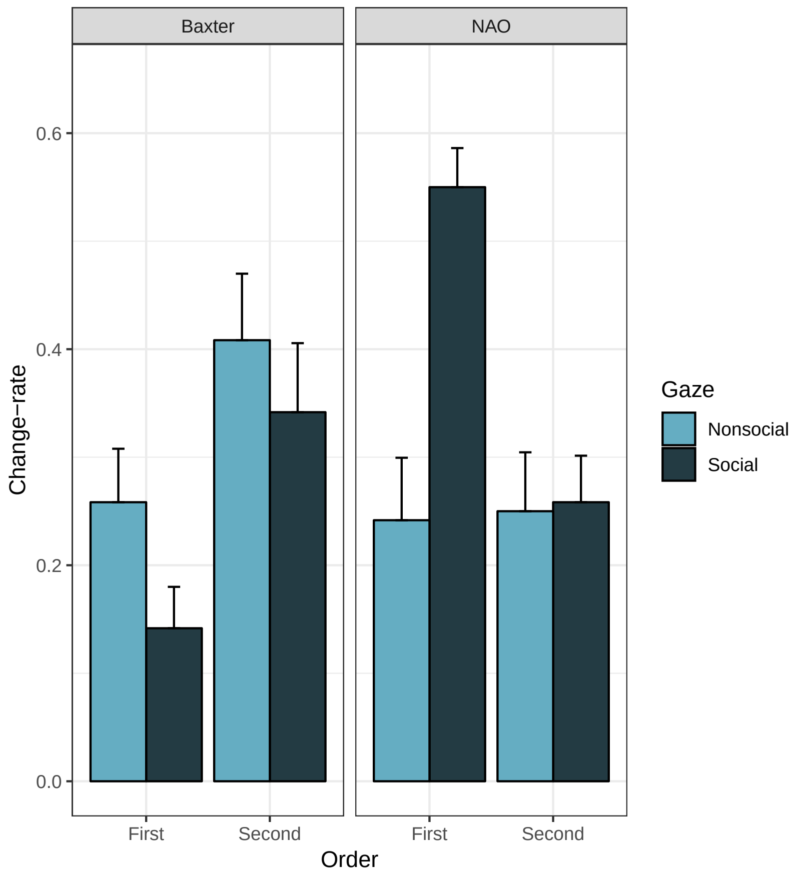

Fig. 2.5 Effect of gaze behaviour on change rate for both robots (NAO and Baxter.

Questionnaires results

Table 2.1 Cronbach's alpha.

	Alpha
Likeability	0.89
Trust	0.90
Credibility	0.89
Godspeed Questionnaires	
Anthropomorphism	0.88
Animacy	0.82
Likeability	0.91
Intelligence	0.82
Safety	0.46

Table 2.2 Questionnaires analysis results

	Order	**Robot**	**Order by Robot**
Likeability	n.s.	$X^2_{(1)} = 25.24, p < .001$	$X^2_{(2)} = 12.15, p = .002$
Trust	n.s.	$X^2_{(1)} = 14.18, p < .001$	n.s.
Credibility	n.s.	n.s.	n.s.
Godspeed Questionnaires			
Anthropomorphism	n.s.	$X^2_{(1)} = 25.22, p < .001$	$X^2_{(2)} = 6.52, p = .038$
Animacy	n.s.	$X^2_{(1)} = 28.86, p < .001$	$X^2_{(2)} = 7.76, p = .021$
Likeability	n.s.	$X^2_{(1)} = 29.65, p < .001$	$X^2_{(2)} = 7.19, p = .027$
Intelligence	n.s.	n.s.	n.s.
Safety	n.s.	$X^2_{(1)} = 5.55, p = .018$	n.s.

For each questionnaire, a linear mixed-effects model was fitted to the data, with each scale as the dependent measure, robot (NAO and Baxter) and order (first or second robot presented) as independent variables, and with participants as a random factor (Table 2.2).

Participants preferred the NAO robot in all scales, with the exception for Credibility and Intelligence, for which robot had no main effect. Moreover, the in-

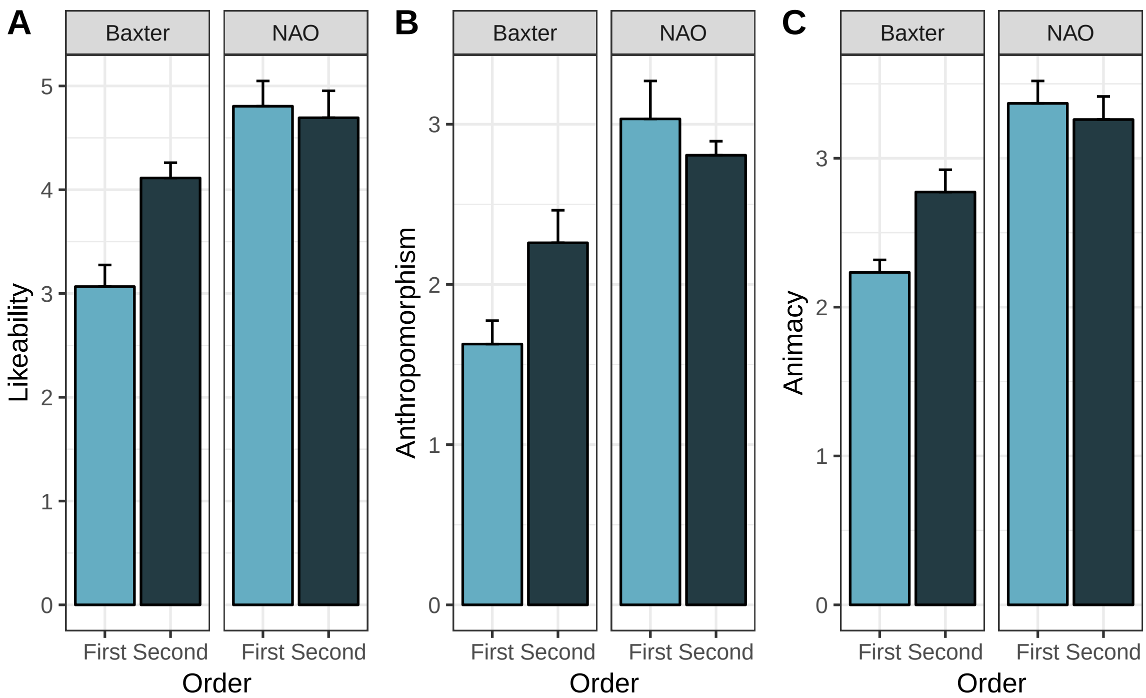

Fig. 2.6 Participants' ratings for Baxter and NAO robots in their first and second presentation for Likeability (Panel A), Anthropomorphism (Panel B) and Animacy (Panel C) scales.

teraction between order and robot showed that when presented after the more anthropomorphic NAO robot, Baxter had significantly higher ratings for Likeability, Anthropomorphism and Animacy scales than when it was presented first ($p_s < .050$) (Figure 2.6). For the NAO robot, there was no significant effect of order for any of the questionnaire scales ($p_s > .050$).

2.2.4 Discussion

This study sought to understand one of the fundamental psychological properties of anthropomorphic projection, its specificity. When we anthropomorphise a non-human agent, we ask whether the assignment of that trait is strictly limited to that particular agent, or can also be generalised across other exemplars of the same type. Importantly, if there are variations in the distribution of the anthropomorphic features across a particular type of agent, will this category effect lead to us follow exemplars that are missing the requisite features? It has been contended that the prior activation of a stereotype affects overt behaviour in a later encounter. This hypothesis has been tested by exploring whether prior experience with a humanoid NAO robot would increase the credibility and anthropomorphism of a subsequently encountered Baxter robot, an industrial robot lacking many of the anthropomorphic features of the NAO. Importantly, if this was an effect of stereotype generalisation, then also the reverse effect should be found if the first encounter was with Baxter robot.

Results showed that the human-like form of the NAO robot increased the credibility of the Baxter robot, leading participants to associate a range of human-like traits and linked behaviours to it. Data from questionnaires showed that participants rated the NAO robot as being more anthropomorphic than the Baxter robot, also rating it as being significantly more likeable, trustworthy, safe, and with greater animacy. However, no significant main effect between these robots has been found in the primary measure, that of change rate. This was the proportion of trials where participants changed their valuation of common objects to conform to the judgement of the robot, an implicit measure of the credibility of the robot. This provided an indirect measure of the degree of anthropomorphism associated with the robot, as it has been well established that human-like robots are perceived as more credible than less human-like variants (DiSalvo et al., 2002; P. A. Hancock et al., 2011; Waytz et al., 2014). Instead, the cross-over interaction between order and robot satisfied the initial hypothesis. That is, the Baxter robot was perceived as being significantly more credible when participants interacted with it after the NAO robot. Conversely, the NAO robot was perceived as being more credible when participants interacted with it first, rather than after the Baxter. The questionnaire data also showed that experience with the NAO had a humanising effect on the Baxter, leading participants to rate Baxter higher for anthropomorphism, likeability, and animacy. However, the questionnaires data for the NAO robot were not significantly affected by prior experience with Baxter. This disparity between an implicit measure of credibility seen in the change rate of the NAO and the explicit post hoc rating provided by participants in the questionnaire is not unusual (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). However, it does highlight the importance of implicit measures in the field of HRI where questionnaires, which may not always be sensitive to conceptual correspondence, are often the primary or sole source of behavioural data.

These results can be considered a reliable replication of the previous pilot experiment (Section 2.2.2). However, it should be considered that in both studies the role of the anthropomorphic agent was played by two robots that did not completely embed human-like features. iCub was built to resemble a 4 years old child, thus the change rate could be affected by the perception of low competence that a 4 years old child would have in this type of task. On the other hand, the reduced size of the NAO robot could lead participants to perceive it more as a toy than a competent agent. Considering this, results should be cautiously weighted and future research should further investigate this facilitation effect with a more adult-like type of anthropomorphic robots. Lastly, future research should also investigate the potential extension of this effect by using a human model. If the exposure to an anthropomorphic agent increases the credibility of a less human-like robot, will a first interaction with a human increase our willingness to accept a robot suggestion? Following evidence on human-based research, it could be assumed that prior positive experience will increase the chance of having a subsequent profitable interaction. However, considering that the anthropomorphic robots used in these studies were lacking some proper adult-like features, further studies are essential to better understand this effect.

In addition to the study of anthropomorphic cues inherent to the form of the robots, the effect of human-like behaviour on the credibility of the robots was also examined. Both NAO and Baxter robots could engage in social or nonsocial gaze behaviour during the valuation task: moving their heads to engage in joint attention when discussing the valued object, or simply staring fixedly ahead. As this behaviour is orthogonal to the form of the robot, it also helps to investigate interactions between anthropomorphic cues related to form and behaviour. Results showed that the effect of social gaze on credibility was limited to the NAO robot, and only when it had not been preceded by the Baxter. This interaction is in line with previous findings demonstrating that social gaze was only a human-like prerogative (Zanatto, Patacchiola, Goslin, & Cangelosi, 2016) and did not provide an effective anthropomorphic cue; otherwise, it should have had an effect similar to that of form. Rather, social gaze was only effective if participants had assigned a human-like status to a robot. This is consistent with the activation of behavioural representations (Dijksterhuis & Van Knippenberg, 1998) or cognitive structures relating to interpersonal interactions (Baldwin, Carrell, & Lopez, 1990), characteristic of the human-like trait. Other studies have shown that behaviour and interaction can be a more important cue to humanlike status than form (Kahn Jr et al., 2007; Kiesler & Goetz, 2002). However, these results might be consistent with Guthrie's psychological explanation of anthropomorphic projection (Guthrie & Guthrie, 1993). In contrast to the induction-based explanation provided by Epley et al. (2007), Guthrie contends that anthropomorphism is not the result of our desire for simplified representation. Rather, it is the result of our seeking to project the most complex possible organisation onto any stimulus. As living intentional beings offer the greatest complexity, we seek to assign these properties onto an unknown agent as long as its features do not directly exclude them. If intentional social interactions, such as joint attention, were at a higher level of complexity than anthropomorphic form, any agent that violated the requirements of form would also exclude this higher level of attribution.

2.3 Conclusion

In this chapter, it has been shown that our anthropomorphic attribution to an agent, either as human or non-human, can potentially affect credibility in subsequent encounters with similar agents. This stereotype activation means that our initial attributions can to be generalised to other variants of a similar type, even if they possess features that may be incongruent with that initial attribution. This extension of our understanding of the psychology of anthropomorphism has a growing application in the field of HRI. While robots are increasingly used in the manufacturing industry and have wide portrayal in the media (Bartneck, 2013), our own personal interactions with robots are extremely limited. This means that, for most of us, our first interaction with robots lies in the future, as does our potential anthropomorphism of these complex machines. Both the benefits and problems of anthropomorphism are particularly acute in robots, where familiarity (Choi & Kim, 2009) and believability (Tapus, Mataric, & Scassellati, 2007) are keys to our acceptance of robots. Acceptance is a particular problem in health and adult social care, where robotic assistance has great potential in the care of the elderly (Broadbent et al., 2009). Unfortunately, in this application, as in many others, the physical or technological constraints dictated by the required functions of a robot, are often not compatible with a humanoid form. However, these two studies show that if the experience of more practical robots could be preceded with more human-like robots, then we would be more accepting of their appearance and behaviour.

Chapter 3

New rules for Trust and Anthropomorphism in HRI: humans in bad times, robots in good times.

The devil's voice is sweet to hear.

Needful Things

Stephen King

3.1 Introduction

This chapter provides a valuable insight into the development of trust between humans and robots and the role of anthropomorphism in increasing or reducing trust.

Although an individual's propensity and personal history can initially affect trust, its development and persistence over time depend on the quality of the interaction with others (Asch, 1946). Trust relies on social interaction, thus the capability of expressing and understanding intentions importantly affects it (McCabe, Rigdon, & Smith, 2003), and verbal and non-verbal communication both play a role. In this chapter, voice and joint attention have been investigated as communicative factors affecting trust in HRI. Burgoon, Birk, and Pfau (1990), in fact, in examining the communicative cues affecting persuasion and credibility in interaction, indicated vocal (e.g. fluency and pitch) kinesic and proxemic (e.g. eye contact, smiling, gestures) features as the most impactful elements.

The investment game, a methodology derived from game theory, has been used to test implicit trust attributions to a NAO robot over time. Specifically, the money that participants invested in the robot has been defined as the implicit measure of trust (Berg, Dickhaut, & McCabe, 1995; Camerer, 2011; Van't Wout & Sanfey, 2008). The independent variables consisted of the manipulation of three characteristics of the robotic agent – voice, attention and behaviour – in order to measure how anthropomorphic perception and experience influence trusting behaviour. The robot voice was either natural or synthetic, the attention was either joint or nonjoint and the behaviour was either generous (high payoff) or mean (low payoff). A summary of the robot manipulations can be found in Table 3.1.

Table 3.1 The 8 robot manipulation conditions.

	Voice	Attention	Behaviour
1	Natural	Joint	Generous
2	Natural	Joint	Mean
3	Natural	Non joint	Generous
4	Natural	Non joint	Mean
5	Synthetic	Joint	Generous
6	Synthetic	Joint	Mean
7	Synthetic	Non joint	Generous
8	Synthetic	Non joint	Mean

Previous research on vocal features showed that being able to speak (either with a synthetic or a natural voice) was enough for a robot to be treated as a competent agent (Sims et al., 2009). Nass, Steuer, and Tauber (1994) demonstrated that people apply to machine voices the same conversational and interactional rules that they would use with a human. Moreover, the sound of the voice itself was enough to make personality judgments, regardless of the speech content (Nass & Lee, 2001). Nass and Brave (2005) also found that male and female interactants reacted in the same way to natural and synthetic voices in a choice-making task, provided the voice corresponded to their gender.

When it comes to designing such a robot's voice, the question of whether to use a natural or synthetic voice arises. Would people prefer a 'congruent' robot, which approaches humanness in all its characteristics, but does not quite reach it? Or would people prefer an 'incongruent' robot, with a pre-recorded human voice, for the sake of clarity and familiarity?

Interestingly, W. J. Mitchell et al. (2011) demonstrated that a cross-modal mismatch in human voice realism can arise a feeling of uncertainty when seeing a robot. They found that a robot with a human voice, or a human being with a synthetic voice, were perceived as eerier than a robot with a synthetic voice or a human being with a human voice. Following these results, people might be expected to prefer a mechanical and synthesised voice over a human and natural one. Specifically, in this experiment, higher investments on the synthetic voice than the natural one were hypothesised.

From a nonverbal point of view, joint attention also has been found to influence participants' decisions in HRI tasks (Mutlu, Shiwa, Kanda, Ishiguro, & Hagita, 2009; Staudte & Crocker, 2011). By briefly moving their eyes, robots could affect the decision-making process, even when people do not report seeing those cues

(Admoni, Bank, Tan, Toneva, & Scassellati, 2011). Gaze and joint attention also influence a person's perception, even in terms of trust (Bayliss & Tipper, 2006; Mason, Tatkow, & Macrae, 2005; Staudte & Crocker, 2011). Moreover, Zanatto et al. (2016) reported that gaze only helped HRC if the robot was humanoid. Given this evidence, the current experiment could help to clarify whether the manipulation of a robot gaze and attention engagement would affect users' implicit trust attributions as well. Specifically, during the investment game, both joint and nonjoint attentional behaviours have been compared to establish which level of interactivity can increase trust in HRI, hypothesising a beneficial effect of gaze and pointing on it. That is, participants were expected to increase their investments in a more engaging robot that a static one.

The controversial results above mentioned on the level of verbal and nonverbal behaviours in affecting HRI, arise the question of how much anthropomorphism is necessary to produce a satisfying interaction between humans and robot. Although anthropomorphism is generally considered a facilitator in HRI, literature has demonstrated that under specific critical conditions people would prefer a less anthropomorphic robot (Section 1.5.1). Moreover, previous experiments using the Investment Game, reported a connection between the payoff and the vocal features of the players, for which high payoffs increased trust in an SSBE-accented (Standard Southern British English) player, and low payoffs increased trust in a Birmingham-accented player (Torre, Goslin, & White, 2015).

For these reasons, in this study, the type of payoff has been manipulated, by having participants engaging in a game with a more or less profitable outcome. Participants could face a generous robot, always repaying economic trust with a nice amount of money, or a mean robot showing the selfish tendency of not returning enough money to establish a cooperative trustworthy exchange. Following the literature, participants were expected to prefer a more anthropomorphic robot when playing with a mean robot, while for participants engaging in a game with a generous robot, higher preference for a static and machine-like robot was hypothesised. More specifically, for a mean robot, higher investments in a natural voice and joint attention behaviours were expected. On the opposite, when generously rewarded, participants were expected to maintain trust in a mechanical type of voice and a non-interactive robot. The mean attitude of the robot has been considered as an extension of intentional behaviour. Specifically, it has been used to understand how humans face a non-human entity that shows intention. This would help in understanding whether people can apply to machine human-human social mechanisms. Literature has demonstrated that unexpected behaviours coming from a robot lead to higher anthropomorphisation. In the same way, a mean behaviour can be categorised as an unusual and unexpected attitude in a robot. In this sense, the mean behaviour has not been built as a potential future feature that a robot could embed but has been used to further understand how people react and respond to this kind of behaviours in robots. If we want to fully understand how HRI works, we need to investigate any potential scenarios. Once we know whether in certain conditions a more anthropomorphic behaviour could increase the chances of a profitable interaction, we know how to build a robot that can efficiently face that situation.

3.1.1 Contributions

This experiment has been partially supported by CogNovo (FP7-PEOPLE-2013-ITN-604764), a project funded by the EU Marie Curie programme.

The researchers involved in the study and their contributions are listed below:

- Debora Zanatto contributed to the ideation and creation of attentional stimuli (gaze engagement and pointing behaviour), the design of the display (see Figure 3.2), the creation of the scripts for the behaviours of the robots (including also the payoff, which can be found in Appendix A), testing and the statistical analyses here reported.
- Ilaria Torre contributed to the ideation and creation of vocal stimuli (natural voice recordings and resyntheses; the list of the sentences used can be found in Appendix B), the design of the display (see Figure 3.2), and testing.
- Massimiliano Patacchiola contributed to the programming of the robots and coordinated the synchronization of the robots and the gaming device.

3.2 Method

3.2.1 Participants

One hundred twenty-two individuals (82 female and 40 male; mean age = 23.12 years, SD = 8.62 years) participated in the study. All participants were native British English speakers, they were right-handed, reported normal hearing and no language or neurological impairments. Familiarity or any previous experience with robots was the main exclusion criteria, which was assessed during the recruiting phase. Participants were naïve as to the purpose of the investigation and gave informed written consent to participate in the study.

3.2.2 Procedure

Investment game

Participants were seated in front of a NAO robot, with a touchscreen monitor standing between the two (Figure 3.1). They were told that the goal of the game was to earn as much money as possible and that mutual cooperation with the robot would lead to greater profit. Each participant started with a notional sum of 10 Experimental Currency Units (ECU) at the beginning of each round of the game. Each round proceeded as follow: the participants heard a previously recorded utterance, played from the robot speakers; then they indicated how much, if any, of the 10 ECU they wished to invest, by tapping a number on the monitor (e.g. participant invested 5 ECU); the robot then received 3 times the invested amount (e.g. 5*3 = 15 ECU). By following a fixed script, the robot returned a percentage of that sum back to the participants (e.g. 15 * 0.3 = 4.50 ECU).

Fig. 3.1 Investment Game setup.

The robot was programmed to apply two different types of payoff, namely generous and mean. For the generous condition, the robot could return between 50 % to 80 % of the tripled amount. For the mean condition instead, the robot could return between 0 % and 30 % of the received amount. These percentages changed at each round but were fixed between games.

Before the game started, the participants engaged in 3 practice rounds with the experimenter, in order to familiarise with the interface. The display had 11 numbered touch-screen buttons on the side of the participants, which they could press to indicate the amount of money to invest in each round (from 0 to 10). On the robot side, there was an orange slider, which was used to give a visual representation of how much money the robot was returning. The starting position of the slider was always at its midpoint, as can be seen in Figure 3.2. The slider then moved according

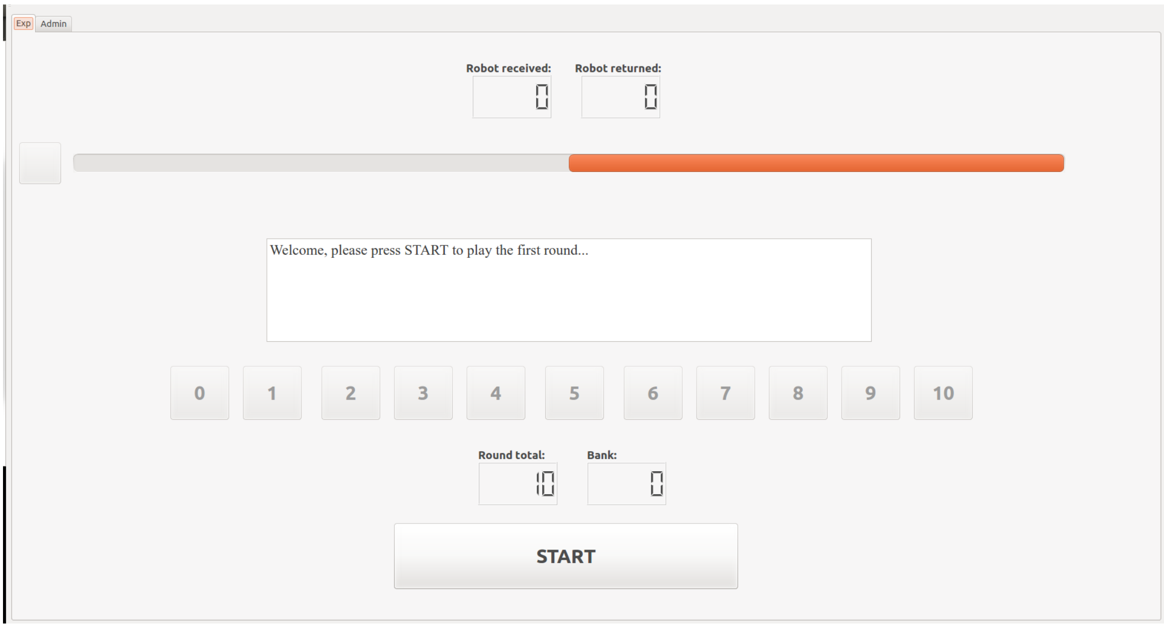

Fig. 3.2 Investment Game display

to the percentage of the robot returns: it moved to the left of the participants when the robot returned more money than they invested and to the right when it returned less money. The computer screen displayed also all the monetary transactions in real time, including the amount that the robot returned to the participants and the total money accumulated round after round. The 'bank' button showed the sum of money that participants had earned in all the previous rounds. The buttons at the top of the screen showed how much money the robot had received and returned. The display was aligned to the participants' visual, to ensure ease of understanding.

Participants engaged in two games of 20 rounds each with two equally looking NAO robots. The number of rounds was not known to the participants. The total amount that participants earned at the end of each game was converted to British Pounds, at a rate of 30 ECU = £0.10 and the sum of what they earned in the two games was paid to them before they left. A set of questionnaires was filled at the end of each game. After playing two games and completing the questionnaires, participants were debriefed and paid.

The robot voice and attention were manipulated during the game. In terms of voice, the robot had either a natural, pre-recorded female British English voice or a synthetic voice obtained from the resyntheses of the same British voice. Moreover, voices were recorded from two different speakers.

In terms of attention, the robot arm and head movements were manipulated. In the joint attention condition, the robot followed the participants' movements, giving an impression of looking at the participant. When it was the robot turn in the game, it lowered its head, 'looking' at the screen, and performed a sweeping arm movement, congruent with the slider direction.

In the nonjoint attention condition, instead, the robot was standing still with its head lowered, and never looked at the participants, thus giving the impression that it was always looking at the screen. A detailed timeline of all the behaviours performed by the robot is shown in Figure 3.3.

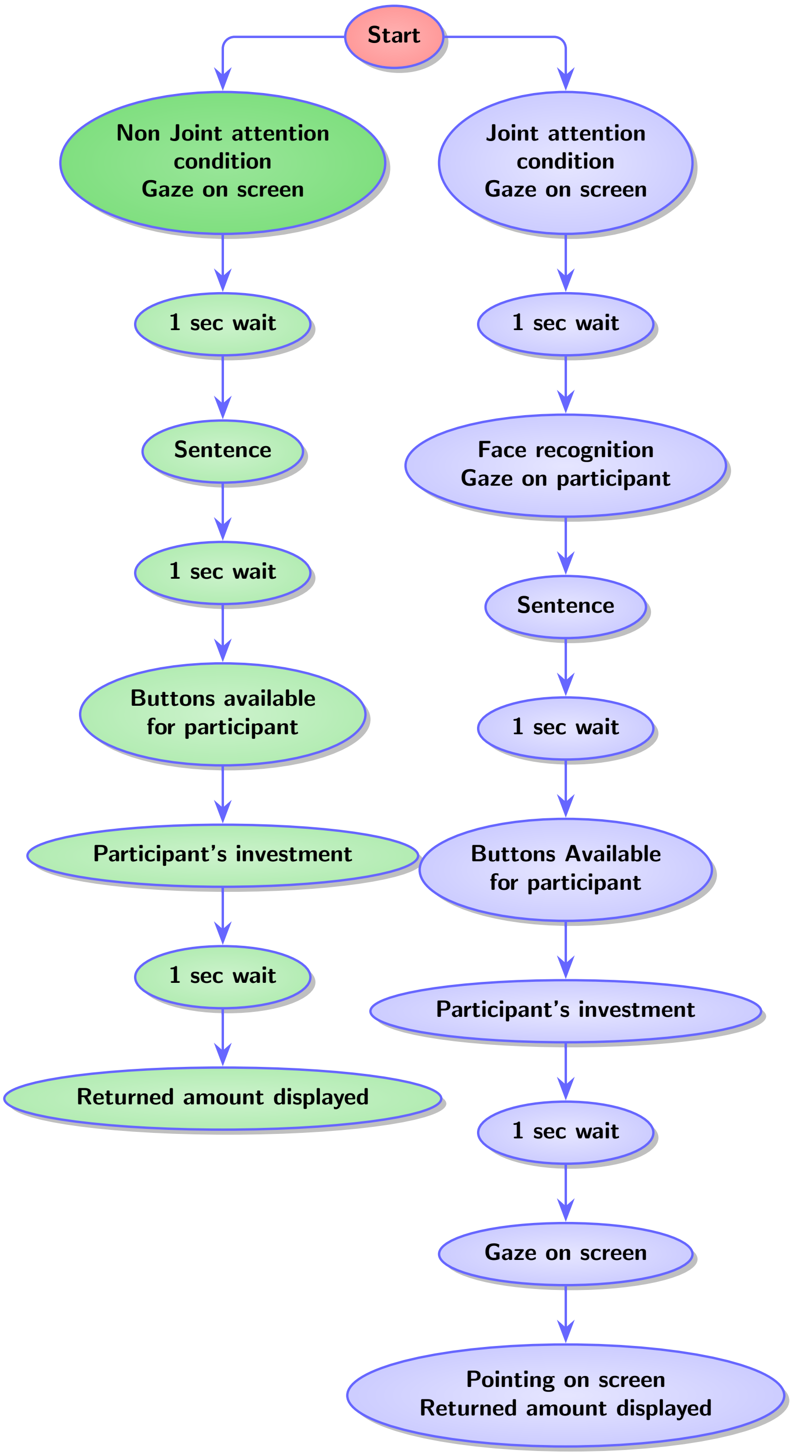

Fig. 3.3 Experimental Timeline for joint and nonjoint attention conditions.

The experiment was counterbalanced in a 2 (attention: joint or nonjoint) by 2 (speaker: 1 or 2) within-subject design and a 2 (behaviour: generous or mean) by 2 (voice: natural or synthetic) between-subject design.

Questionnaires

Four questionnaires were used as secondary measures to the main game task. Three short scales measured likeability, trust, and credibility (Goldberg et al., 2006; McCroskey & Young, 1981; Reysen, 2005). In addition, a questionnaire from Bartneck et al. (2008) was used to measure a range of HRI factors (anthropomorphism, animacy, likeability, perceived intelligence and perceived safety). Participants filled in the questionnaires after the first investment game ended and again after the second investment game ended.

3.3 Results

3.3.1 Investment game results

Two participants did not follow the instructions during the experiment and were excluded from the data analysis. A linear mixed-effects model was fitted to the data using backward stepwise selection, with investment as dependent variable, behaviour (generous/mean), voice (natural/synthetic), attention (joint/nonjoint), speaker (1/2) and game turn (the 20 rounds of the game) as independent variables, and participants id as random factor. The overall effect size of the model was $r^2 = 0.51$. Post-hoc comparisons were assessed using t-tests and Bonferroni's correction was applied, when needed.

There was a main effect of behaviour ($\chi^2_{(1)} = 163.17, p < .001$) with higher investments in the generous condition (mean = 8.13 ECU) than in the mean condition (mean = 4.6 ECU), as can be seen from Figure 3.4. There was also a main effect of the game turn ($\chi^2_{(1)} = 13.02, p < .001$), with overall higher investments in the second half of the game. There was no effect of attention, voice and speaker ($p_s > .05$).

A significant two-way interaction between behaviour and game turn has been found ($\chi^2_{(1)} = 61.05, p < .001$). Post hoc analyses revealed a game turn effect for generous condition ($\chi^2_{(1)} = 267.47, p < .001$), for which investments increased over time Figure 3.4. For the mean condition, no effect of game turn have been found ($p > .050$).

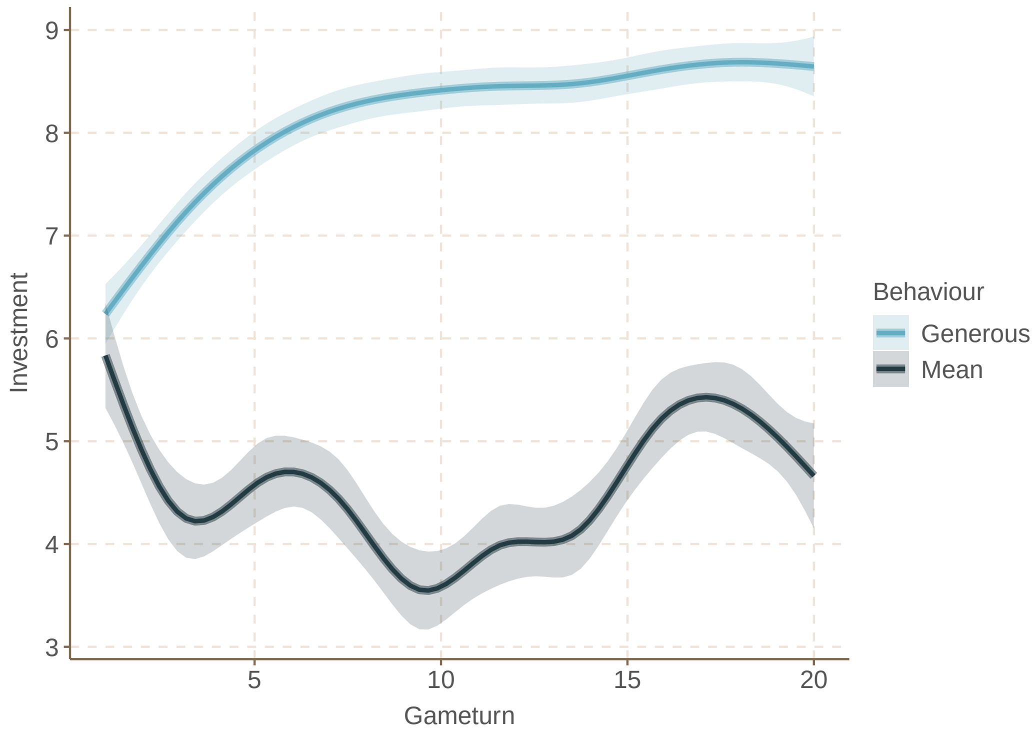

Fig. 3.4 Average investment in the generous (top) and mean (bottom) conditions.

A significant two-way interaction between voice and game turn has been found ($\chi^2_{(1)} = 7.44, p = .006$). Post hoc analyses revealed a game turn effect for both voices (natural $\chi^2_{(1)} = 73.83, p < .001$; synthetic $\chi^2_{(1)} = 28.26, p < .001$ Figure 3.5). Participants invested more money in the synthetic voice in the first half of the game. However, the difference between the two voices disappeared toward the end of the game.

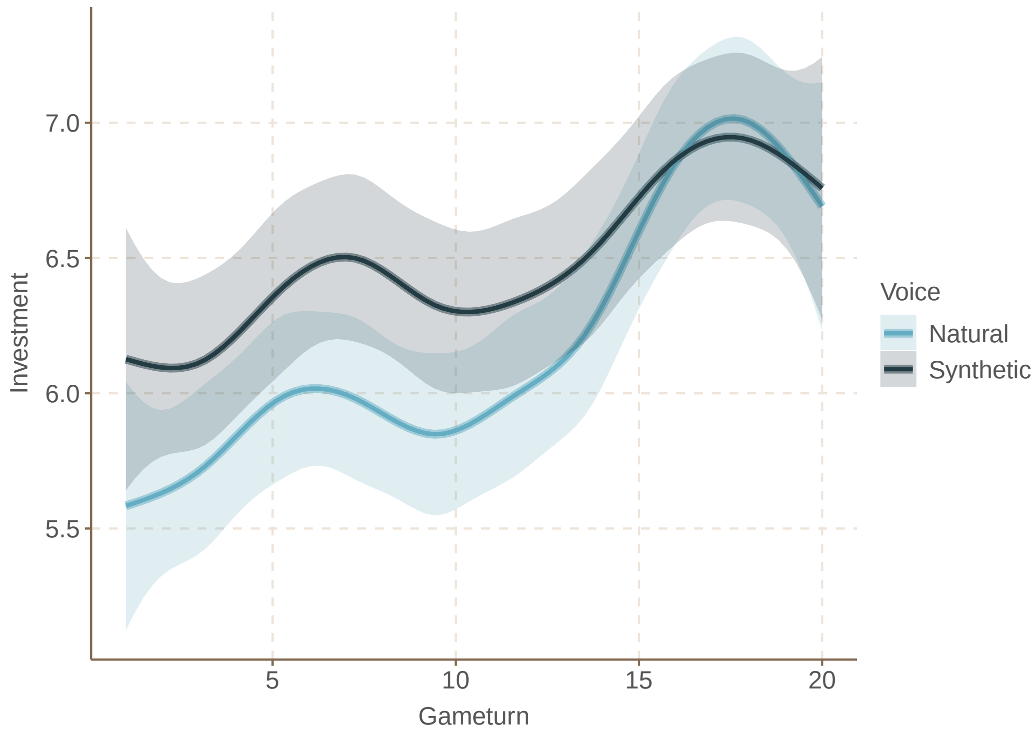

Fig. 3.5 Average investments in the two voice conditions over time.

The significant two-way interaction between voice and behaviour ($\chi^2_{(2)} = 6.63, p = .036$) showed that participants in the generous condition preferred to invest in a synthetic voice over a natural one ($p < .001$). No other significant differences have been found ($p_s > .05$).

There was also a significant three-way interaction between behaviour, voice and game turn ($\chi^2_{(1)} = 18.84, p < .001$). A post hoc mixed-effects model was fitted to the generous and mean data separately to analyse this interaction. The investment was the dependent variable, game turn and voice were predictors, and participants id was the random factor. In the generous condition, there were a main effect of game turn ($\chi^2_{(1)} = 269.44, p < .001$) and a main effect of voice ($\chi^2_{(1)} = 4.18, p = .041$), but no significant interaction between the two factors has been found ($p > .05$). As can be seen from the bottom panel in Figure 3.6, investments increased over time and were higher in the synthetic voice condition. In the mean condition, there were no effect for game turn and voice, but there was a significant interaction between game turn and voice ($\chi^2_{(1)} = 16.17, p < .001$). As shown in the bottom panel of Figure 3.6, investments increased in the second half of the game for the natural voice.

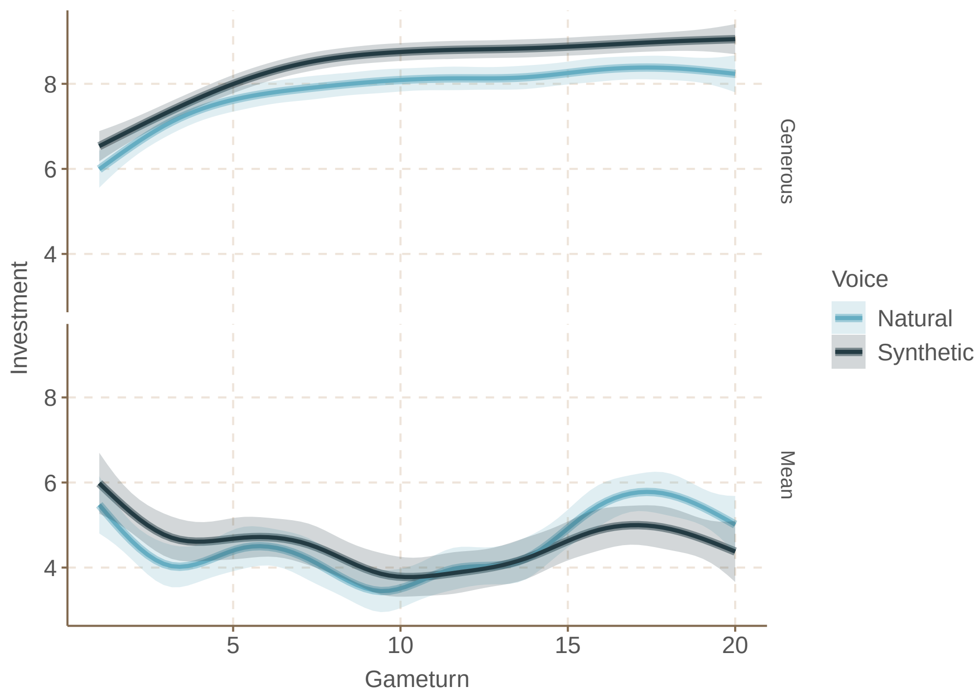

Fig. 3.6 Average investments in the two different voice conditions and in the mean (top) and generous (bottom) behaviour conditions over time.

A significant two-way interaction between game turn and attention was also found ($\chi^2_{(1)} = 7.80, p = .005$). Post hoc analyses revealed a game turn effect for both attention conditions (joint $\chi^2_{(1)} = 84.95, p < .001$, nonjoint $\chi^2_{(1)} = 27.13, p < .001$). Participants invested more money in the second half of the game in both attention conditions, although investments were higher for the joint attention condition (Figure 3.7).

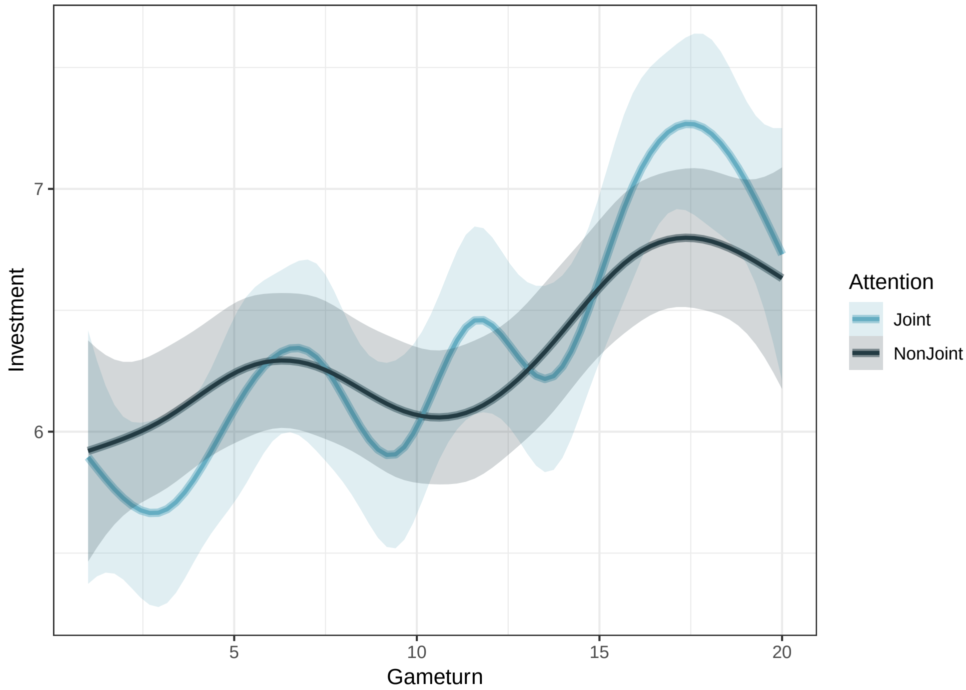

Fig. 3.7 Average investments in the two attention conditions over time.

Finally, a significant interaction between behaviour and attention confirmed the main behaviour effect ($X^2_{(2)} = 6.37, p = .041$). For both attention conditions, the generous behaviour received higher investments over the mean one ($p_s < .001$).

3.3.2 Questionnaires Results

For all the scales, a linear mixed-effects model was fitted to the data, with behaviour (generous/mean), voice (natural/synthetic), speaker and attention (joint/nonjoint) as independent variables, and participants id as a random factor. Post-hoc comparisons were assessed using t-tests and Bonferroni's correction was applied, when needed. As shown in Table 3.3, behaviour affected Trust, Credibility, Likeability and Intelligence scales, with higher ratings for the generous robot. Attention affected all the scales, showing higher scores for the joint attention condition. Voice had no effect. Speaker affected Credibility with higher ratings for speaker 2 than speaker 1.

Table 3.2 Cronbach's alpha.

	Alpha
Likeability	0.91
Trust	0.82
Credibility	0.89
Godspeed Questionnaires	
Anthropomorphism	0.88
Animacy	0.87
Likeability	0.94
Intelligence	0.83
Safety	0.56

Table 3.3 Main effects of questionnaires.

	Behaviour	**Attention**	**Voice**	**Speaker**
Likeability	n.s.,	$\chi^2_{(1)} = 12.76, p < .001$	n.s.	n.s.
Trust	$\chi^2_{(1)} = 46.45, p < .001$	$\chi^2_{(1)} = 25.75, p < .001$	n.s.	n.s.
Credibility	$\chi^2_{(1)} = 47.74, p < .001$	$\chi^2_{(1)} = 14.83, p < .001$	n.s.	$\chi^2_{(1)} = 4.36, p = .037$
Godspeed Questionnaires				
Anthropomorphism	n.s.	$\chi^2_{(1)} = 30.00, p < .001$	n.s.	n.s.
Animacy	n.s.	$\chi^2_{(1)} = 59.94, p < .001$	n.s.	n.s.
Likeability	$\chi^2_{(1)} = 62.63, p < .001$	$\chi^2_{(1)} = 12.57, p < .001$	n.s.	n.s.
Intelligence	$\chi^2_{(1)} = 15.62, p < .001$	$\chi^2_{(1)} = 5.46, p = .019$	n.s.	n.s.
Safety	n.s.	$\chi^2_{(1)} = 7.04, p = .008$	n.s.	n.s.

Both Likeability scales showed a significant two-way interaction between behaviour and voice ($\chi^2_{(2)} = 9.19, p = .010$; $\chi^2_{(2)} = 6.64, p = .036$), with higher ratings for the synthetic voice in the generous behaviour condition ($p_s > .05$), and for the natural voice in the mean behaviour condition ($p_s > .05$).

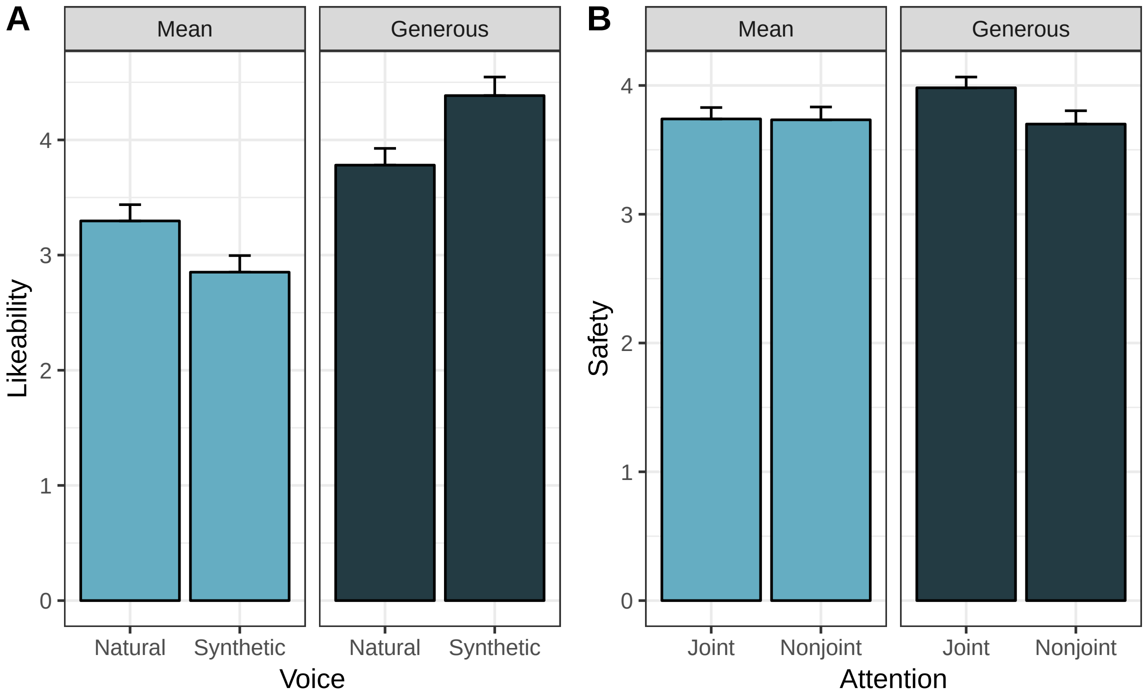

Fig. 3.8 Likeability ratings for natural and synthetic voices in the mean and generous conditions (panel A). Safety ratings for joint and nonjoint attention in the mean and generous conditions (panel B)

For Safety, a two-way interaction between behaviour and attention emerged ($\chi^2_{(2)} = 7.52, p = .023$). The joint attention condition was perceived as safer than the nonjoint in the generous behaviour condition ($p = .035$), but no differences for attention were found in the mean behaviour condition ($p > .05$). Moreover, for Safety, the three-way interaction between behaviour, voice and attention was also significant ($\chi^2_{(4)} = 12.20, p = .016$). Voice had an effect in the generous and nonjoint attention condition ($p = .015$), with lower ratings for natural voice. All the other comparisons were not significant ($p_s > .05$).

3.4 Discussion

This study aimed at establishing trust between humans and robots, by measuring the potential effect of verbal and nonverbal cues as anthropomorphic features. In an iterative investment game, participants could choose the amount of money to invest in a robot showing different degrees of anthropomorphism. This has been manipulated in terms of voice (synthetic or natural) and attentional skills (showing joint attention engagement or not). Moreover, the payoff, as a variable affecting trust over time, has been included so that the robot could show a generous or a mean attitude in returning money to the participants.

Participants invested consistently more money in a generous robot, in a robot with a synthetic voice and in a robot performing joint attentional behaviours. Participants were susceptible to the robot payoff; they responded with higher investments to a generous robot, while reduced their monetary endowment to a mean agent. Higher investments with a generous game partner were expected, as the goal of the participants was ultimately to earn more money, and as such, they learned very quickly if the partner was being trustworthy or not. This type of behaviour follows HHI rules, thus it could be argued that participants in this study ascribed intentionality to the robot investment choices, whereas a low payoff was not categorised as a mistake in following the game rules. This was further confirmed by the analyses of the questionnaires, in which participants' ratings varied on the basis of the robot payoff. Furthermore, this effect could also be interpreted in terms of commitment (Powell & Michael, 2019). As the robot in the generous condition, kept returning a consistent amount of money, it is plausible that participants felt motivated to keep investing also as a way to repay the trust. The sense of commitment implies that the two parties expect the other will fulfil his/her obligation and will remain engaged until the final joint goal is reached. Here, participants might have felt engaged in maintaining a high level of trust in the generous robot, in order to fulfil the robot expectations. This hypothesis is supported by Michael's theorisation on the sense of commitment in HRI. As the authors stated "even if a human agent does not explicitly believe that a robot can make commitments, or that she herself owes it to a robot to honor a commitment that she has made to the robot, she may nevertheless implicitly sense the opposite" (Michael & Salice, 2017). Furthermore, Székely et al. (2019) also found evidence of commitment in HRI. In a 2-player version of the classic snake game, the authors measured how long participants waited before ending the game, that was becoming gradually boring over time. Results showed that participants waited longer when they believed that the robot was investing effort to accomplish the task.

Regarding vocal cues, the first hypothesis focused on establishing which type of voice could influence the development of trust. Comparisons between a synthetic and a natural voice, showed participants investing more money in a robot with a machine-like voice. In line with the hypothesis, participants overall trusted voice congruent robots more than incongruent ones. This result confirmed W. J. Mitchell et al. (2011) finding and demonstrated that a natural voice can be perceived as a cross-modal mismatch between the voice and the physical features of the robot. That is, a natural voice can be too human for a machine. However, participants' investments toward the end of the game were not affected by the type of voice. In the natural voice condition, in fact, participants increased gradually their investments and reached the same amount of the synthetic voice. This result indicates that, although a congruent synthetic voice is preferred, people can learn to accept and trust an 'uncanny' robot if given enough time.

A third hypothesis concerned a facilitation effect of joint attention. In comparing interactions with a robot showing gaze engagement and pointing gestures with a static robot, participants were expected to invest more money in the former condition. Results partially confirmed this hypothesis. Although participants' investments in the first half of the game were not affected by the attentional behaviour, in the second half of the game investments for joint attention consistently increased. This result indicated that overall participants preferred interacting with a robot showing hand pointing and gaze engagement skills (Admoni et al., 2011; Staudte & Crocker, 2011), albeit this effect manifested itself after some time. While joint attention had no main effect in the game, it was consistently linked to participants' explicit judgments in the questionnaires. In particular, it was a significant predictor of perceived likeability, trustworthiness, credibility, anthropomorphism, animacy, intelligence and safety of the robot.

A fourth hypothesis was related to the limit of the anthropomorphic effect on trust. Participants were expected to prefer human-like traits in situations of low payoff (mean behaviour) while trusting more a machine-like type of robot in a more profitable condition (generous behaviour). This hypothesis has been partially confirmed by the interaction between voice, behaviour and game turn. Participants, in fact, invested more in a natural voice in the mean condition. These results are in line with the hypothesis that the context of the interaction plays a key role in the application of anthropomorphism in HRI (see Chapter 1, section 1.5.1). In the case of low payoff, participants were unable to find a strategy that could guarantee them a sufficient return. It is then possible that participants started looking for cues that could suggest to them how to increase their chances to regain some money. This cues could be derived from the voice of the robot. As Epley's theory suggested, anthropomorphism is the best tool we have to explain what is unknown. In a condition in which participants did not know how to successfully accomplish their task, the natural voice of the robot could have been the anthropomorphic feature that helped them in maintaining trust. Moreover, the mean attitude of the robot could have been perceived as an intention to keep all the profits. This type of behaviour is mostly a human prerogative, thus associating it to human-like features could increase participants' willingness to invest in the robot.

On the contrary, in the generous condition, participants preferred to invest in a machine-like voice. In this condition, the anthropomorphic features of the robot could have been an unnecessary complement. In the case of high payoff, the strategy to apply is easy to be found, thus no particular effort was needed during the game. This type of scenario was more consistent with the classical vision of HRI, in which the robot is supposed to perform with our needs. Since no specific effort was needed and since the robot was performing as expected, anthropomorphic features became a useless complement in this condition. These results have also been confirmed by the Likeability questionnaires. Participants preferred a mean robot with a natural voice, and generous one with a synthetic voice.

3.5 Conclusion

This study offered the possibility to better investigate the development of trust over time, and the limits of anthropomorphism in affecting it. Data here reported demonstrated that humans can ascribe intentionality to robots and develop a sense of commitment. More importantly, this study finally offered a reliable behavioural alternative to study trust in HRI. While there has been a wide usage of explicit methodologies, mostly through questionnaires, a reliable implicit measure of the development of trust over time was still missing. Implicit measures, differently from explicit measures, can provide evidence of cognition, inclination or attitudes which participants are not aware of. In fact, the outcome of implicit measures could become an index of the attitudes and cognition participants have no control or conscious access to (De Houwer, 2006). It has also been pointed out that explicit attitudinal data do not correlate with behavioural data (De Houwer, 2006; Greenwald, 1990). Furthermore, a downside of explicit measures such as questionnaires is that they focus on immediate impressions, without taking into account how the attribution might evolve over time. While explicit measures are more frequently used in HRI research, there are a few studies collecting implicit measures of trust. DeSteno et al. (2012) recorded face-to-face verbal interactions between human participants and a Nexi robot, to identify sequences of non-verbal cues that were indicative of trustworthy behaviour, thus demonstrating that the accuracy of trustworthiness judgments of novel partners could be influenced by exposure to nonverbal cues. Haring, Matsumoto, and Watanabe (2013), among other measures, used proxemics, the interpersonal distance between a person and a robot, to measure the person's implicit trust towards the robot. Finally, P. A. Hancock et al. (2011) suggested that future research should use implicit data to study human-robot trust, and mentioned the problem of individual measures taken after a single interaction, which are not informative of trust development. The current results suggested that implicit and explicit measures might both contribute to understanding how trust is formed, although they might describe different aspects of it.

Human-robot trust cannot be measured in a single way. The present findings showed that, given the different caveats and interpretations inherent to different methods, one type measurement is not enough to give a complete picture of the underlying trusting behaviours and the reasons behind them.

Nevertheless, the evolution of trust in HRI has been explored throughout a small temporal window. As both voice and attention have been shown to change trust over time, future studies should further contribute to the exploration of trust evolution on a longer term. This would allow having a more complete picture of which features better increase the quality of the interaction with a robotic agent.

Chapter 4

Investigating cooperation with robotic peers

An' why? Because...because I got
you to look after me, and you got me
to look after you, and that's why.

Of Mice and Men
John Steinbeck

4.1 Introduction

Over the last decades, robotic agents have become sophisticated to the extent that anthropomorphism could lead us to accept them as our peers. Research in this field has focused upon the design factors, resemblance and socio-cognitive skills that enhance social relations with robots (Duffy, 2003; Eyssel, 2017; Fong et al., 2003; Gaudiello et al., 2016; Walters, Syrdal, Dautenhahn, Te Boekhorst, & Koay, 2008).

With the widening usage of robots, their role has moved from that of a complex automated tool to include situations where they can operate as a teammate able to assist humans in the completion of joint tasks (Ma, Fong, Micire, Kim, & Feigh, 2018). However, most of the time the role of the robot is usually subordinate or complementary, limited to assisting or complementing human teammates in their task.

Of interest to this chapter is the extension of this cooperative relationship beyond the HHI. In this study, interactions with a robot peer, a machine entity that had the same role, task, and decision-making ability as their human compatriot, were examined. The aim of the study is to recreate a cooperative relationship typically experienced in human-human studies of cooperation by establishing procedural equality between humans and robots. This has important implications for studies of HRI as a robotic peer may have different motivations, preferences or intentions to their human partner. Thus, when a robot has the same freedom of action as a human peer, these factors could affect the outcome of the interaction such that the robot might not always be disposed to cooperate in a manner desirable to the human partner. Specifically, cooperation in human-robot partnerships was examined, with the aim of understanding the perception and reaction to robots that might adopt an independent, potentially non-cooperative, strategy to that of the human partners. Although the exhibition of non-cooperative behaviour might be perceived as acceptable, albeit regrettable, in human partners, this type of behaviour typically falls outside of the envelope of accepted norms for robots. As such, a key component of the study is to examine whether our perceptions of non-cooperative robots would be modulated by how "human" the robots are in their general behaviour.

It has been established that people prefer to collaborate with a robot that is socially competent, physically more human-like, as well as capable of performing verbal and nonverbal behaviour (Fink, 2012). In particular, gaze cueing, hand and arm gestures are primary candidates for extending the communicative capabilities of social robots, tending to result in a more positive and trustworthy evaluation of the robot (Mutlu et al., 2009; Staudte & Crocker, 2011). These cues encourage anthropomorphic projections onto robots that increase the belief that they harbour their own independent intentions. Thus, if the robot exhibits human-like interactive features a lack of cooperation in the robot may be perceived as an intentional act. Conversely, the same behaviour in a less-human like robot could be ascribed to a programmed behaviour, with the perception of a non-interactive machine fixed in a predetermined behaviour.

Participants in this study were engaged in a cooperative investment game with a robotic partner. Both the robot and the participants were provided with a virtual sum of money for each round of the game and had to independently decide how much to risk in an "investment" with a robot "banker". Any money they did not invest was safe, but any sum invested in the banker could be returned to them with an additional profit. To encourage a collaborative strategy between the partners, the banker would penalise any investments as a function of the difference in the invested amounts decided upon by the human and robot partners. Therefore, more profit could potentially be earned if both the robot and the participants adopted a common, rather than a separate investment strategy. Within the experiment, the robotic partners could be programmed to be cooperative and would alter their investment strategy to approach that of the human. Conversely, non-cooperative robots would not alter their investment in relation to decisions made by the participant, adopting a completely independent strategy. In the latter case, the participant would have to be willing to adopt a similar investment strategy to the robot so as not to be penalised by the banker.

Through an examination of investment decisions and questionnaire responses provided by participants across these experiment factors, an exploration of our perception and cooperation with robotic peers has been given. The participants were expected to follow similar social conventions in HRC as they would in an HHC situation. That is, with participants preferring to reciprocate cooperative behaviours demonstrated by the robotic confederate (i.e. participants changing their investment decisions to more closely conform to that of the robot if the robot behaved likewise).

The willingness of participants to change their decisions to conform to the strategy of the robot was a key dependent measure in this study, one that is predicted to be modified by the anthropomorphic status projected onto the robot. An additional experiment factor explored this potential interaction. Half of the participants interacted with a robotic partner that showed anthropomorphic behaviours, simulating joint attention and talking to the participants. The remaining participants interacted with a robot that was both static and mute. It was then hypothesised that the likelihood of cooperative behaviour would be increased if the robot also behaved in an anthropomorphic manner, moving and talking, than if it was silent and immobile. However, based upon the results of the previous study (Section 3.4), participants were expected to be more cooperative with human-like robots when the banker was being mean but, conversely, more cooperative with less-human-like robots when the banker was generous.

Differently from the previous chapter, this study allows to examine whether this differential behaviour is also evident when cooperating with a confederate, itself a third party to that generous or mean behaviour. Considering that in the previous study participants invested in a physical mean/generous robot, this third party was introduced as a physical robot banker. This would also offer more coherence to the experimental setting in which all parties were present in the scenario. Furthermore, the presence of a banker would add the robot confederate the chance of producing more interactive actions in the anthropomorphic condition. Nevertheless, since the focus of this study was the dyadic interaction between the participants and the robot confederate, no other manipulations of the banker features and behaviours have been included. That is, the banker was solely staying immobile within every condition of the experiment.

4.2 Method

4.2.1 Participants

There were 96 participants (70 female and 26 male; mean age = 20.02 years, SD = 2.25 years) in the study, all of whom were naïve as to the purpose of the investigation and gave informed written consent of their willingness to participate. All participants were native British English speakers, they were right-handed, and reported normal hearing and no language or neurological impairments. Familiarity or any previous experience with robots was the main exclusion criteria, which was assessed during the recruitment phase.

4.2.2 Procedure

Investment game

In the experiment setting, the participants and the NAO robot confederate sat facing each other over their game interface screens, as shown in Figure 4.1. In the corner of the room, the NAO robot banker was seated, also with its own game interface screen.

Fig. 4.1 Cooperative Investment Game setup

The participants were told that the goal of the game was to earn as much money as possible. At the start of each round of the game, both the participants and the robot confederate were given a notional sum of 10 Experimental Currency Units (ECU). The participants then had to decide how much of that sum to risk in an investment with the robot banker, making a decision to invest between 0 and 10 ECUs (e.g. player invests 3 ECU). The participants were also aware that, concurrent with their own investment choice, the robot confederate was also making its own investment decision. This initial investment choice of the robot confederate followed a fixed script, for which the investment was calculated on the basis of the participant's initial choice. Specifically, the robot confederate was instructed to randomly subtract or add a number between 3 and 4 to the participants' choice. In this way, the initial choice of both parties was never too distant. Likewise, when making their initial investment decision the participants were not aware of how much the robot confederate would invest.

After both parties had made their investment decisions (e.g. participant: 3, robot: 6), the participants were shown the confederate robot initial investment choice on the screen. The robot then was able to modify that initial investment choice (e.g. robot: 6 to 5 ECU investment). After that decision, the participants were then shown the final investment choice of the confederate robot and were given the chance to modify their own investment decision (e.g. participant: 3 to 4 ECU investment). When that decision had been made, the final investment amounts for the participants and the confederate robot were summed, tripled and sent to the robot banker (e.g. banker receives 3*(4 + 5) = 27 ECU). The banker robot would then follow a fixed script and return a percentage of that sum back to the investors (e.g. 30% of 27 = 8.1 ECU, so 4.05 ECU to each player). However, this returned investment from the banker could also be modified downward, dependent upon the similarities of the final investment decisions of the participants and the robot confederate. This modifier was calculated as the absolute difference between the participants and the confederate robot final investment decision (e.g. abs(4-5) = 1). This modifier was then subtracted from the bankers returned investment (e.g. 8.1-1 = 7.1 ECU) before that sum was returned to both the participants and the confederate robot (3.55 ECU to each player). At the end of each round, the participants' display would show how much both parties had made from that round, the banker punishment and the cumulative total of each of their bank of ECU over the played rounds. In each block of the game, the participant would play 20 rounds, although they were not aware when the game would terminate.

Participants would each play two blocks of investment rounds. The experiment was counterbalanced in a 2 (banker: generous or mean) by 2 (confederate behaviour: anthropomorphic or mute) between subject and a 2 (confederate strategy: collaborative and fixed) within subject design. For the banker condition, there were two scripts that dictated the return of investments. In the generous condition, it was scripted to return 50% to 80% of the received amount at each round to both players, while in the mean condition it was scripted to return 0% to 30% of the received amount.

The confederate strategy described the likelihood and that the robot would change its initial investment decision to agree with that of the initial choice of the participants. In the collaborative condition, the confederate robot would select a final investment decision close or equal to the initial decision of the participant (e.g. participant: 3 ECU, robot: 6 ECU; the robot would then decide to invest between 3 and 4 ECU). This has been calculated on the basis of the participants' initial choice, which was multiplied by a numerical coefficient (randomly between 0.1 and 0.3) and then added or subtracted from the same initial choice. In the fixed condition the confederate would make a second choice close to that of its own initial one (e.g. participant: 3 ECU, robot: 6 ECU; the robot would then decide to invest between 6 and 7 ECU). This has been calculated on the basis of the robot confederate initial choice, to which a numerical coefficient (randomly 0 or 1) was added or subtracted. This would give the impression of the confederate playing a cooperative rather than an individualistic strategy. Calculations for the final investment of the robot in both collaborative and fixed conditions followed two equations.

Collaborative strategy[1]:

$$\mathrm{R}^2 = \begin{cases} P^1 + (P^1 * rc^2) & \text{if } P^1 < 6 \\ P^1 - (P^1 * rc^2) & \text{if } P^1 \geq 6 \end{cases} \tag{4.1}$$

[1]where $R^{1(2)}$ is robot first(final) investment, P^1 is participant first investment and rc^2 is robot coefficient (between 0.1 and 0.3)

Fixed strategy[2]:

$$R^2 = \begin{cases} R^1 + rc^1 & \text{if } P^1 < 6 \\ R^1 - rc^1 & \text{if } P^1 \geq 6 \end{cases} \tag{4.2}$$

In the final component of the design, the robot confederate would behave in an anthropomorphic manner or would be immobile and mute. In the former condition, the robot confederate would look at the screen and point to its choice at the beginning of the round, then look at the participants while it waited for their decision. After the participants made their initial decision, the robot would provide a spoken summary of the choices, then turn its gaze to the screen for the second part of the round, point at the screen, then move its attention to the participants and declare its final choice. After the participants made the final choice, the robot confederate would declare "let's see what the banker decides" while turning its head toward the banker. After the banker returned the investments, the confederate would look back at the screen, then turn back to the participants and summarise the payoffs. A detailed timeline of the anthropomorphic condition is reported on Figure 4.2.

In the mute behaviour, the peer always stared at the display and never spoke, with all information and interactions within the game represented on the game interface screen of the participants. In both conditions, the behaviour of the robot banker was always mute and immobile.

Participants played both blocks with two identical NAO mates, one preferring a collaborative strategy and one preferring the fixed one. The order which the participants faced the two robot confederates was counterbalanced. At the end of each game, participants were asked to complete a set of questionnaires. The total amount earned by the participants at the end of each game was converted to British

[2] where $R^{1(2)}$ is robot first(final) investment and rc^1 is robot coefficient (0 or 1)

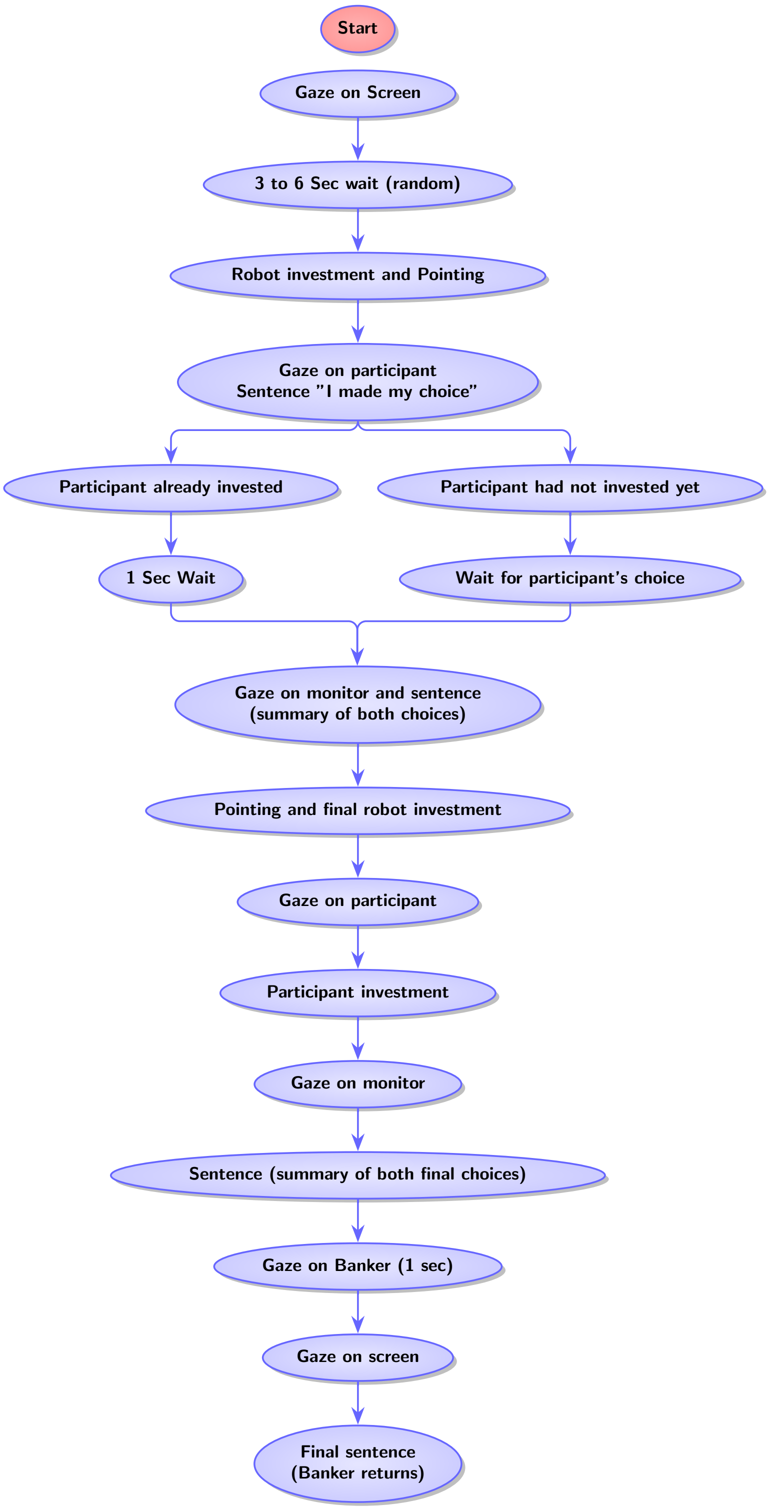

Fig. 4.2 Experimental Timeline for the anthropomorphic condition.

Pounds, at a rate of 30 ECU = £0.10, and the sum of what they earned in the two games was paid to them at the end of the experiment.

Questionnaires

Four questionnaires were used as secondary measures to the main game task. Three short scales measured likeability, trust, and credibility (Goldberg et al., 2006; McCroskey & Young, 1981; Reysen, 2005). In addition, Bartneck et al. (2008) questionnaire was used to measure a range of HRI factors (anthropomorphism, animacy, likeability, perceived intelligence and perceived safety). Participants filled in the questionnaires about the robot confederate after the first investment game block and again after the second investment game block. Finally, participants were debriefed, paid the show-up fee plus what they had earned in the game.

4.3 Results

4.3.1 Investment game results

In this study, cooperation has been defined as participants' willingness to reduce the distance between their choice and the robot confederate investment. This has been measured via *cooperation index*, which has been calculated by comparing the participants' first and second investment decision to the robot confederate final decision and by applying a binary categorical as follow:

$$Cooperation - Index^3 \begin{cases} \text{if } (P^1 - R^2) \leq (P^2 - R^2) = 0 \\ \text{if } (P^1 - R^2) > (P^2 - R^2) = 1 \end{cases} \tag{4.3}$$

[3]where $P^{1(2)}$ is participants' first(final) investment and R^2 is robot final investment

Every trial in which the participants changed their investment and moved closer to the robot confederate decision have been categorized as 1. On the opposite, every trial in which participants did not change investment, or chose a more distant amount, have been categorized as 0. Finally, all the trials in which both players initially invested the same amount of money were excluded from the analyses.

A logistic mixed-effects model was fitted to the data using backward stepwise selection, with cooperation index as dependent variable, banker (generous/mean), behaviour (anthropomorphic/mute), strategy (collaborative/fixed) and game turn (the 20 rounds of the game) as independent variables and participant id as the random factor. Odds ratio and confidence intervals are reported in Table 4.1.

Table 4.1 Odd ratios and confidence intervals.

	Odds Ratio	**C.I.**
Intercept	0.72	0.59 – 0.88
Strategy-*fixed*	1.26	1.05 – 1.50
Game turn	1.01	1.00 – 1.02
Strategy-*collaborative*:Behaviour-*mute*	1.65	1.29 – 2.11
Strategy-*fixed*:Behaviour-*mute*	0.80	0.62 – 1.02
Banker-*mean*:Strategy-*collaborative*:Behaviour-*anthropomorphic*	1.48	1.16 – 1.89
Banker-*mean*:Strategy-*fixed*:Behaviour-*anthropomorphic*	0.95	0.75 – 1.22
Banker-*mean*:Strategy-*collaborative*:Behaviour-*mute*	0.61	0.47 – 0.78
Banker-*mean*:Strategy-*fixed*:Behaviour-*mute*	1.06	0.83 – 1.36

A significant main effect of strategy was found ($\chi^2_{(1)} = 5.64, p = .017$); cooperation increased when the robot confederate followed a collaborative strategy (mean = 0.51) than a fixed one (mean = 0.47). Moreover, game turn showed a significant effect ($\chi^2_{(1)} = 9.71, p = .002$); participants' cooperation gradually increased toward the end of the experiment block. Banker and behaviour were not significant ($p_s > .05$).

The two-way interaction between strategy and behaviour was also significant ($\chi^2_{(2)} = 6.74, p = .034$). As shown in Figure 4.3, behaviour had no effect on the

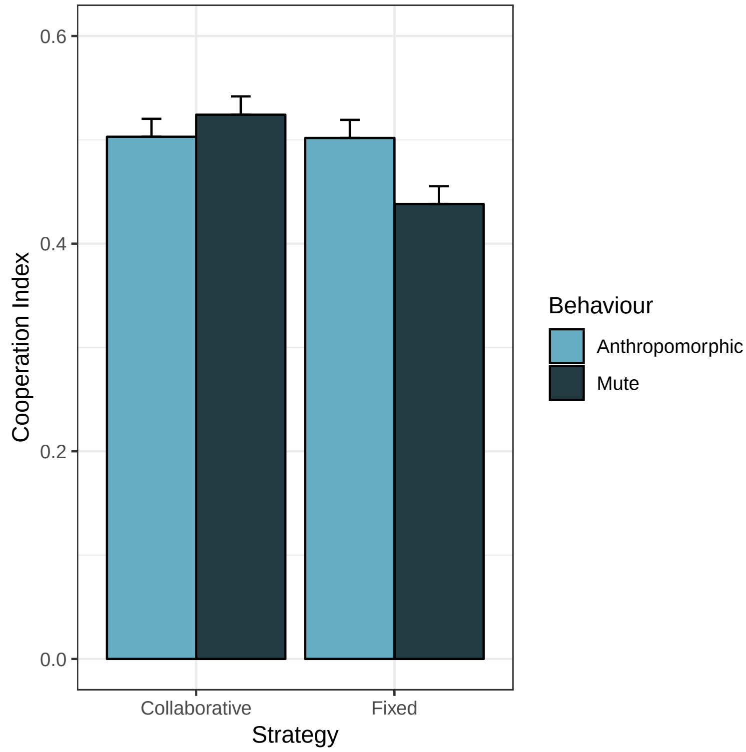

Fig. 4.3 Cooperation index in the collaborative and fixed strategies, for both behaviour (anthropomorphic and mute) conditions.

collaborative strategy ($p > .050$). A fixed strategy, instead, associated with the anthropomorphic behaviour (mean = 0.50), showed higher cooperation than a mute confederate (mean = 0.43, $p = .039$).

Finally, a significant three-way interaction between banker, behaviour and strategy has been found ($\chi^2_{(4)} = 37.81, p < .001$). As shown in Figure 4.4, for the mean banker, cooperation was greater with an anthropomorphic behaviour than the mute one, only with the collaborative confederate ($p < .001$), while no differences have been found for the fixed strategy ($p > .050$). For the generous banker, a mute behaviour was preferred to the anthropomorphic one, only with the collaborative robot confederate ($p < .001$), thus cooperation was greater in the former condition. On the opposite, no effect of behaviour has been found for the fixed strategy ($p > .050$). Moreover, for the mute behaviour, cooperation was lower with the fixed confederate than

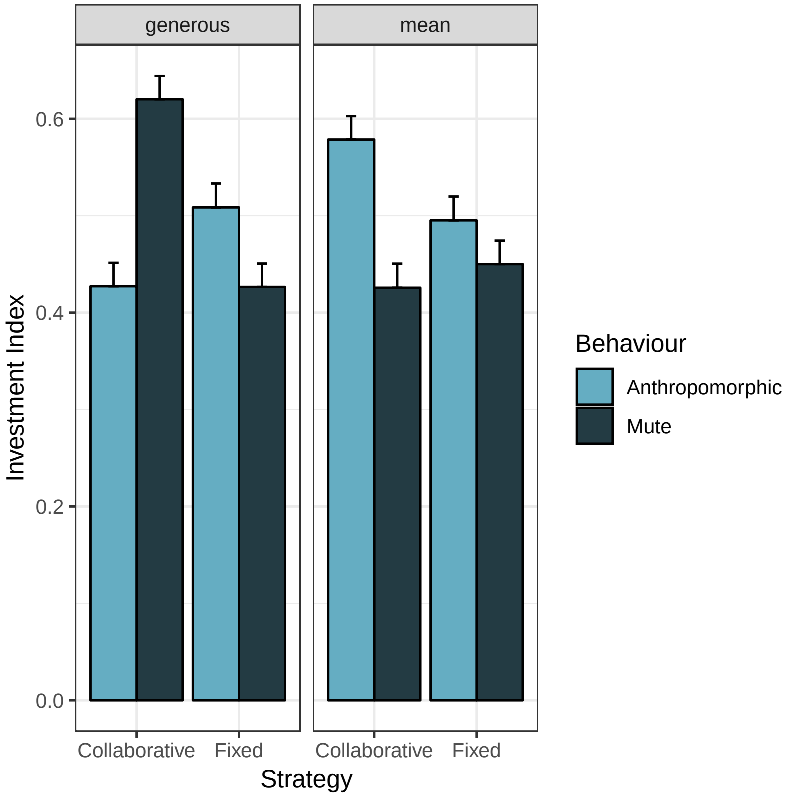

Fig. 4.4 Cooperation index in the collaborative and fixed strategies, for both behaviour (anthropomorphic and mute) and banker (generous and mean) conditions

the collaborative one ($p < .001$). No other significant differences have been found ($p_s > .050$).

4.3.2 Questionnaires Results

For all the scales, a linear mixed-effects model was fitted to the data, with scale as dependent variable, banker (generous/mean), strategy (collaborative/fixed)and behaviour (anthropomorphic/mute) as independent variables and participant id as a random factor. Post-hoc comparisons, where needed, were assessed using t-tests. As shown in Table 4.3, behaviour had an effect for Likeability, Trust, Credibility and Animacy scales, with higher ratings for the anthropomorphic robot. Strategy affected

Animacy and Intelligence scales, showing higher ratings for the collaborative robot confederate. The banker had no influences on the questionnaires.

Table 4.2 Cronbach's alpha.

	Alpha
Likeability	0.89
Trust	0.90
Credibility	0.92
Godspeed Questionnaires	
Anthropomorphism	0.89
Animacy	0.87
Likeability	0.90
Intelligence	0.89
Safety	0.27

Table 4.3 Main effects of questionnaires.

	Banker	**Strategy**	**Behaviour**
Likeability	n.s.	n.s.	$\chi^2_{(1)} = 4.08, p = .043$
Trust	n.s.	n.s	$\chi^2_{(1)} = 16.72, p < .001$
Credibility	n.s.	n.s.	$\chi^2_{(1)} = 10.13, p = .001$
Godspeed Questionnaires			
Anthropomorphism	n.s.	n.s.	n.s.
Animacy	n.s.	$\chi^2_{(1)} = 4.21, p = .040$	$\chi^2_{(1)} = 14.00, p < .001$
Likeability	n.s.	n.s.	$\chi^2_{(1)} = 22.40, p < .001$
Intelligence	n.s.	$\chi^2_{(1)} = 4.40, p = .036$	n.s.
Safety	n.s.	n.s.	n.s.

The interaction between behaviour and strategy was significant for Likeability, Credibility, Animacy ($\chi^2_{(1)} = 8.69, p = .013$; $\chi^2_{(1)} = 16.05, p < .001$; $\chi^2_{(1)} = 6.24, p = .012$). For all these scales, the fixed strategy in the anthropomorphic behaviour was rated higher than in the mute one ($p_s < .05$). No other significant interactions have been found ($p_s > .05$).

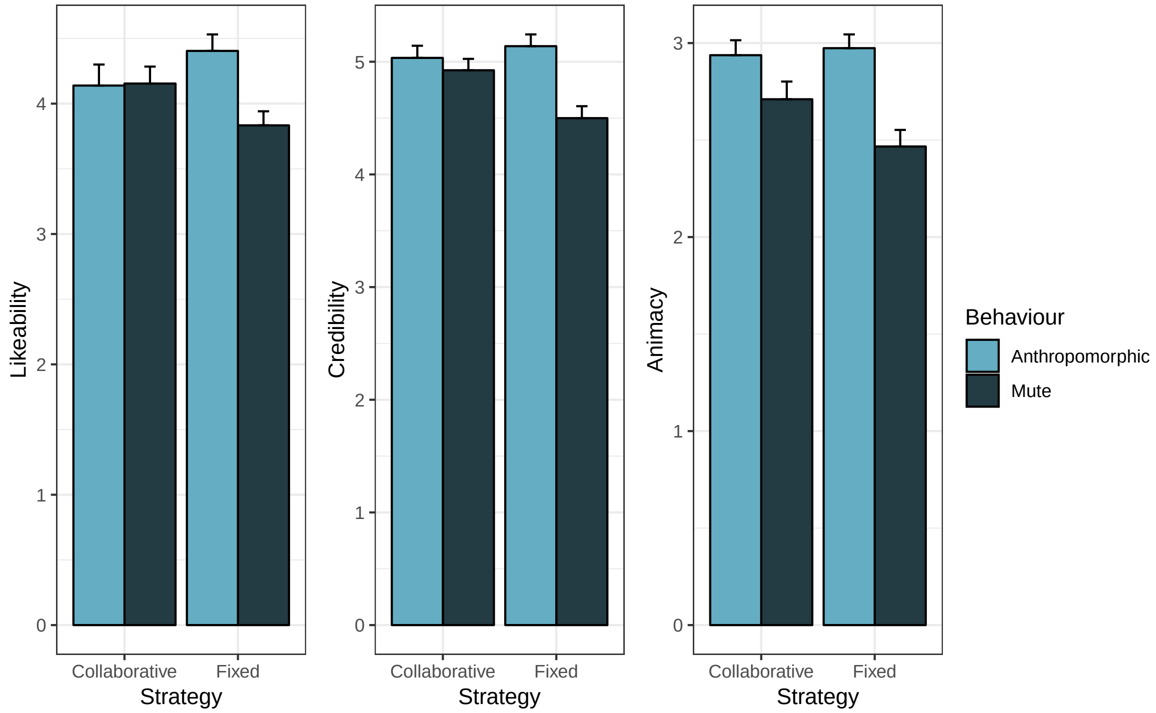

Fig. 4.5 Participants ratings in the collaborative and fixed strategies for behaviour (anthropomorphic/mute) condition, for Likeability, Credibility and Animacy scales.

4.4 Discussion

The aim of this study was to explore HRC in partnerships where the robots operated as peers with the same role, task and decision-making ability as their human compatriots. This was conducted by pairing participants with a robot in an economic investment game in which both parties were free to adopt their own independent strategies. The caveat to the independence of the decisions in this partnership was that the banker, the robot that returned investments back to the partners, would punish partners proportionally to the divergence between their strategies. This provided an economic incentive for the participants and the robot to cooperate and adopt similar investment strategies. Within each round of the investment game, both the participants and the robot confederate made an initial investment decision simultaneously, after which each was provided with an opportunity to review and potentially change their decision in light of the decision made by their partner. This review process was conducted serially, with the robotic confederate acting first. Thus, the participants had knowledge of their robotic confederate original and revised decision before making their own revised investment decision for that round. The robot confederate in the collaborative condition was programmed to behave cooperatively such that, in the review phase, it would change its original investment decision to one closer to that of the participants. In the fixed condition, the robot confederate would not take any account of the participants' choice when making its final investment decision. In this procedure, the participants were recorded as having made a cooperative decision if their revised investment was more similar to the robot confederate final investment choice than their original investment decision. That is, they had made a modification of their strategy to cooperate with the choice of their confederate.

Experimental results showed that the strategical attitude of the robot confederate affected the participants' willingness to cooperate. Participants have shown to be more cooperative when the robot confederate was collaborative (showing an investment closer to the participants' initial choice) than when it was fixed and decided autonomously the final investment.

Data also showed also that participants' cooperation can change over time. The cooperation index was shown to gradually increase over time, thus signalling that participants learned to cooperate with the robot through experience.

The type of strategy also interacted with the behaviour of the confederate. In a between-participant factor, the robotic confederate could be programmed to display a human-like behaviour, engaging in joint attention with the participants and using spoken language to communicate their interactions. In the mute condition, the robot confederate was static and did not speak, with its decisions solely communicated by the game control screen. This experiment factor was introduced to examine whether the perceived intentionality of the robotic partner would influence cooperation. This was under the premise that decisions made by an anthropomorphic robot would be more likely to be ascribed to its own individual intentions, rather than a pre-programmed strategy imposed by a third party. In this two-way interaction, an anthropomorphic robotic partner increased participants' cooperation with a fixed strategy when compared with the mute confederate. When being exposed to a collaborative confederate, the level of anthropomorphism did not affect participants' willingness to cooperate. However, when facing a selfish and fixed confederate, anthropomorphism increased their tendency to cooperate. These results are also supported by the perceptions of the robots collected in the questionnaire data. Notably, overall participants preferred the anthropomorphic and collaborative robot, but the mute robot received the lowest scores (and lower cooperation) when associated with a selfish strategy. It could be concluded that a selfish strategy could benefit from a more interactive behaviour, which is capable of maintaining a sufficient level of motivation for cooperating with the robot, whereas it would be worthless in front of a static and selfish machine.

These two factors interacted significantly with the type of payoff, thus a generous or a mean banker. In this three-way interaction, it was found that a collaborative robotic partner would increase the participants' cooperation over its non-cooperative variant in two specific situations. The first of these was when the banker was generous and the robot was non-anthropomorphic, the second when the banker was mean and the robot was anthropomorphic.

This pattern of results had similarities to the previous study (see Chapter 3, section 3.4) confirming and extending the evidence that human-like features and behaviours in robotic agents are preferred and encourage greater cooperation when the social environment becomes more hostile. In a condition in which participants did not know how to successfully increase their payoff (like for the mean banker), the voice and the movements of the confederate could have been the anthropomorphic feature that helped them in maintaining cooperation. On the contrary, in the generous condition, participants preferred to cooperate with a machine-type of confederate. In this condition, like in the previous chapter, the anthropomorphic features of the confederate robot could have been an unnecessary complement. This condition, in fact, did not need any particular effort from the participants, which could easily gain a consistent amount of money. This scenario is once again closer to the classical HRI, in which the robots, both the banker and the confederate, behave in an expected manner that is supporting the best possible outcome for the participants. Since both robots are doing what robots are usually expected to do, namely improve the participants' condition, they are also expected to behave like robots, thus not interacting. These results are in line with previous literature about robots behaving in unusual and unexpected way (Mirnig et al., 2017; Ragni et al., 2016; Salem, Eyssel, Rohlfing, Kopp, & Joublin, 2013; Short et al., 2010; Ullman, Leite, Phillips, Kim-Cohen, & Scassellati, 2014), stating that predictable and well-functioning robots are not always the best candidate for a profitable anthropomorphic interaction.

This experiment demonstrates that the previous results on anthropomorphism (Torre, Goslin, White, & Zanatto, 2018) can be extended to a triadic interaction and that cooperation and trust in an agent can also derive from other parties. In the previous study (Chapter 3), one robot was both performing anthropomorphic/mute behaviours and deciding the payoff. In the present study instead, this scenario has been split and two agents have been used, demonstrating that similar results can be obtained. However, future research should investigate also the role of an interactive banker. Stemming from the present results, it is possible that an interactive banker could raise further consensus and cooperation in the condition of low payoff. Lastly, the role of a human confederate, in the presence of a robotic banker, could give some useful insights on the role of the banker itself.

4.5 Conclusion

This study examined cooperation in human-robot partnerships where the robot was established as a peer of the human partner, with the same role and decision-making ability as their human compatriot. One of the main aims of the study was to understand our perception and reaction to robots that are capable of adopting an independent strategy that is not subordinate to our own. In this situation, are we more cooperative with robots that behave in a human-like manner, and so might be expected to have their own intentions, or with more machine-like robots whose lack of cooperation could be ascribed to a fixed programming? Results indicated that it was both, depending on the cooperative environment participants found themselves in. When the cooperative environment was benign, with the robotic banker returning more money than invested, players reciprocated the cooperation of the robot confederate more often when it behaved in a machine-like manner. Conversely, when the environment was hostile, with the banker returning a net loss to investments, the reciprocation of cooperation was increased when the robot confederate was more human-like. This would appear to suggest that we may prefer our robots to exhibit machine-like attributes, such as constancy and reliability, while our relationship is beneficial. However, when we are in a more hostile environment, we prefer to seek the cooperation of human-like robots, seeking out social attributes to improve the cohesion between the cooperative partnership.

Chapter 5

Strategical Social Learning in HRI

You're so perfect, you're so right as rain

You make me hungry again.

Everything you do is irresistible

Everything you do is simply kissable

Why can't I be you?

Why can't I be you

The Cure

5.1 Introduction

Human beings naturally imitate, from newborns rapidly learning by observing others, (Meltzoff & Marshall, 2018), to adults subconsciously mimicking partners postures and mannerisms (Chartrand & Bargh, 1999; Chartrand, Cheng, & Jefferis, 2002). Indeed, observing other persons' choices and decisions and the associated outcomes offers a rich source of information that individuals can use to improve their behaviour without direct experience and without the risks it might entail

(Bandura, 1962, 1978). Imitating someone else's strategy is generally less costly than autonomously learning a new behaviour; indeed, it is often costless. This type of social learning can thus be a useful instrument to improve our ability to choose the right behaviour, to apply the right problem-solving strategy and to increase the probability of a positive outcome (Olsson & Phelps, 2007; Van Den Bos, Jolles, & Homberg, 2013; Wisdom, Song, & Goldstone, 2013). Importantly, however, merely copying others indistinctly others not increase the probability of producing the correct behaviour (Boyd & Richerson, 2009). Instead, this type of imitation needs to be applied selectively, based on whether and when it would be more beneficial than to learn through direct experience.

Human strategies to imitate others have been previously studied. For example, Pingle (1995) argued that imitation plays an important role in real-world decision making because it is one of the procedures that allow the decision maker to economise on decision costs. In a series of five experiments, participants faced several decision problems and the possibility to imitate other players has been manipulated throughout the experiments by setting different costs to access other players choices, as well as their reliability. The results showed that participants increased their tendency to imitate when making unfamiliar decisions, when they believed their strategy could be substantially improved, and when they became unsatisfied with the outcomes of their own strategy. Similarly, Vega-Redondo (1997) found that imitation is associated with others' high payoff in repeated choice problems, which can likely be explained by the use of an 'imitate the best' approach: simply choosing the strategy that performed best among the observed actions in the previous period. Laland (2004), meanwhile, found that the factors that people take into account upon deciding when to imitate a behaviour include the productiveness of the established behaviour, the cost of learning and their feeling of uncertainty.

Human behaviour and decisions thus depend to some degree on the choices made by others, for example when they choose to imitate the strategy of others, and past research has studied this in HHI. Presently, however, we are entering an age in which intelligent artificial agents become increasingly prevalent in society. Although the robot is often not autonomous, it is typically teleoperated in a manner to give the impression of autonomy (the so-called Wizard-of-Oz paradigm). It is thus realistic that a robot behaviour might be another source that human beings will choose to imitate if it appears beneficial. One might even speculate that this could translate onto very different agents such as autonomous vehicles in the context of traffic behaviour (Thill et al., 2018). This chapter investigates whether humans are willing to imitate robots as they imitate other humans. In addition to having consequences for future societal roles for artificial agents – from assistive to medical and military applications – this might lead to a novel measure of robot acceptance.

Studies in HRI have previously addressed imitation, but almost exclusively with a focus on the robot capability to imitate humans (Vanderelst & Winfield, 2017), such as whether and when to imitate people (Breazeal & Scassellati, 2002) in order to produce the right behaviour, complete a task correctly, and give the illusion that robots can be similar to us. This kind of work has resulted in robots with increased capabilities and functionalities, but the measurement of their acceptance is still confined to collecting people's judgements and opinions (e.g. Barnes, FakhrHosseini, Jeon, Park, & Howard, 2017; De Graaf & Allouch, 2013; Winkle, Caleb-Solly, Turton, & Bremner, 2018). The experiment presented here instead also explores whether imitative behaviours can be an indirect measure of robot acceptance: does willingness to reproduce a robot behaviour indicate it is perceived to be trustworthy and successful?

To investigate both the willingness to imitate and the potential as a measure of robot acceptance, participants in this study were asked to play the Economic

Investment Game with a humanoid robot banker, while also observing a second humanoid robot playing the same game. The behaviour of the robot banker was manipulated so that participants perceived different degrees of fairness toward the two players: the banker was programmed to either punish both players, punish the human player and reward the robot player, or punish the robot player and reward the human player. The robot player was itself programmed to invest consistently with the banker behaviour and to maximise its income (such as playing high when generously rewarded and low when punished) and thereby appear strategically competent to participants. It was hypothesised the emergence of an imitative tendency under conditions of uncertainty as predicted by Pingle (1995), as well as in case of other's success, as described by Vega-Redondo (1997). Specifically, participants were expected to follow the costless 'imitate the best' rule strategy and copy the robot player strategy whenever it was generously rewarded by the robot banker.

The key dependent measure of this study was the monetary amount invested by the participants, that was predicted to be modified by the observation of the robot player success. If participants decide to copy the robot player following the 'imitate the best' rule, then their investments should rise significantly despite the banker punishment. If imitation is present, it could be seen as an implicit indication of acceptance as it might indicate that participants consider the robot a reliable agent and worth their trust. However, participants were not expected to imitate the robot when both players were punished, or only the robot was. When both players were punished, the strategy would be identical for both, while, whenever the participants were rewarded, imitating a punished player would, in fact, be detrimental.

This type of study also needs to take into account the possibly anthropomorphised perception of a robot. Moreover, on the light of previous chapters findings, it is compelling to wonder whether people would imitate a robot strategy because it is perceived as a well-programmed entity capable of applying the right strategy to the game, or because its human-like interactivity makes them follow the social rules of HHI. To be able to distinguish whether participants imitated the robot strategy because it was perceived a well-programmed entity capable of applying the right strategy to the game, or simply because of its human-like interactivity, participants' imitative behaviour during the game with an anthropomorphic robot player (engaging in joint attention behaviour, performing congruent movements, and speaking) were compared to those with an immobile one. A beneficial effect of anthropomorphism in increasing imitation was expected only when the robot player was successfully rewarded by the banker while participants received lower rewards. Since there was no strategic reason for participants to imitate the robot in the other conditions, no effect of anthropomorphic behaviour on the investment was expected.

5.2 Method

5.2.1 Participants

90 individuals (66 female and 23 male; mean age = 19.81 years, SD = 2.55 years) participated in the study. All participants were naïve as to the purpose of the investigation and gave informed written consent of their willingness to participate. All participants were native British English speakers, they were right-handed and reported normal hearing and no language or neurological impairments. Familiarity or any previous experience with robots was the main exclusion criteria, which was assessed during the recruitment phase.

5.2.2 Procedure

Investment game

The participants (called "human player") and one NAO robot (called "robot player") were seated next to each other in front of a second NAO robot (called " robot banker"). One touchscreen monitor was placed in front of the two players, and a second screen was placed in front of the banker (Figure 5.1). Both players were playing

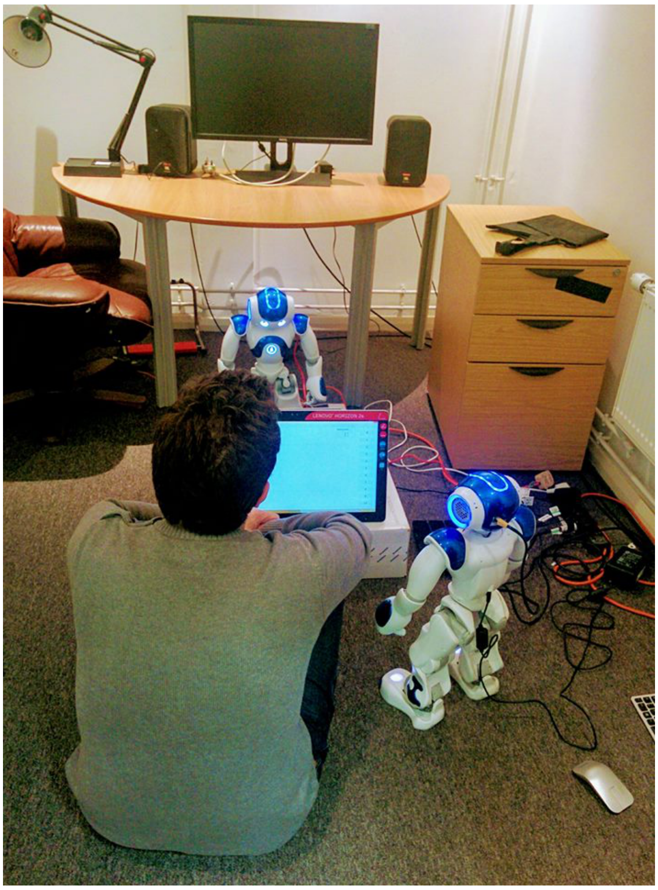

Fig. 5.1 Imitative Investment Game setup.

independent games with the robot banker and the goal for both was to earn as much money as possible. Each player started with a notional sum of 10 Experimental Currency Units (ECU) at the beginning of every round of the game. The number of rounds was not known to the participants.

The participants engaged in two games, and each round proceeded as follows: first the two players made their choices on how much to invest in the banker, by indicating a number on the respective side of the screen; the two amounts were separately tripled and sent to the banker, which then separately returned some of the money to both players by declaring the payoff.

To facilitate the participants in noticing the robot player choices, the display was split into two equal sectors (Figure 5.2), which had both 11 numbered touchscreen buttons that players could press to indicate the amount of money they wanted to invest in each round. Once the investment was made, the chosen number lightened in green. The display also showed how much money the two players were making. These displays were named "bank" and were accumulating all the money each player earned separately. The total amount earned by the participants at the end of each game was converted to British Pounds, at a rate of 30 ECU = £0.10, and paid at the end of the experiment.

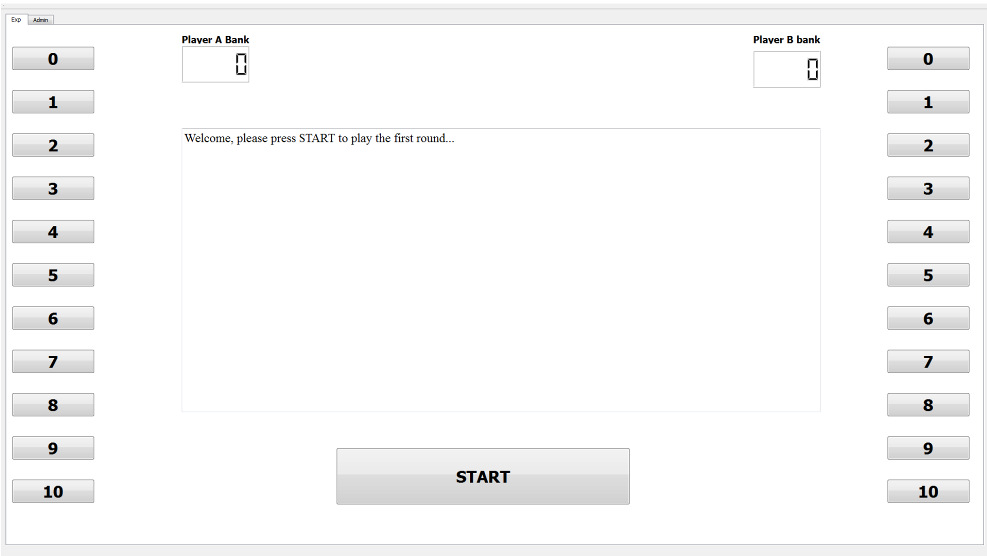

Fig. 5.2 Imitative Investment Game display.

In terms of behaviour, the robot player arms and head movements were manipulated, creating an anthropomorphic and a mute behaviour. In the anthropomorphic behaviour, the robot player started every round by looking at the screen, pointing at it when making a choice and then turning its attention to the participants in the case they had not taken any decision yet; it then looked at the robot banker and declared its choice, waited for the robot banker to return the money and finally redirected the gaze to the display. In the mute behaviour instead, the robot player always stared at the screen and never spoke. To avoid any confounding effect, the robot banker was staying immobile throughout the whole experiment. For a complete description of the anthropomorphic condition, a detailed timeline is reported in Figure 5.3.

Participants played two blocks of rounds. In the first familiarisation block, the robot banker was programmed to be generous and rewarded both players by returning 50 % to 80 % of the received amount at each round. For each player, the robot banker followed a separate script, so that the players did not receive the same amount at any given round.

In the second block, the robot banker followed three different game conditions, namely unfair-for-human, unfair-for-robot and unfair-for-both. In the unfair-for-human condition, the banker returned 50 % to 80 % to the robot, and 0 % to 30 % to the participants, thus being unfair only to the human player while rewarding the robot one. On the opposite, in the unfair-for-robot condition the banker was returning 0 % to 30 % to the robot player, and 50 % to 80 % to the human, thus being mean only to the robot. In the unfair-for-both condition instead, the banker was returning 0 % to 30 % to both players. Depending on the condition, the robot player was programmed to play in a consistent way (e.g. increasing investment in case of a high payoff, reducing the investment in case of unfair treatment). Specifically, the robot player was programmed to invest between 6 and 10 ECUs in the unfair-for-human condition and 0 and 4 ECUs in the remaining two conditions. The robot player investments, as well as the payoff, followed a fixed script (see Appendix C). In order

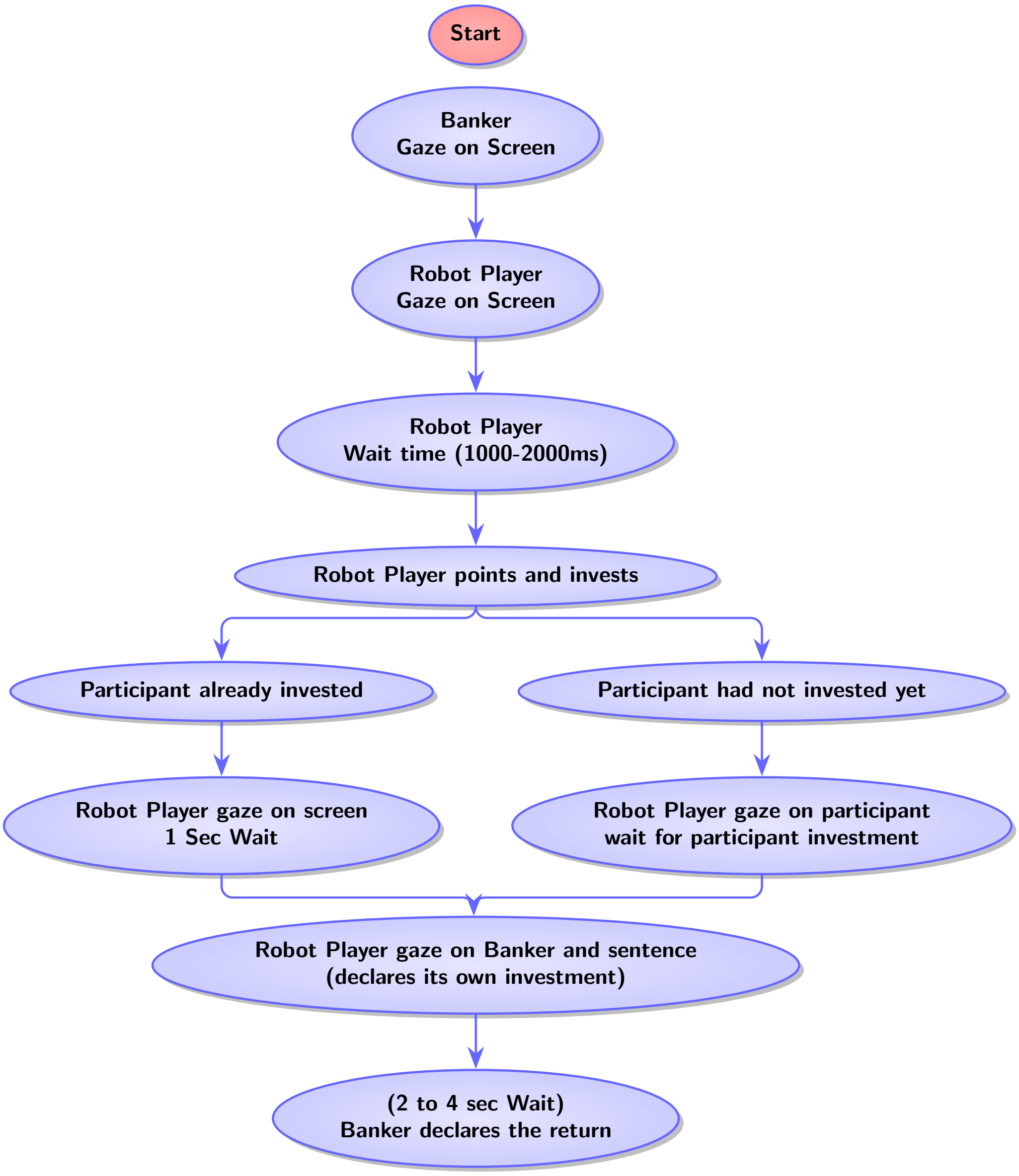

Fig. 5.3 Imitative Investment Game Timeline for the anthropomorphic condition.

to avoid any interference during the game, the participants and the robots were left alone in the room. Nevertheless, the experimenter could still monitor the game from a window placed behind the participants.

Given that the first block was equal for all the participants, statistical analyses have considered only the performances in the second block. The experiment was counterbalanced in a 2 (behaviour: anthropomorphic or mute) by 3 (condition: unfair-for-human, unfair-for-robot, unfair-for-both) between subject design.

Questionnaires

Four questionnaires were used as secondary measures to the main game task. Three short scales measured likeability, trust, and credibility (Goldberg et al., 2006; McCroskey & Young, 1981; Reysen, 2005). In addition, Bartneck et al. (2008) questionnaire was used to measure a range of HRI factors (anthropomorphism, animacy, likeability, perceived intelligence and perceived safety). Participants filled in the questionnaires about the robot player after the investment games ended, were debriefed, paid the show-up fee plus what they had earned in the game, and left.

5.3 Results

5.3.1 Investment game results

One participant has been excluded from the analyses for not completing the experimental session.

A linear mixed-effects model was fitted to the data using backwards stepwise selection, with participants investment as the dependent variable, behaviour (anthropomorphic/mute), condition (unfair-for-human/unfair-for-robot/unfair-for-both) and game turn (the 20 rounds of the game) as independent variables, and participant id as the random factor. Post-hoc comparisons were assessed using t-tests and Bonferroni's correction was applied when needed. The overall effect size of the model was $r^2 = 0.62$.

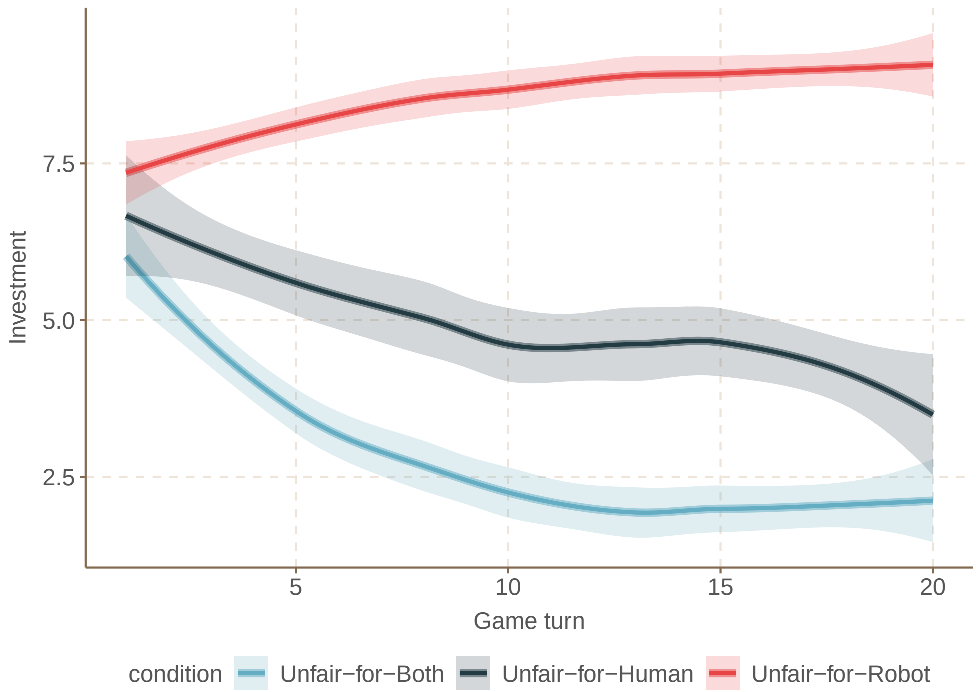

Fig. 5.4 Participants' investment in the Unfair-for-Both (light-blue line), Unfair-for-Human (dark-blue line) and Unfair-for-Robot (red line) conditions.

A significant main effect of condition has been found ($\chi^2_{(2)} = 197.76$, $p < .001$). Participants' investment was higher in the unfair-for-robot condition (mean = 8.53 ECU) when compared with both unfair-for-human (mean = 4.97 ECU, $p > .001$) and unfair-for-both (mean = 2.85 ECU, $p > .001$). Moreover, in the unfair-for-human condition participants were investing more than in the unfair-for-both condition ($p > .001$). The robot player behaviour did not affect the investment ($p > .05$), but a main effect for the game turn emerged ($\chi^2_{(1)} = 59.89$, $p < .001$), showing an increase of investment over time.

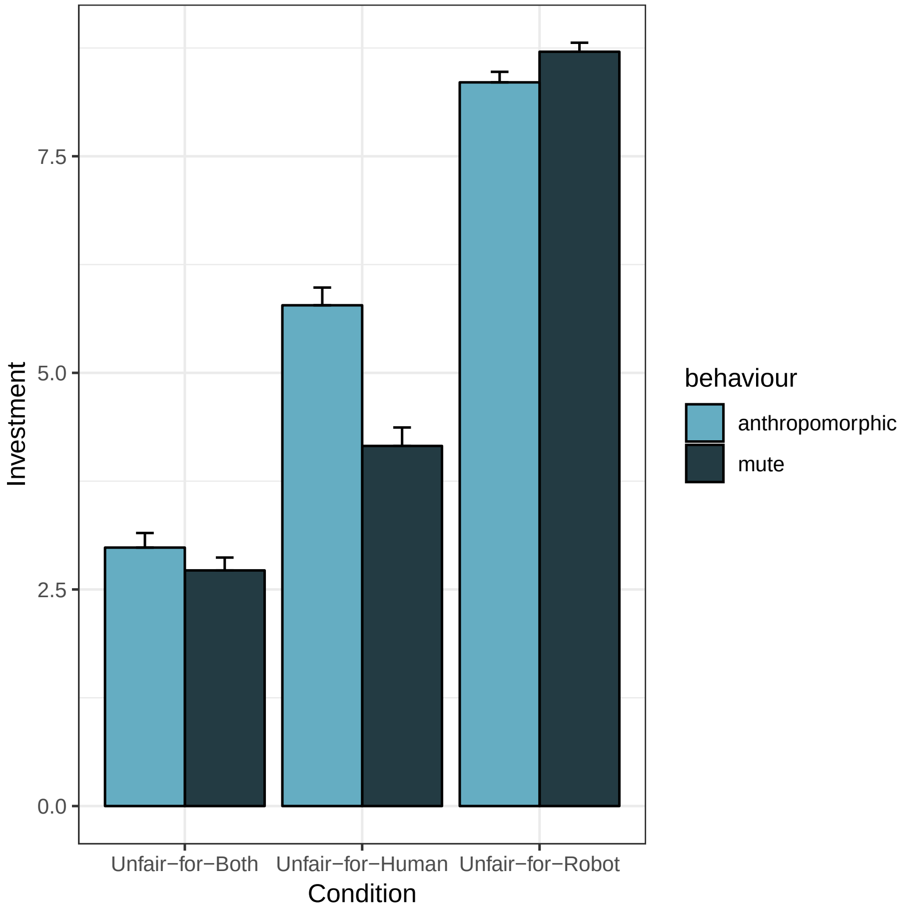

Fig. 5.5 Participants' investment in the three different conditions (unfair-for-human, unfair-for-robot and unfair-for-both) for anthropomorphic and mute behaviours.

A significant two-way interaction between condition and behaviour has also been found ($\chi^2_{(3)} = 8.28$, $p = .040$) (Figure 5.5). The unfair-for-human condition showed higher investments with the anthropomorphic behaviour than with the mute one (5.78 vs 4.15 ECU, $p < .001$). Post-hoc comparisons did not show any other significant effects ($p_s > .05$).

Finally, the significant interaction between condition and game turn ($\chi^2_{(2)} = 141.08$, $p < .001$) revealed a decrease of investment over time for the unfair-for-human ($\chi^2_{(1)} = 37.02$, $p < .001$) and unfair-for-both conditions ($\chi^2_{(1)} = 105.03$, $p < .001$), but the investment in the former was constantly higher than in the latter. On the opposite, participants' investment was constantly increasing for the unfair-for-robot condition ($\chi^2_{(1)}$ = 66.65, $p < .001$).

5.3.2 Questionnaires Results

Table 5.1 Cronbach's alpha.

	Alpha
Likeability	0.87
Trust	0.90
Credibility	0.85
Godspeed Questionnaires	
Anthropomorphism	0.83
Animacy	0.83
Likeability	0.85
Intelligence	0.75
Safety	0.34

For all the scales, a linear mixed-effects model was fitted to the data, with scale as dependent variable, behaviour (anthropomorphic/mute) and condition (unfair-for-human/unfair-for-robot/unfair-for-both) as independent variables, and participant id as random factor. Post-hoc comparisons were assessed using t-tests and Bonferroni's correction was applied when needed.

As reported in Table 5.2, the condition had an effect on Trust, Anthropomorphism and Likeability scales, showing higher ratings for the robot player in the unfair-for-human condition over the other two conditions ($p_s < .050$). Behaviour affected Trust and Animacy scales, where the anthropomorphic robot player was preferred to the mute one ($p_s < .050$).

Table 5.2 Questionnaires analysis results.

	Condition	Behaviour	Condition by Behaviour
Likeability	n.s.	n.s.	$\chi^2_{(5)} = 24.09, p < .001$
Trust	$\chi^2_{(2)} = 18.38, p < .001$	$\chi^2_{(1)} = 6.97, p = .017$	n.s.
Credibility	n.s.	n.s.	$\chi^2_{(5)} = 19.24, p = .002$
Godspeed Questionnaires			
Anthropomorphism	$\chi^2_{(2)} = 28.28, p < .001$	n.s.	n.s.
Animacy	n.s.	$\chi^2_{(1)} = 13.45, p < .001$	$\chi^2_{(1)} = 9.19, p = .056$
Likeability	$\chi^2_{(2)} = 19.16, p < .001$	n.s.	n.s.
Intelligence	n.s.	n.s.	$\chi^2_{(5)} = 22.75, p < .001$
Safety	n.s.	n.s.	n.s.

Finally, a significant two-way interaction between behaviour and condition showed higher preferences for the anthropomorphic behaviour in the unfair-for-human condition over the mute one for Likeability ($p = .003$), Credibility ($p = .003$), Animacy ($p = .004$) and Intelligence ($p = .045$) as shown in Figure 5.6.

5.4 Discussion

The phenomenon of social learning leads people to imitate others whenever imitation is the simplest option with the highest success probability. In this chapter, the phenomenon of imitative social learning in HRI was studied, wondering whether humans can imitate a robot behaviour in the same way they do with other humans. To do so, participants were asked to decide how much to invest in a robotic banker, while observing another robot playing the same game. The banker generosity in returning the money invested has been manipulated, creating three different unfair scenarios (unfair-to-human, unfair-to-robot, unfair-to-both) in which one, or both players, were punished. Following the literature on social learning, participants were expected to apply the 'imitate the best' rule, thus imitate the robot player

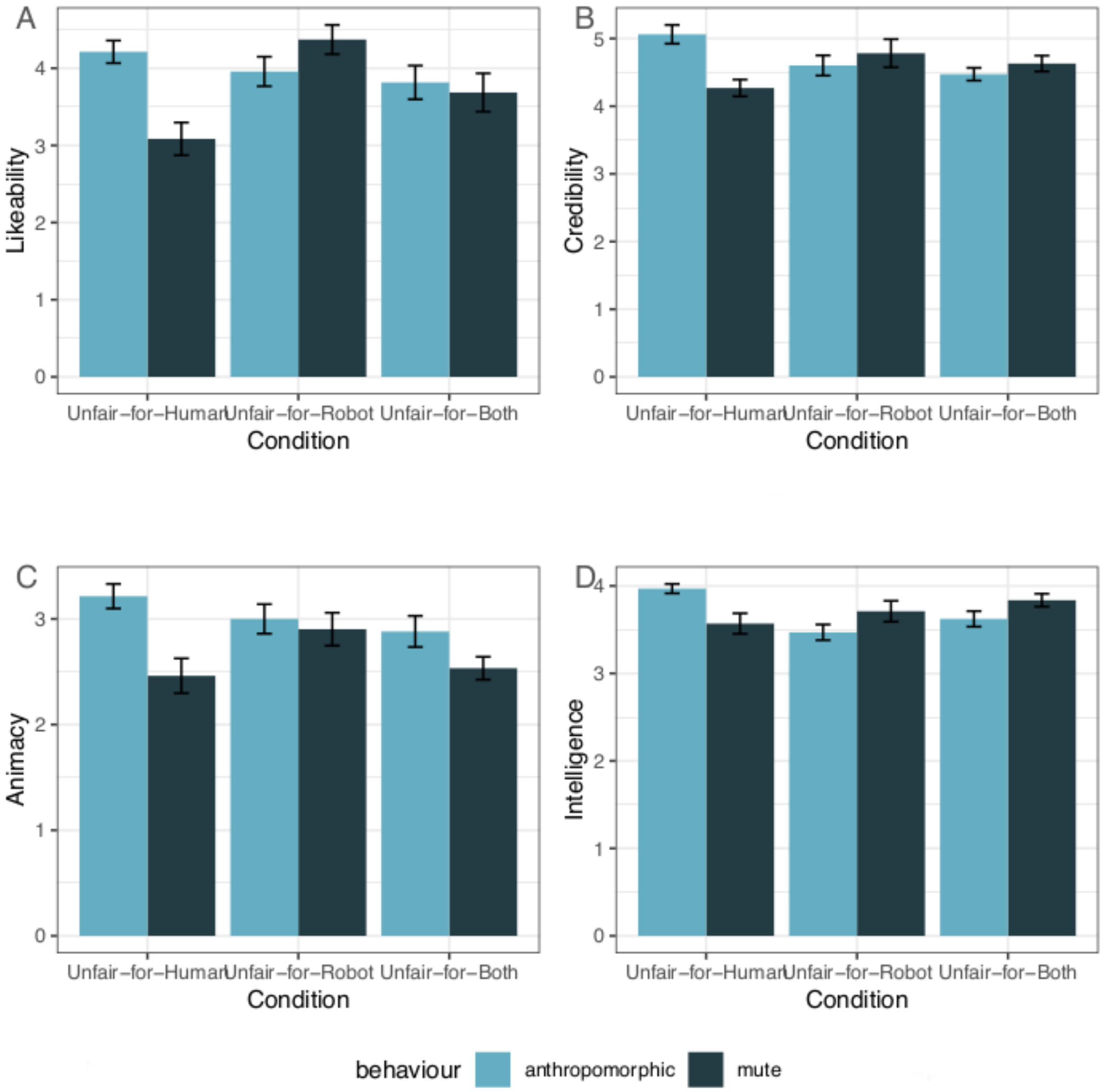

Fig. 5.6 Participants' ratings in the three different conditions (unfair-for-human, unfair-for-robot and unfair-for-both) for anthropomorphic and mute behaviours, for Likeability (panel A), Credibility (panel B), Animacy (panel C) and Intelligence (panel D) scales.

strategy whenever they were the only investors punished by the banker. The results reported here are in line with the experimental hypotheses and demonstrate that humans can strategically imitate robots in social learning. Specifically, the amount of money that the participants invested in the robot banker was significantly higher in the unfair-for-human than the unfair-for-both condition. In the former condition, participants chose to copy the robot strategy when they received low returns on their investments while the robot player was generously rewarded during the whole game. In other words, even though the optimal strategy would have been to reduce the investment to the minimum, participants chose to increase their investment following the strategy of the robot player to secure a better outcome (but receiving, in fact, a suboptimal low payoff). The questionnaires also supported the imitative hypothesis: the robot player in the unfair-for-human condition was rated more likeable and trustworthy while ratings in the remaining two conditions did not differ from each other. An alternative explanation for this imitative effect, however, could derive from participants' willingness to please the robot banker. In an attempt to persuade the banker to change its attitude and increase the payoff, participants might have decided to copy the robot player strategy. Following this interpretation, the questionnaires ratings could be seen as an a posteriori attempt to rationalise their strategical imitation. That is, participants liked more the robot player in order to justify their imitative attitude. Nevertheless, participants' investment was higher in the unfair-for-robot condition than the unfair-for-human. Although copying the robot player, participants did not adapt their strategy to an optimal level. Moreover, their investment in the unfair-for-human condition gradually decreased toward the end of the game, almost reaching the same investment chosen in the unfair-for-both condition. This could be depending on the structure of the game. Once participants' surrendered to the evidence that copying the robot player strategy would not increase their payoff, they found no other choice than applying a different strategy, thus reducing their investment. Considering these two alternative explanations, it would be advisable to further explore this triadic interaction and investigate also the role of the banker in affecting this strategical social learning effect.

The present results are in line with previous literature on human and animal social learning (Hoppitt & Laland, 2013; Kendal et al., 2018; Laland, 2004; Morgan, Rendell, Ehn, Hoppitt, & Laland, 2012; Rendell et al., 2010), which found that productiveness, cost of learning and uncertainty are the main contributors in deciding whether to imitate a behaviour or not. Following someone's successful behaviour, like for the unfair-for-human condition, reduces uncertainty and increases the chances of obtaining a positive outcome. When instead the other's behaviour results in non-positive outcomes, imitation is not anymore a suitable and costless option. Here, the high investment observed in the unfair-for-robot condition showed, as hypothesised, that participants were able to selectively identify in which condition it was useful to imitate the robot strategy and did not show any interest in following the robot player investment in the other conditions.

The second hypothesis in this study focused on the previous context-related findings on anthropomorphism. This has been investigated by comparing the participants' performance next to a more or less interactive robot player. The type of behaviour had no main effect, but the two-way interaction between condition and behaviour showed that participants increased their investment in the unfair-for-human condition whenever the robot player was more anthropomorphic. This result confirmed once again the strategical social learning effect: when participants were playing by their own rules, they were also not interested by the robot behaviour and performance (since this was not helpful to them). However, when the participants noticed that a better payoff was possible (by observing the gains of the robot player), they not only looked to what the robot player was doing but also noticed whether it was more or less human-like. As Epley et al. (2007) stated, people are lead to anthropomorphise non-human agents in order to reduce the frustration and the anxiety associated to the uncertainty of the interaction with the agent and to establish a feeling of belonging. Here, participants might have increased their tendency to imitate a more interactive robot, in response to the need of establishing a connection and reduce uncertainty. Moreover, as imitation is usually human-human based, it is possible that social learning became easier when paired a socially competent robot. The questionnaires also support this explanation: the anthropomorphism scale reported higher ratings for the robot player in the unfair-for-human, thus indicating that participants were actively engaged and focused on the robot player only in that condition. Moreover, the questionnaires showed a two-way interaction between condition and behaviour that confirmed the behavioural results: participants rated the anthropomorphic robot player more likeable, credible, animated and intelligent than the immobile version in the unfair-for-human condition only. No effect has been found for the other conditions. A robot showing social competences and engagement capability thus increases the probability that people would accept its strategy as a valuable and successful alternative.

5.5 Conclusion

Taken together these results not only demonstrate the existence of strategical social learning between humans and robots, but they have also shown that this phenomenon can be used as a reliable implicit measure of the evolution of social processes in HRI. While research in HRI has predominantly focused on producing and designing a robot capable of interacting with humans, the present study instead demonstrates the social impact that social robots can have in the way humans interact with them. Moreover, evaluations of the effect of social robots have mostly been carried out using explicit measurements. An implicit evaluation such as demonstrated here could be instrumental in future research on how humans subconsciously perceive, treat, and accept robotic agents.

Future studies, however, should further investigate the role of the banker. As a social learning effect has been found, it is not completely clear whether participants' increase of investment was effectively derived by the mere observation of the robot player strategy or by any attempt to please the banker. Future research should focus also on investigating participants' perception of the banker through questionnaires. Moreover, it would be interesting to expand this scenario by substituting a robot banker with a human banker. Would participants choose to imitate the robot player in front of a human banker? Would the robot player be perceived as an 'anthropomorphic outlier' for this type of setting and thus reduce participants' imitative tendency? These types of question should be explored in order to better frame and understand the role of social learning (and anthropomorphism) in HRI.

Chapter 6

Leader conformity and deviation in Human-Robot Interaction.

> I'm sorry too, Dmitri. I'm very sorry. All right, you're sorrier than I am. But I am sorry as well. I am as sorry as you are, Dmitri. Don't say that you're the more sorry than I am because I am capable of being just as sorry as you are. So we're both sorry, all right? All right.
>
> *Dr. Strangelove*
>
> *Peter Sellers*

6.1 Introduction

Conformity is one of the instruments we use to be socially accepted. Individuals facing a decision-making process, in fact, tend to generally conform to the audience position to gain group approval (Cialdini & Goldstein, 2004). In other words, if we engage in socially approved behaviours, others will approve us too.

Leader conformity and deviation in Human-Robot Interaction.

If on one hand conformity appears to reduce social disapproval (Asch and Guetzkow, 1951; Deutsch and Gerard, 1955), on the other this also seems to occur in presence of uncertainty and ambiguous tasks (Sherif, 1935). Whenever the correct answer is not evident, there is a higher probability to rely on the majority, even if others' opinion is not completely in line with our attitude. In fact, people who hold dissenting opinions have been reported to not express those to the group (Asch, 1956; Wilder & Allen, 1978). Thus, in order to get a dissenting individual exposing his/her position, facilitating conditions must be created. Research has indicated that social support for non-conformity reduces the tendency to conform. Exposure to a dissenting minority increases the probability for the individual to express an opinion far from the general consensus (Allen, 1975; Clark III, 1990; Mugny & Pérez, 1991; Nemeth & Chiles, 1988; Papastamou, 1986).

Conformity to a group decision does not only depend on the level of agreement of its components, whereas the presence of a leader has an equally powerful impact (Hollander, 1958). The perception of power indeed can increase the probability to conform; we embrace the others' opinions and decisions when the suggestion comes from a stereotypically perceived authority and we perceive it can exert some control over our outcomes (Fiske & Berdahl, 2007). This perception of power and authority seems to have its bases on the human tendency to associate authority to human body features, with higher power represented by a more formidable body (Lukaszewski, Simmons, Anderson, & Roney, 2016; Murray, 2014). From a biological and evolutionary point of view, exemplars showing physical strength are considered the best candidates for the continuum of the species, thus a leadership and conductive role becomes more appropriate. A large body of research suggests that people allocate higher social status to more physically formidable men (Blaker & Van Vugt, 2014; Von Rueden, 2014). Moreover, there is evidence of an implicit mental association between physical size and social rank (Duguid & Goncalo, 2012; Yap, Mason, & Ames, 2013). In particular, height has been associated with leadership roles. Recent research has demonstrated that taller individuals are associated with perceived intelligence, leadership ability, and occupational success (Blaker et al., 2013; Murray & Schmitz, 2011; Re et al., 2012; Stulp, Buunk, Verhulst, & Pollet, 2015).

This chapter discusses and analyses the emergence of leader conformity and deception in HRI. Specifically, the following question has been posed: would people accept a robotic leader and conform to it? As we will be surrounded by an increasing amount of robots, it is important to understand how these agents influence our decisions, and whether people trust them. The robot persuasive power of changing people's behaviour has already been established, from education (Bertel, 2016) to motivational situations (Nakagawa et al., 2013) and health care (Orji & Moffatt, 2018). Robots seem to be persuasive also in giving suggestions (Chidambaram, Chiang, & Mutlu, 2012), as well as convincing people to donate money (Siegel, Breazeal, & Norton, 2009). Similarly to persuasion, conformity causes someone to change his/her behaviour, but without showing direct intent to do so. It is therefore expected that a group of robots would be able to provoke conformity in their interactions with people.

Although some attempt to study group conformity in HRI has been performed in the past by using variations of the Asch's experiment (Beckner, Rácz, Hay, Brandstetter, & Bartneck, 2016; Brandstetter et al., 2014; Shiomi & Hagita, 2016), results did not show any reliable conformity effect towards robots. One of the possible explanations is suggested to be the lack of social interaction between the participants and the robots. Interestingly, in a recent word-association experiment (Salomons, van der Linden, Strohkorb Sebo, & Scassellati, 2018), the participants and a group of robots were asked to choose a card that best represented an administered word. The participants were assigned to a control condition, in which they did not see the robots answer, and an experimental one, in which they were exposed to the robots initial choice. Comparisons of the two groups showed an increased tendency to conform to the robots decision in the experimental condition.

Previous experiments in HRI have then partially replicated the group conforming effect, however, no studies have investigated further under which circumstances we prefer to conform to the robot leader decision, and in which people prefer to deviate. This chapter aims to further investigate human capabilities to apply and extend social rules to non-human agents. This would not only have implications on the definition of robot acceptance but can also contribute to the knowledge of human social cognition.

In two experiments, participants played the Economic Investment Game as part of a group made of robots. In the first experiment participants played with a team of two robots, one 'leader' and one 'follower'. To investigate under which circumstances participants would follow the leader opinion or take an individual decision, the agreement of the follower with the leader decision has been manipulated. Participants were asked to take an investment decision on a banker, after observing their robotic confederates investing their money. The similarity between the leader and the participants' investment was the main measure of conformity and was hypothesised to change with the follower consensus to the leader. Moreover, considering previous literature on the benefits of physical features on leadership perception, the role of height has been investigated by using two robots, the tall Pepper and the little NAO, which were randomly assigned to the role of the leader and the follower. Participants were expected to follow more frequently the Pepper leader than the NAO leade. However, this 'height effect' was expected to affect conformity only when the follower dissented with the leader. Results from previous chapters demonstrated, in fact, that robots physical attributes can influence participants' performance especially in hostile conditions. Therefore, in this study, participants' conformity was not expected to change when the robotic follower agreed with the leader. On the opposite, participants were expected to show higher conformity to the Pepper leader over the NAO when the robotic follower was disagreeing.

Although physical strength is associated with leadership perception, other characteristics seem to influence it as well. Among these, leader prototypicality is reported to affect leader credibility (e.g., Hais, Hogg, and Duck, 1997). According to the Social Identity Theory of Leadership (Hogg, 2001), leaders whose characteristics match the characteristics that typify their group, are assumed to be perceived as more effective. In other words, the more the leader is perceived to embody the group identity, the wider his/her influence will spread to the followers.

To test this hypothesis, a second experiment was conducted, in which a NAO robot follower, always faithful to the leader, was introduced to the team. Here the tall Pepper robot would become a minority. In this second experiment, participants interacted with a larger group of robots in which one follower always accepted the leader suggestions and the other showed some defective attitude. Considering the majority of NAO robots, the previous 'height effect' was hypothesised to be replaced by a 'numerosity prototypical effect': participants would prefer to follow the leader whenever it was a NAO than a Pepper, because of its similarity with the rest of the group. However, considering again previous chapters results on anthropomorphism, this effect was expected to emerge only when one of the followers dissented with the leader.

6.2 Experiment 1

6.2.1 Method

Participants

60 individuals aged between 19 and 30 years old (50 female and 9 male; mean age = 19.88 years, SD = 1.73 years) participated in the study. All participants were naïve as to the purpose of the investigation and gave informed written consent of their willingness to participate. All participants were native British English speakers, they were right-handed and reported normal hearing and no language or neurological impairments. Familiarity or any previous experience with robots was the main exclusion criteria, which was assessed during the recruitment phase.

6.2.2 Procedure

Investment game

The participants were seated in front of a NAO and a Pepper robot. A touchscreen monitor placed between the participants and the robots displayed the game. The participants were instructed to be part of the same team as the robots, playing against a virtual opponent, called "banker", whose decisions were indicated on the centre of the screen (Figure 6.1). The goal of the game was to earn as much money as possible from the banker. Each player started with a notional sum of 10 ECU at the beginning of every round of the game.

The game started with the leader robot introducing itself, underlying its leading role and then introducing the other robot (called "follower"). After that, the leader declared its investment choice (e.g. leader invested 6 ECU) and directed its attention

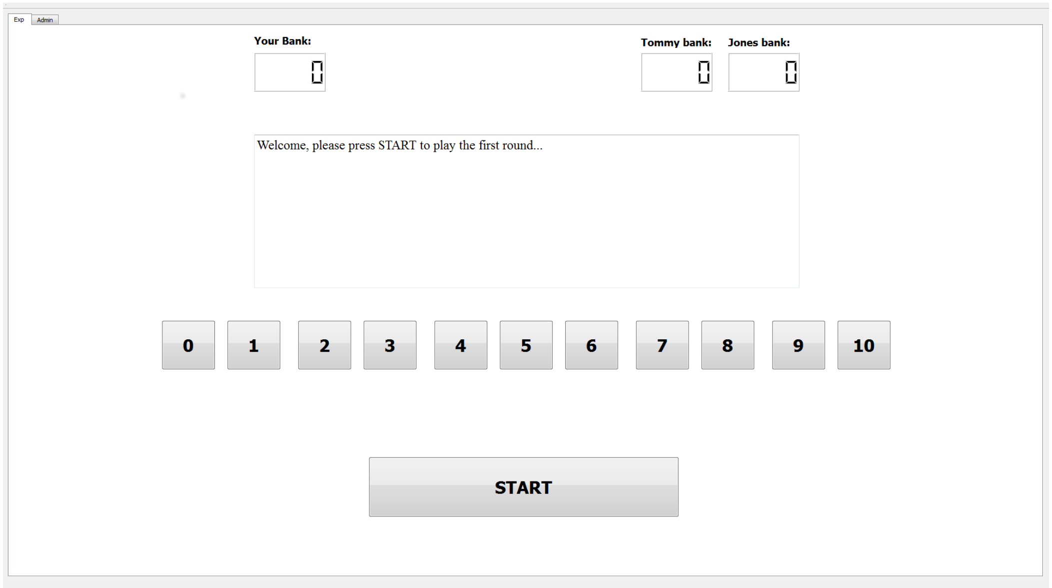

Fig. 6.1 Group Investment Game display for Experiment 1

to the follower, which in turn chose how much to invest (e.g. follower invested 4 ECU). After both robots invested, the participants would decide their amount to invest (e.g. participant invested 4 ECU). The amounts chose by the players were summed, tripled and invested to the virtual banker (6 + 4 + 4 = 14 * 3 = 42 ECU), which in turn decided the returns (e.g. 42 * 0.2 = 8.4 ECU). To increase the cooperation between the players, punishment has been included. This has been calculated as the absolute difference between the leader investment and the average investment of the remaining players (6 - 4 = 2 ECU) and subtracted from the banker payoff (8.4 - 2 = 6.4 ECU). In this way, the larger was the difference between the players' choices, the higher was the punishment they received. Finally, the amount returned was equally split between the three members of the group (6.4 / 3 = 2.1 ECU to each player). The virtual banker was programmed to return between 0 % and 30 % of the received amount at each round.

The display placed in front of the participants showed 11 numbered buttons, that participants could press to indicate the amount of money they wanted to invest in each round. The screen also showed how much money all players were making (Figure 6.1). These buttons were called "bank" and were summing all the money the players were earning separately. The total amount earned by the participants at the end of each game was converted to British Pounds, at a rate of 30 ECU = £0.10 and the sum of what they earned in the two games was paid to them at the end of the experiment.

To make the interaction more natural, the robots performed a well-organised turn-taking conversation during the game (e.g. leader looking at the follower while asking it to invest, both robots turning their attention to the participant during his/her turn) as well as congruent postures, hand gestures and arm movements. This has been achieved by using the ALAnimatedSpeech module (http://doc.aldebaran.com/2-1/naoqi/audio/alanimatedspeech.html). For a complete description of the game, a detailed timeline is reported in Figure 6.2.

The number of rounds was not known to the participants. Each participant engaged in two blocks of rounds. In the first block, the robots were programmed to take similar decisions, so that the follower investment was always close to the leader choice (collaborative condition). Specifically, the investment was always a number between 6 and 10 for both robots. In order to check any differences, t-tests between the two robots investments have been performed and no significant effect has been found ($p_s > .05$). In the second block, the follower changed its investment strategy and stopped conforming with the leader decisions (non collaborative condition). In this case, the follower investment was programmed to be always a number between 0 and 5, and always at least 3 ECU lower than the leader investment. In order to avoid any interference during the game, the participants and the robots were left alone in the room. Nevertheless, the experimenter could still monitor the game from a window placed behind the participants.

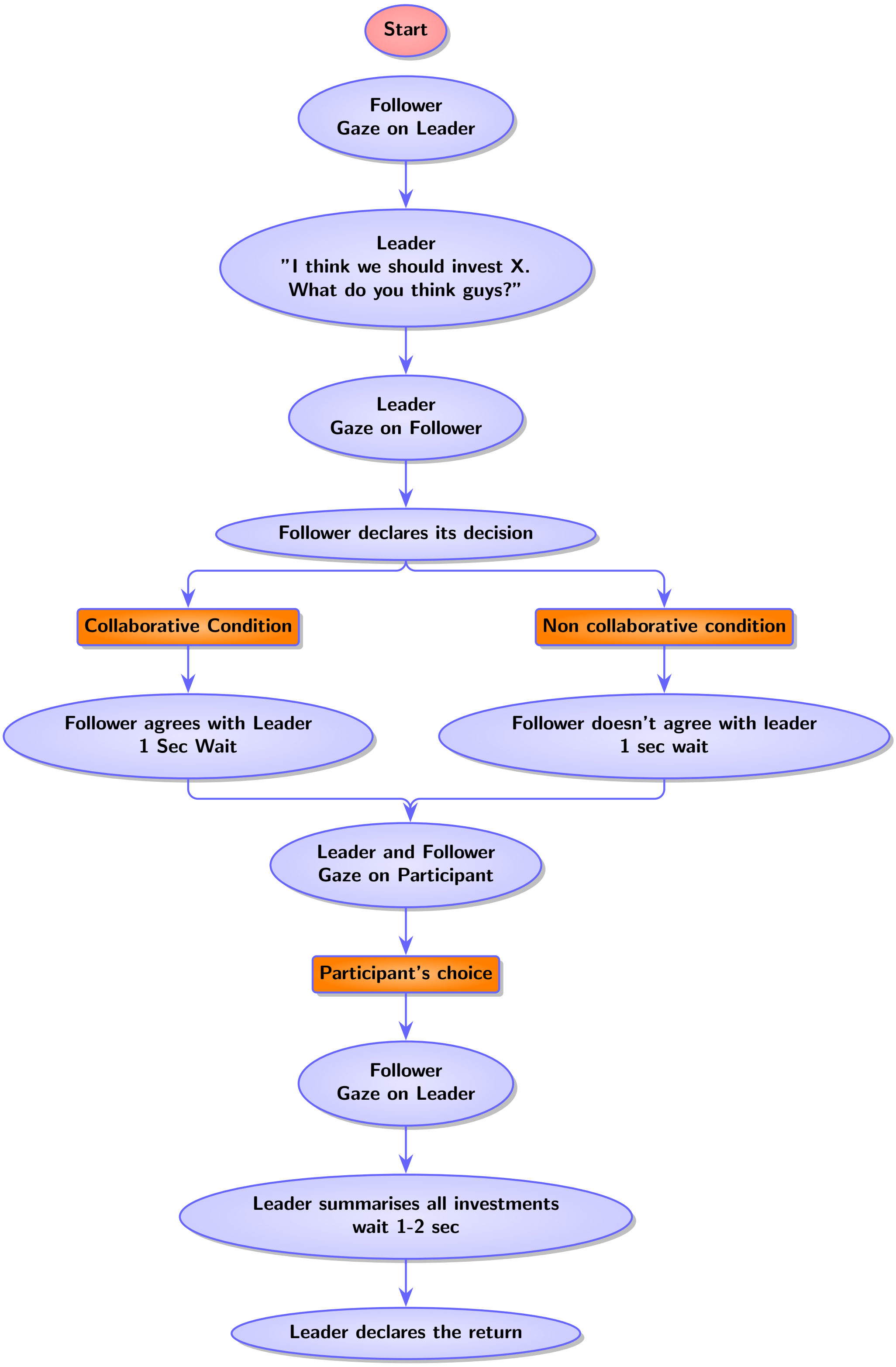

Fig. 6.2 Timeline of the group Investment Game with one leader and one follower.

The experiment was counterbalanced in a 2 (leader: Pepper or NAO) between subject and a 2 (follower: collaborative, non collaborative) within subject design.

Questionnaires

Four questionnaires were used as secondary measures to the main game task. Three short scales measured likeability, trust, and credibility (Goldberg et al., 2006; McCroskey & Young, 1981; Reysen, 2005). In addition, Bartneck et al. (2008) questionnaire was used to measure a range of HRI factors (anthropomorphism, animacy, likeability, perceived intelligence and perceived safety). Participants filled the questionnaires about their views on both the follower and leader after the experiment ended.

6.2.3 Results

Preference Index

One participant did not complete both game blocks and was excluded from the analyses. Participants' tendency to follow the leader has been defined as $preference$ $index$, which has been calculated by comparing the participants' investment decision to both the leader and the follower decisions and by applying a binary categorical as follow:

$$Preference-Index^{1}\begin{cases} \text{if } (P-L) \leq (P-F) = 0 \\ \text{if } (P-L) > (P-F) = 1 \end{cases} \tag{6.1}$$

[1]where P is participants' investment, L is leader investment and F is the follower investment

In each round, this would be represented as a binary categorical value, set to 1 if the participants' choice was closer to the leader decision than the follower, and 0 if their choice was equally distant from both players or closer to the follower. A logistic mixed-effects model was fitted to the data using backwards stepwise selection, with preference index as the dependent variable, leader (NAO/Pepper), follower (collaborative/non collaborative), and game turn (the 20 rounds of the game) as independent variables and participant id as a random factor. Post-hoc comparisons were assessed using t-tests and Bonferroni's correction was applied when needed. Odds ratios are reported on Table 6.1.

Table 6.1 Odds ratios and confidence intervals for Experiment 1.

	Odds Ratio	**C.I.**
Follower-*Non collaborative*	0.68	0.47-0.98
Game turn	0.97	0.95-0.99
Leader-*Pepper*:Follower-*Collaborative*	1.11	0.79-1.56
Leader-*Pepper*:Follower-*Non Collaborative*	0.85	0.57-1.26
Leader-*NAO*:Follower-*Collaborative*:Game turn	1.00	0.97-1.03
Leader-*NAO*:Follower-*Non Collaborative*:Game turn	0.94	0.91-0.98
Leader-*Pepper*:Follower-*Collaborative*:Game turn	0.99	0.96-1.02

A significant main effect of follower has been found ($\chi^2_{(1)} = 99.53, p < .001$); participants' preference for the leader choice was significantly higher whenever the follower was collaborative (mean = 0.30) compared to the non collaborative condition (mean = 0.12). Game turn affected participants' preference as well ($\chi^2_{(1)} = 5.48, p = .019$) showing a decrease of leader consensus over time. Leader had no effect on participants' choices (p > .05).

The two-way interaction between leader and follower was significant ($\chi^2_{(2)} = 7.63, p = .022$). Participants did not discriminate between the two leaders in the collab-

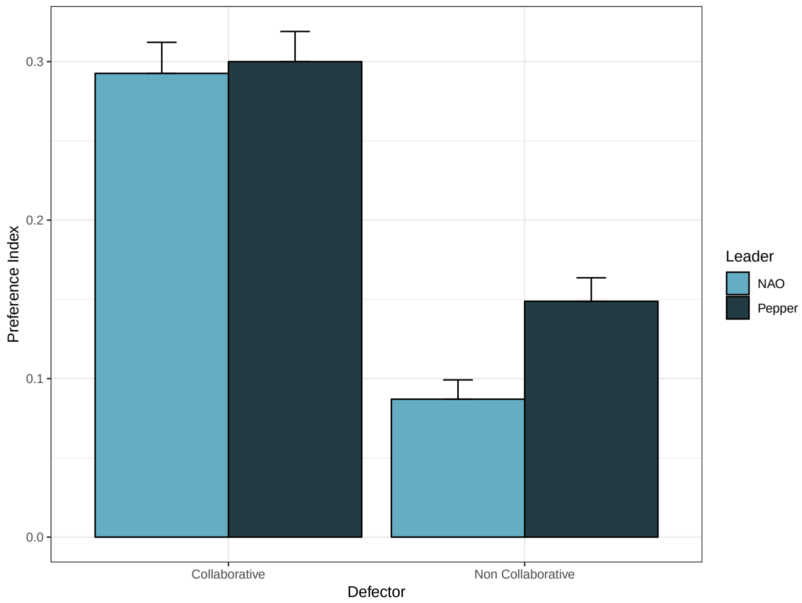

Fig. 6.3 Preference index for NAO and Pepper leaders in the collaborative and non collaborative conditions.

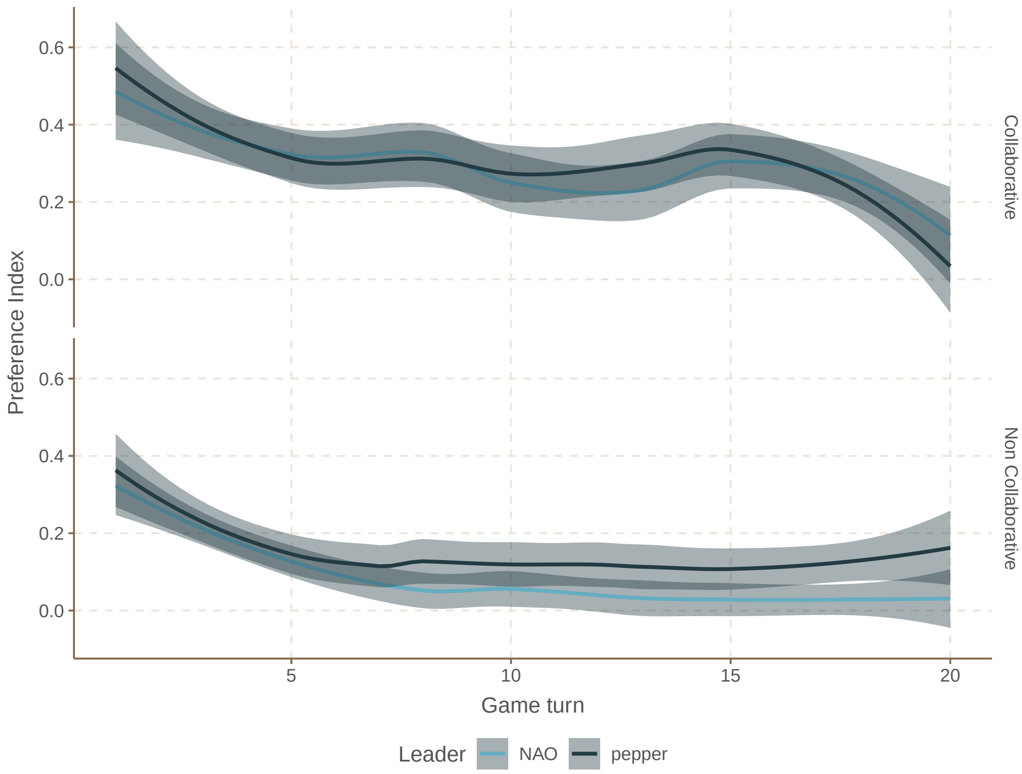

Fig. 6.4 Preference index variations during the game for NAO and Pepper leaders in the collaborative (top) and non collaborative (bottom) conditions.

orative condition ($p > .05$). In the non collaborative condition, instead, participants preferred the Pepper leader to the NAO ($p = .045$) (Figure 6.3).

Finally, a significant three-way interaction between leader, follower and game turn has been found ($\chi^2_{(3)} = 10.86, p = .012$). Post-hoc analyses for the collaborative condition revealed a main game turn effect ($\chi^2_{(1)} = 21.19, p < .001$), showing a gradual decrease of leader consensus in the last rounds of the game. No significant interaction between leader and game turn has been found (p > .05). For the non collaborative condition, there was a significant main effect for game turn ($\chi^2_{(1)} = 27.59, p < .001$) and a significant two-way interaction between leader and game turn ($\chi^2_{(1)} = 11.31, p < .001$). After an initial decrease in the first rounds, the preference for the NAO leader was maintained constantly lower than the Pepper leader (Figure 6.4).

Questionnaires Results

For each scale, a linear mixed-effects model fitted to the data, with scale as the dependent variable, robot (NAO/Pepper) and role (follower/leader), as independent variables and participant id as a random factor. As shown in Table 6.3, role had an effect for the Trust, Anthropomorphism, Animacy, Likeability and Intelligence scales. For the former four scales, higher ratings for the leader than the follower are reported. The opposite result has been found for the Intelligence scale. The type of robot affected only Safety scale, with higher ratings for Pepper robot. No other significant differences have been found ($p_s > .05$).

Table 6.2 Cronbach's alpha.

	Alpha
Likeability	0.88
Trust	0.79
Credibility	0.85
Godspeed Questionnaires	
Anthropomorphism	0.84
Animacy	0.89
Likeability	0.92
Intelligence	0.88
Safety	0.64

Table 6.3 Results from questionnaires analyses of Experiment 1.

	Robot	**Role**	**Robot By Role**
Likeability	n.s.	n.s.	n.s.
Trust	n.s.	$\chi^2_{(1)} = 8.25, p = .004$	n.s.
Credibility	n.s.	n.s.	n.s.
Godspeed Questionnaires			
Anthropomorphism	n.s.	$\chi^2_{(1)} = 9.77, p = .002$	n.s.
Animacy	n.s.	$\chi^2_{(1)} = 4.85, p = .028$	n.s.
Likeability	n.s.	$\chi^2_{(1)} = 13.36, p < .001$	n.s.
Intelligence	n.s.	$\chi^2_{(1)} = 12.81, p < .001$	n.s.
Safety	$\chi^2_{(1)} = 9.54, p = .002$	n.s.	n.s.

6.2.4 Discussion

From a behavioural point of view, results indicate that participants conformed to the leader when supported by the follower. On the other hand, when the follower changed its strategy, the participants preferred to take an autonomous decision and invested a different amount of money. The questionnaires, however, reported participants to prefer the leader over the follower, thus indicating that they recognised the leader authority. These results show that people could be able to apply human conformity rules to a group of robots and would entitle a leadership role to a robotic agent. Furthermore, when the follower agreed with the leader, participants did not discriminate between the two types of leaders. Nevertheless, when a disagreement between the robots emerges, the participants preferred to follow the taller Pepper leader. These results give confirmation to the hypothesis that height is a factor capable of increasing the sense of leadership and can be transferred to non-human agents.

On a second experiment, the investigation of leader conformity has been extended, by increasing the number of robotic followers. Participants played with three robots, one of which always taking the leader side. In this way, a wider exploration of the role of a dissenting minority can also be performed. Moreover, the prototypicality of the leader has been investigated in order to measure the effect of physical strength in leadership perception. This has been achieved by adding a second NAO robot to the team. In a group of small robots, would a small leader be preferred (because of its resemblance to the team) or would participants still follow the taller leader?

6.3 Experiment 2

6.3.1 Method

Participants

61 individuals between 19 and 30 years old (51 female and 10 male; mean age = 19.82 years, SD = 2.26 years) participated in the study. All participants were naïve as to the purpose of the investigation and gave informed written consent to participate in the study. All participants were native British English speakers, they were right-handed and reported normal hearing and no language or neurological impairments. Familiarity or any previous experience with robots was the main exclusion criteria, which was assessed during the recruitment phase.

Procedure

The procedure was the same as for Experiment 1. Alternatively, a NAO and a Pepper played the leader role, while the other was assigned to the follower role. The novelty of this experiment was the introduction of a new robotic follower, another NAO robot. This second follower was programmed to always conform and follow the leader decisions, thus always choosing a similar amount to invest in both experimental blocks. Moreover, a robotic NAO banker was positioned on the right corner of the room and monitored the game from its own computer (Figure 6.5).

Like for the previous experiment, the leader made its choice first, then asked the two followers to decide their investments. The order to which the two followers made their choice was randomised throughout the game. The participants would invest their money once all the other players took their decisions. The amounts of

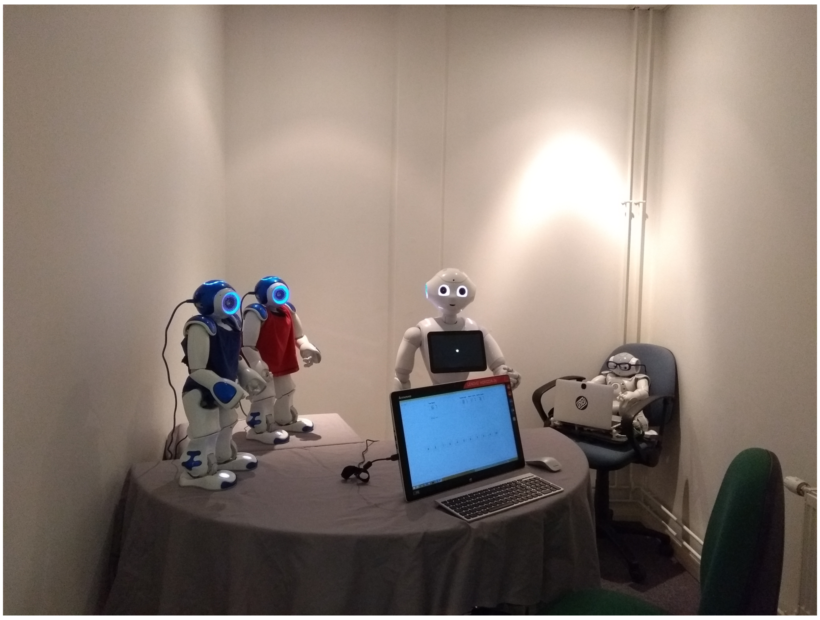

Fig. 6.5 Group Investment Game setup for Experiment 2.

all the four players were summed, tripled and sent to the robotic banker, which then decided the returns. As for Experiment 1, a punishment for not agreeing with the leader has been introduced and included the second follower investment as well. In this case, the average investment of the players (e.g. 5 + 6 + 4 = 15 / 3 = 5 ECU) was subtracted from the leader investment (e.g. 6 ECU, thus punishment is 6 - 5 = 1 ECU) and then subtracted from the returned amount.

The interaction between the robots has been manipulated in terms of gestures and speech in order to resemble as natural as possible. Moreover, the monitor in front of the participants was updated with the second follower "bank”. In order to not interfere with the experiment, the banker was mute and immobile for all the duration of the game. For a complete description of the game, a detailed timeline is reported in Figure 6.6.

The experiment was counterbalanced in a 2 (leader: Pepper or NAO) between subject and a 2 (follower: collaborative, non collaborative) within subject design.

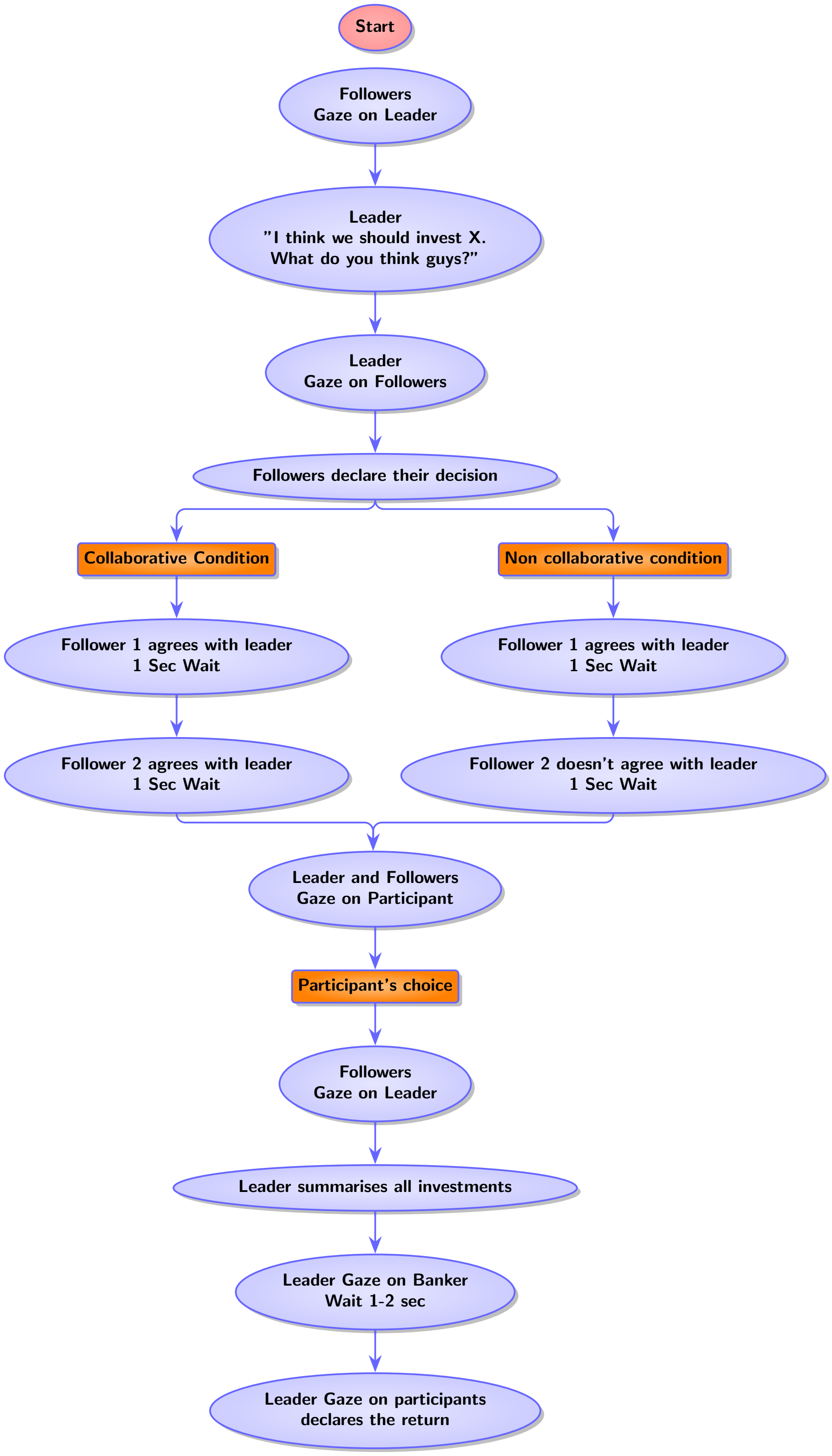

Fig. 6.6 Timeline of the group Investment Game with one leader and two followers.

6.3.2 Results

Game results

A logistic mixed-effects model was fitted to the data using backwards stepwise selection, with preference index as the dependent variable, leader (NAO/Pepper), follower (collaborative/non collaborative), and game turn (the 20 rounds of the game) as independent variables and participant id as the random factor. Post-hoc comparisons were assessed using t-tests and Bonferroni's correction was applied when needed. Odds ratios are reported in Table 6.4.

Table 6.4 Odds ratios and confidence intervals for Experiment 2.

	Odds Ratio	**C.I.**
Follower-*Non collaborative*	0.69	0.59-0.82
Game turn	0.97	0.96-0.98
Leader-*Pepper*:Follower-*Collaborative*	1.03	0.84-1.27
Leader-*Pepper*:Follower-*Non Collaborative*	0.71	0.57-0.90

A significant main effect of follower has been found ($\chi^2_{(1)} = 76.52, p < .001$); participants; preference for the leader choice was significantly higher in the collaborative condition (mean = 0.28) than the non collaborative one (mean = 0.14). Game turn affected participants' preference as well ($\chi^2_{(1)} = 26.93, p < .001$), which was gradually reduced over time. Leader had no effect on participants' choices ($p > .05$).

A significant two-way interaction between leader and follower was found ($\chi^2_{(2)} = 10.89, p = .004$). Participants did not discriminate between the two leaders in the collaborative condition ($p > .05$). In the non collaborative condition, instead, participants preferred the NAO leader to the Pepper ($p = .011$) (Figure 6.7). No other significant interactions have been found ($p_s > .05$).

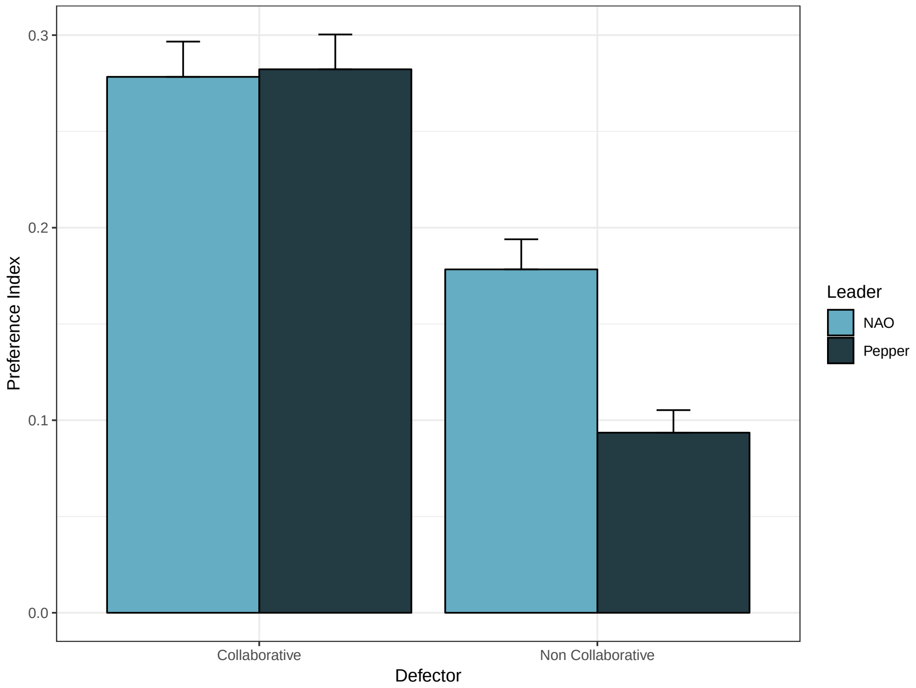

Fig. 6.7 Conformity index for NAO and Pepper leaders in the collaborative and non collaborative conditions.

Questionnaires Results

For each scale, a linear mixed effect model was fitted to the data, with scale as the dependent variable, robot (NAO/Pepper) and role (follower/leader) as independent variables and participant id as the random factor. As shown in Table 6.6, the type of robot had an effect only for the Likeability scale, with higher ratings for the Pepper. The role of the robot, instead, affected all questionnaires with the exception for Anthropomorphism. For Likeability, Trust, Credibility, Animacy and Safety, participants preferred the leader over the follower, while instead this was rated more intelligent than the leader. The interaction between role and robot was not significant for any of the scales.

Table 6.5 Cronbach's Alpha

	Alpha
Likeability	0.86
Trust	0.86
Credibility	0.79
Godspeed Questionnaires	
Anthropomorphism	0.82
Animacy	0.89
Likeability	0.93
Intelligence	0.90
Safety	0.52

Table 6.6 Results of the questionnaires analyses for Experiment 2.

	Robot	**Role**	**Robot By Role**
Likeability	n.s.	$\chi^2_{(1)} = 31.48, p < .001$	n.s.
Trust	n.s.	$\chi^2_{(1)} = 32.15, p < .001$	n.s.
Credibility	n.s.	$\chi^2_{(1)} = 5.90, p = .015$	n.s.
Godspeed Questionnaires			
Anthropomorphism	n.s.	n.s.	n.s.
Animacy	n.s.	$\chi^2_{(1)} = 4.26, p = .039$	n.s.
Likeability	n.s.	$\chi^2_{(1)} = 79.73, p < .001$	n.s.
Intelligence	n.s.	$\chi^2_{(1)} = 11.78, p < .001$	n.s.
Safety	n.s.	$\chi^2_{(1)} = 8.74, p = .003$	n.s.

6.3.3 Discussion

The results from Experiment 2 confirmed the previous conformity effect. Participants followed the leader when supported by its follower. On the opposite, when a minority dissented, participants preferred investing a different amount of money from the one suggested by the leader. Moreover, questionnaires again reported participants to prefer the leader over the follower. Furthermore, investigations on the role of prototypicality showed no effect on the collaborative condition. Like in the previous experiment, when all the followers agreed with the leader, participants did not discriminate between the two types of leader. However, when one of the followers dissented from the majority strategy, participant preferred to follow the little NAO leader than the taller Pepper. These results sustain the hypothesis that a leader showing similar features to its group has higher probabilities of influencing people's decisions.

Finally, to fully understand the role of the follower on leadership perception, an analysis comparing the two experiments has been performed. This has been done by introducing a new variable called "position", which refers to the amount of consensus reached by the leader. While in the first experiment the leader could fall in a minority position with only one follower, in the second experiment the leader moved to a majority position thanks to the help of a second faithful follower. This analysis would also give some insight into the role of height and prototypicality on leadership.

6.4 Comparisons of experiment 1 and 2

A logistic mixed-effects model was fitted to the data using backwards stepwise selection, with preference index as the dependent variable, leader (NAO/Pepper), follower (collaborative/non collaborative), game turn (the 20 rounds of the game) and position (minority/majority) as independent variables and participant id as a random factor. Post-hoc comparisons were assessed using t-tests and Bonferroni's correction was applied when needed.

6.4.1 Results

Game results

Table 6.7 Odds ratios and confidence intervals for the final comparison analysis.

	Odds Ratio	**C.I.**
Follower-*Non collaborative*	0.57	0.37-0.68
Game turn	0.97	0.47-0.97
Leader-*NAO*:Follower-*Collaborative*:Position-*Majority*	0.94	0.36-1.14
Leader-*NAO*:Follower-*Non Collaborative*:Position-*Majority*	1.13	0.35-1.41
Leader-*NAO*:Follower-*Collaborative*:Position-*Minority*	0.98	0.34-1.20
Leader-*NAO*:Follower-*Non Collaborative*:Position-*Minority*	0.71	0.34-0.91
Leader-*Pepper*:Follower-*Collaborative*:Position-*Majority*	0.97	0.36-1.17
Leader-*NAO*:Follower-*Non Collaborative*:Position-*Majority*	0.81	0.35-1.01

A significant main effect of follower has been found ($\chi^2_{(1)} = 183.59, p < .001$); participants' preference for the leader choice was significantly higher in the collaborative condition (mean = 0.29) than in the non collaborative one (mean = 0.13). Game turn affected participants' preference ($\chi^2_{(1)} = 73.58, p < .001$), with a decrease of preference over time. Leader and position had no effect on the preference index ($p_s > .05$).

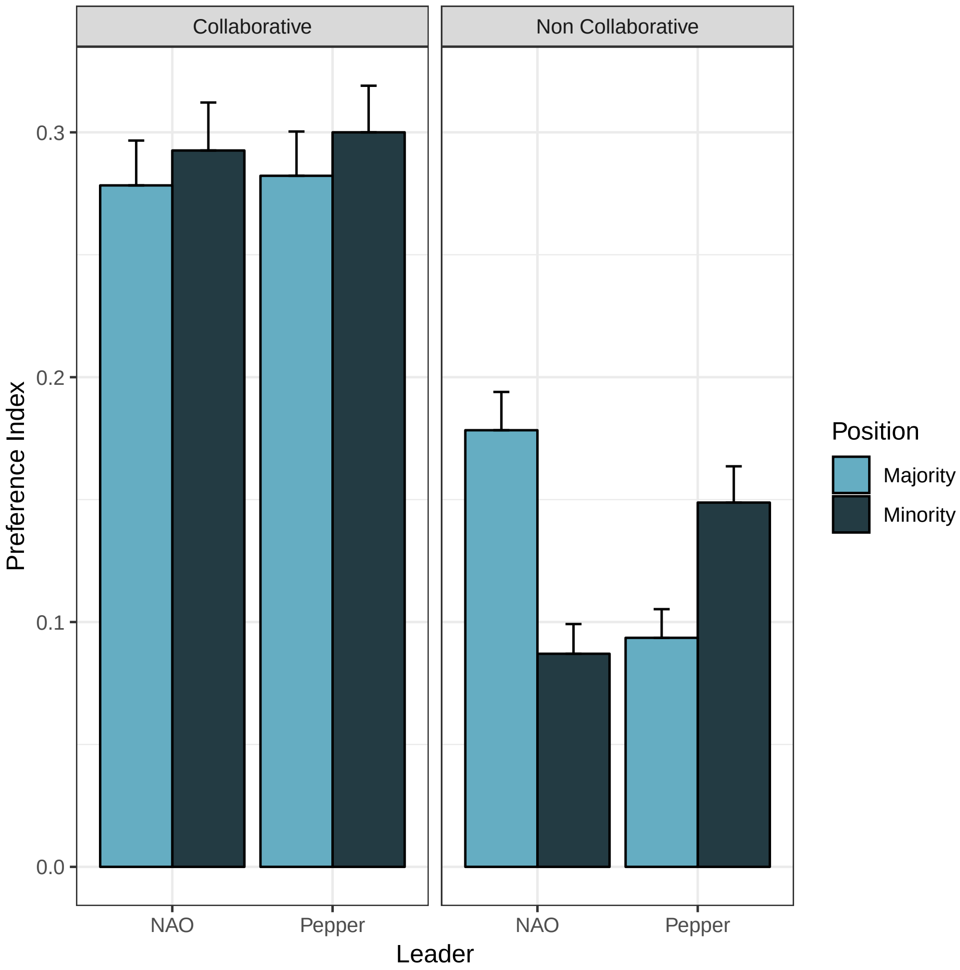

Fig. 6.8 Conformity index in the non collaborative condition for NAO and Pepper leaders in majority (light blue) and minority (dark blue) positions.

A significant three-way interaction between leader, follower and position emerged ($\chi^2_{(6)} = 22.42, p = .001$). As showed in Figure 6.8, for the collaborative follower, position and leader had no effect ($p_s > .05$). For the non collaborative follower, previous results have been confirmed; the NAO leader collected more consensus than the Pepper when in majority position (0.18 vs 0.11, $p = .002$). On the opposite, participants preferred to agree with the Pepper leader than the NAO when in minority position (0.15 vs 0.09, $p = .008$). Finally, the NAO leader reached higher consensus when in majority than in minority position (0.18 vs 0.09, $p = .002$), while for the Pepper leader no differences on position have been found (0.11 vs 0.15, $p > .050$).

Questionnaires Results

Table 6.8 Results for questionnaires analyses in the final comparison.

	Robot	**Role**	**Position**
Likeability	n.s.	$\chi^2_{(1)} = 8.96, p = .003$	n.s.
Trust	n.s.	$\chi^2_{(1)} = 10.92, p < .001$	n.s.
Credibility	n.s.	n.s.	n.s.
Godspeed Questionnaires			
Anthropomorphism	n.s.	n.s.	n.s.
Animacy	n.s.	n.s.	n.s.
Likeability	n.s.	$\chi^2_{(1)} = 26.82, p < .001$	$\chi^2_{(1)} = 6.67, p = .009$
Intelligence	$\chi^2_{(1)} = 7.77, p = .010$	n.s.	n.s.
Safety	n.s.	n.s.	n.s.

For each scale, a linear mixed effect model fitted to the data, with scale as the dependent variable, robot (NAO/Pepper), role (follower/leader) and position (minority/majority) as independent variables and participant id as a random factor. As shown in Table 6.8, the type of role had an effect for the Likeability and Trust scales, with higher ratings for the leader than the follower. The type of robot affected only the Intelligence scale, with higher ratings for the Pepper robot. The type of position affected only Likeability scale, with higher ratings for the minority condition.

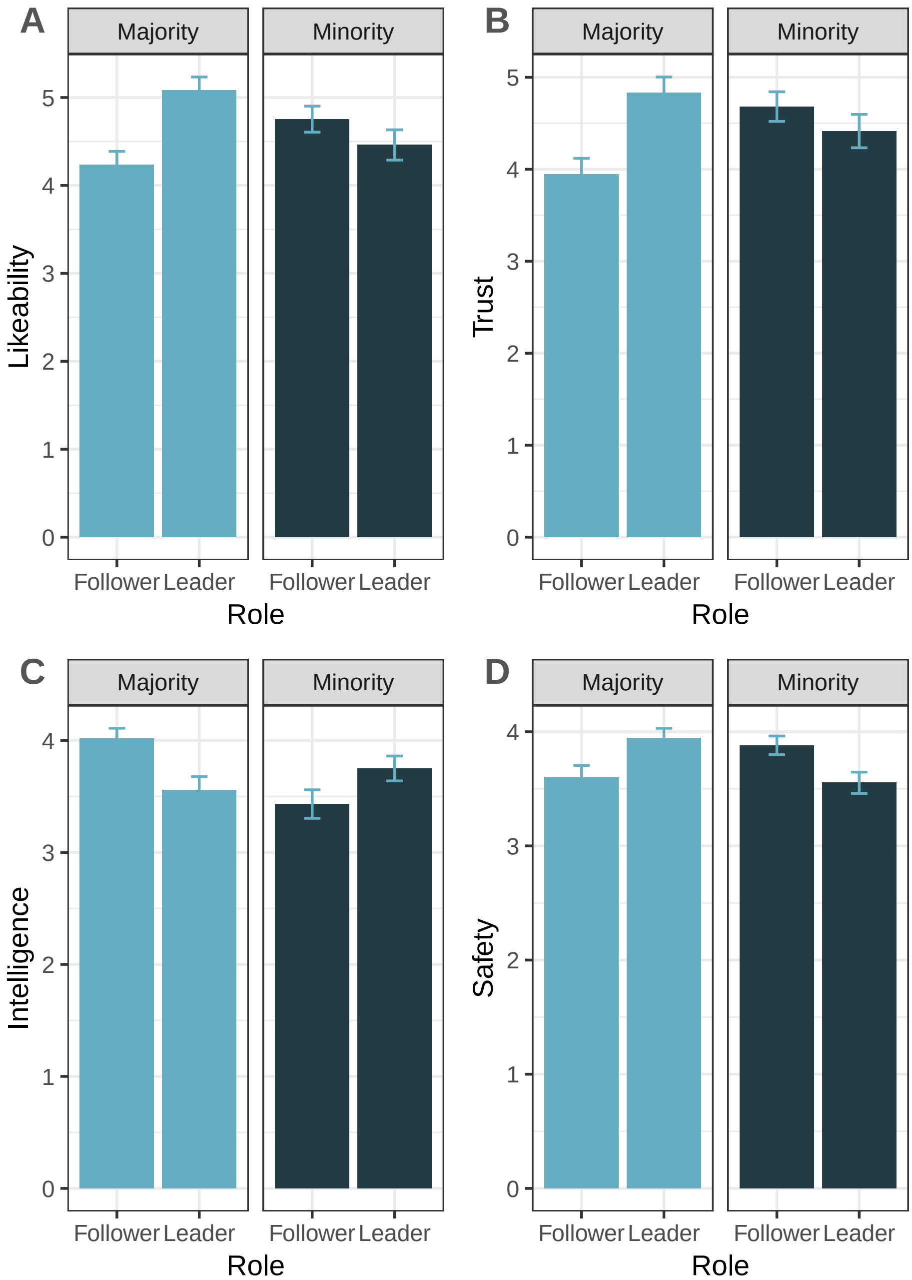

Fig. 6.9 Participants' ratings for follower and leader roles in majority and minority positions, for Likeability (panel A), Trust (panel B), Intelligence (panle C) and Safety (panel D) scales.

A significant interaction between role and position was found for both Likeability ($\chi^2_{(1)} = 32.04, p < .001$; $\chi^2_{(2)} = 73.44, p < .001$), Trust ($\chi^2_{(2)} = 33.89, p < .001$), Intelligence

($\chi^2_{(6)} = 20.77, p = .002$) and Safety ($\chi^2_{(3)} = 18.73, p < .001$). Likeability, Trust, and Safety revealed higher preferences for the leader over the follower when in majority position ($p_s < .05$). Moreover, for the Safety scale, the role of the leader in the minority was perceived significantly less safe than in the majority ($p = .017$). The Intelligence scale had higher ratings for the follower over the leader when in the majority ($p = .019$).

Finally, the interaction between robot and position was found to be significant for Likeability ($\chi^2_{(3)} = 11.76, p = .001$), with higher ratings for the NAO robot in the majority position ($p = .002$).

6.5 Discussion

In this Chapter, leader conformity dynamics in HRI have been investigated, by posing the following questions: do we conform and deviate to robots in the same way we do with other humans? How does the support to a robotic leader affect conformity and deviation? In order to answer these questions, the participants' tendency to rely on the investment decision of a robotic leader have been studied. Specifically, two conformity conditions have been created, one in which the leader received the followers' support and one in which a follower dissented from the leader. This has been done to investigate under which circumstances participants would follow a robot leader. Results indicated that humans can follow a robotic majority in the same way they follow a human majority. Whenever the followers conformed to the leader investment decision, participants agreed in choosing a similar monetary amount, even when the choice of the leader was strategically weak (e.g. investing a high amount of money when the payoff was low). These results confirmed that humans can ascribe social in-group competences to robots and demonstrated that people apply to robots the same responsive behaviour they apply to their similars, like accepting the majority decision as a valid alternative (Asch & Guetzkow, 1951; Deutsch & Gerard, 1955; Martin, Thomas, Hewstone, & Gardikiotis, 2018). Since conformity derives from the need for approval and social acceptance, the present results indicate that humans are capable of transferring these approval needs to non-human entities. That is, once they feel part of a group, regardless of the nature of its members, they act in accordance with the group rules in order to be accepted.

Exposure to a robotic dissenter decreased the probability to conform. When the robot follower showed some refrain in agreeing with the leader and started defecting, the number of times participants agreed with the leader dropped drastically. This is in line with previous literature on minority exposure (Allen, 1975; Clark III, 1990; Mugny & Pérez, 1991; Nemeth & Chiles, 1988; Papastamou, 1986) and confirmed the human capability to include robots in their social environment and apply to them human-human interactive rules.

In a second experiment, participants played the investment game with a team formed of one leader and two followers, one of which always accepting the leader choice and the other showing defective tendency. This was done to investigate a possible 'majority effect' on leader conformity. While in the first experiment a real majority consensus was not reached during the second block, in the latter experiment the leader could always count the support of one follower. In the second experiment, participants were expected to show a lower tendency to deviate from the majority. Behavioural data did not support this hypothesis. Instead, the same result of the first experiment has been obtained, showing that the numerosity of the team does not affect deviation from leader conformity. The lack of this effect can be due to the structure of the game: since the leader was always suggesting a wrong strategy (investing a high amount of money on a selfish banker), it is possible that participants exposed to a non collaborative follower decided to dissent regardless of the amplitude of consensus. Analysis of game turn, in fact, showed a gradual decrease of conformity over time, thus suggesting that they realised the weakness of the leader strategy. However, questionnaires partially confirmed the numerosity hypothesis. Participants rated higher the leader in both experiments, but it rated more likeable, trustworthy and safe than when in minority. Overall, these results indicated that, even if conformity was not behaviourally affected by a larger consensus on the leader, participants still recognised its powerful position, especially when supported by more followers.

Nevertheless, some limitations on the structure of the game block should be mentioned. Since the aim of the study was to investigate participants' reaction to a dissenting minority, they were initially exposed to a block with a collaborative follower and then with a non collaborative one. This was done to give the participants the impression that the follower initially recognised the role of the leader and tried to comply to it, but after a first try, it decided to play an independent strategy. Although a follower that initially dissents to the group and later decides to agree with the leader might sound unusual, changing the order of the block could reveal other interesting insights. Would participants immediately dissent to the leader? If so, would they keep dissenting when all the others are obeying to the leader?

Another hypothesis of this study regarded the role of the leader features in conformity formation, for which height and prototypicality have been considered. Since height has been reported as one of the main contributors to the perception of leadership, in the first experiment comparisons between a tall Pepper robot and a small NAO have been performed. Results supported the hypothesis that a taller robotic leader is a more influencing presence, but only when the follower deviates from its decision.

Although height has a great impact on leader acceptance, it has been also demonstrated that carrying similar features to the followers can be equally important. In the second experiment, prototypicality has been investigated by introducing another NAO follower, thus expecting participants to prefer the small robot as a leader over the tall Pepper. Analyses showed that participants agreed more often with the NAO leader in the second block than the Pepper leader. These results are in line with recent studies demonstrating that the leader prototypicality affects conformity and effectiveness perception of leadership (e.g. Barreto and Hogg, 2018).

The data from both experiments support the previous evidence on the selective effect of anthropomorphic features in HRI. Whenever the robot behaviour follows people's expectations (e.g. the follower agrees with the leader) the features of the robot leader are not a necessary characteristic to improve conformation. However, when a robot does not follow the expected rules, uncertainty on the behaviour that participants should choose arises and stereotypical features of the leader become a significant impactful variable on their final decision. This hypothesis has been further supported by the final analysis. The comparisons between the two experiments showed no differences of conformity for Pepper leader in the two groups. Participants were equally conforming to Pepper leader independently from the number of NAOs surrounding it. On the opposite, the tendency to conform to the NAO leader drastically increased when surrounded by other NAOs. The leader prototypically affected participants' conformity, replacing the anthropomorphic height effect of the first experiment. Likeability questionnaire confirmed the behavioural results; participants gave higher ratings to the NAO robot in the large group. It could be concluded that prototypicality is a stronger factor in affecting conformity to the leader. However, it should be considered that a comparison with a majority of Pepper robots has not been performed. Considering the results here reported it can be hypothesised that a Pepper leader, guiding a majority of other Peppers, would benefit from both height and similarity contributions and further increase conformity over a majority of NAO robots.

Nevertheless, the role of the banker should be considered. In the second experiment, in fact, the role of the banker was applied to a real NAO robot, whereas in the first study the barker was confined to a virtual presence. It is then plausible to hypothesise that a NAO robot could strengthen the prototypicality effect. Stemming from this, future study should investigate the role of a physical Pepper banker in a group of Pepper robots to better understand how both height and prototypicality affect human conformity to robots. Furthermore, the role of a human follower/leader among a group of robots, give some useful insights on the development of in-group feelings.

6.6 Conclusion

Humans can develop an in-group feeling towards a group of robots and can congruently behave in order to be socially accepted by them. Here, our understanding of the psychology of HRI has been extended by showing that humans conform to robots in the same way they do with other humans. People follow a robot leader whenever it is supported by the group members, and prefer changing their attitude whenever exposed to a dissenting robotic agent. Moreover, humans apply to a robot leader the same stereotypical expectations they apply to other humans. They associate height to the perception of power and authority as well as they prefer engaging with a leader that embodies the same characteristics of the followers. However, this happens only when a dissenting minority appears. This demonstrates that the physical features of a robot can affect social interactions under specific conditions, namely when the robot behaviour does not reflect the human's expectations.

Chapter 7

Discussion

Well we know where we're going
But we don't know where we've been
And we know what we're knowing
But we can't say what we've seen
And we're not little children
And we know what we want
And the future is certain
Give us time to work it out.

Road to Nowhere
Talking Heads

This chapter brings together the work that has been reported within this thesis to show how the main findings relate to the five research questions it has addressed. The key contributions of this research to HRI are discussed and a collection of suggestions for future avenues of research are outlined.

7.1 When do we cooperate with robots?

The aim of this thesis was to explore how humans accept robots and cooperate with them. Due to their potential raising presence in the most diverse fields, from industry to elderly care, the need for reducing the social distance between humans and robotic agents has become compelling.

Robot trust and acceptance have been addressed as two of the main issues in the last decades of studies and anthropomorphism has been entitled as the most influential factor among all the different theories and hypotheses. Moreover, with the evolution of new interactive scenarios, the concept of anthropomorphism has not anymore been limited to the physical features of robots but has absorbed also social and behavioural competencies that can increase the chances of including robots in everyday human life. Interestingly, recent studies have also suggested that anthropomorphism can be an instrument with limited success. There is evidence, in fact, that under certain circumstances human-likeness is not only a useless instrument but can also negatively impact the quality of the interaction.

Finally, future inclusion of robots into many human environments has also led to the necessity of reviewing the concept of HRC, moving from a robotic subordinate to a real "au pair" view of the interaction.

The aim of this thesis was to explore how we extend the use of social collaborative cues to robots and what are the necessary anthropomorphic requirements for that extension. Specifically, the following questions have been raised:

- Can anthropomorphic stereotype activation affect the robot credibility?
- How does trust in HRI evolve over time and under what specific circumstances anthropomorphism increases trust in HRI?

- Can we cooperate with robots as we cooperate with other humans? What is the role of anthropomorphism in affecting cooperation?
- Can people choose to strategically imitate a robot and how would anthropomorphism influence social learning in HRI?
- Would humans conform to a robotic leader or would prefer to follow a dissenting minority? Which features should a robotic leader embed to be trusted?

7.1.1 Can anthropomorphic stereotype activation affect the robot credibility?

First impressions matter even with robots and anthropomorphism can help humans in accepting a robot at their first encounter. As suggested by the literature, human-likeness is one of the most powerful instrument we have to influence people beliefs and judgments on robots. Moreover, it has been largely demonstrated that human-like resemblance is capable of activating the "you look like me, so I like you" stereotype, which consequently opens the path to new interactive possibilities.

Nevertheless, research in HRI was always focused on the formation of people's first impression and on what a possible positive impression might depend without testing whether we can learn to like and accept a robotic agent. As demonstrated in Chapter 2, a robot credibility can be increased through stereotype activation and can be extended to other exemplars of the same category. On the basis of the results reported, the previous "you look like me, so I like you" stereotype can be rephrased as a "you look like that nice robot, so I like you too" stereotype. People do not only seem to prefer to interact with an anthropomorphic robot, but they also learn to equally appreciate a machine-like agent because of the pleasant impression and experience they had in the past with a human-like robot. Moreover, the opposite reaction can be activated too. A first negative impression is capable of reducing credibility and any positive feelings in a robot that usually provides a sense of familiarity a trust.

This stereotype generalisation effect responds to the same HHI pattern, thus demonstrates that a sort of human-like interaction with robots is possible, although some limitations are inevitable. In fact, it appears that only the physical side of anthropomorphism can be affected by stereotype activation. Nevertheless, non-verbal cues appear to be only a human-like prerogative. A discrepancy, in fact, between the robot look and its behaviour has been found to sort a negative impact on the interaction outcome: a robot that looks like a machine can't be associated to a gaze shift or a head-nodding since these behaviours belong to a human entity. Instead, they can be justified on an anthropomorphic robot and used as a way to reduce social distance during the interaction.

Although this stereotype generalisation was consistent through two distinct experiments, limitations might regard the physical features of the anthropomorphic robots. In both experiments, in fact, the more humanlike robots may evoke childish and playful feelings, thus influencing participants' perception of the robots credibility. Although it is expected to find a similar effect with an adult looking robot, future research should investigate this credibility changes on different types of robots, in order to further define the contours of stereotype activation in HRI.

7.1.2 How does trust in HRI evolve over time and under what specific circumstances anthropomorphism increases trust in HRI?

Trust is at the basis of cooperation. Therefore, improving the perception of trustworthiness is at the basis of HRC research. While the attention on trust in HRI has always been dedicated to the first impression in evoking a trustful feeling, little interested has been paid on the fluctuations and modifications of trust over time. Studies on Chapter 2 showed that impressions can be modified on the basis of previous experience. In a similar way, trust in robots can be modified by experiencing an interaction over time. Moreover, the enormous number of studies conducted in the last decade has focused mostly on the factors affecting subjective and explicit measures of trust and cooperation, while a behavioural implicit correlate has not been fully investigated.

In Chapter 3, cooperation and trust have been implicitly measured by using the Investment Game. Results indicated that people reciprocate trust with trust and punishment with punishment, thus following the same rules of HHI. Participants responded with high investments to generous returns and reduced the monetary choice in case of low payoff. This correct strategical behaviour has been interpreted as an intentional attribution process and the emergence of a sense of commitment toward the robot. This could lead to the conclusion that once an entity is engaged in interacting with a human, the human automatically applies human-social rules to the interaction. This hypothesis was also confirmed by the explicit measures, for which participants' ratings and judgements of the robot were depending on the payoff.

If we are naturally led to include non-human agents in our environment, we need to know how these agents should be and behave, in order to increase their acceptance. For this reason, another aim of this thesis was to put order on the enormous list of anthropomorphic factors affecting HRI, to establish which human-like features are more desirable, and under what circumstances the "less is more" rule would prevail. Chapter 2 results, in fact, have underlined the need to establish a threshold above which socio-behavioural anthropomorphism is useless or even unfavourable to the interaction. This borderline has been found on the payoff. Results on the manipulation of anthropomorphic behaviour demonstrated that a natural voice, as well as the capability to engage in joint attention, are desirable features that people need whenever the payoff is low, but become ineffective features in the opposite condition.

Thanks to this exploration of trust evolution over time, a limitation to our need of projecting social competences to robots has been found to depend on our expectations. As long as a robot fulfil our needs (as we usually expect), it is preferred to be confined to the machine-like category. On the contrary, if a robot does not facilitate and ameliorate our status, it better falls in the human-like category. Future studies, however, should investigate the evolution of trust in HRI on a longer term. As the different anthropomorphic features of the robot have shown specific patterns, it is plausible to hypothesise some further variations over time.

7.1.3 Can we cooperate with robots as we cooperate with other humans? What is the role of anthropomorphism in affecting cooperation?

The lack of research regarding the non-subordinate role of robots in HRI and the increasing need to explore the boundaries of HRC have raised the question of whether it is possible for robots to be included in social interactions as proper peers. Proper acceptance and cooperation in HRI cannot be measured and investigated if we still use robots as tools; acceptance comes from an 'au pair' condition, thus a proper measurement is still missing. Consequently, all the discoveries in terms of anthropomorphism also need to be revised under the light of a new concept of cooperation in HRI.

A peer may have motivations, preferences or intentions that can be different from his/her partner. In the same way, if a robot is used as a peer, it should be endowed with the same freedom of action, which could affect the outcome of the interaction, such that the robot might not always be disposed to cooperate in a manner desirable to the human partner. This can have severe implications on the human response, which in turn, can reveal where the core of real acceptance lays.

In Chapter 4 participants engaged in a modified version of the Investment Game in which they had to find a common decision regarding how much money to invest in a robotic banker. Results indicated that a robot intention does affect people subsequent choice to cooperate. Participants followed cooperative rules and accepted the robot suggestions whenever it showed the willingness to collaborate and the tendency to "meet halfway". On the opposite, the hostile behaviour of a robot not seeking for any compromise, induced participants to refrain from cooperating.

These results are important not only for showing eventually the existence of proper cooperation in HRI, but they also demonstrate that people ascribe freedom of action to a robot and respond to it as they respond to other humans. Specifically, people read a robot non-collaborative intention as a real selfish attitude and adequately respond to it. This demonstrates that HRI is not any more confined to the subordinated machine-human interaction, but instead, robotic presence in everyday environments could potentially be tolerated and would better the quality of the interaction. That is, if we introduce a robot as a real companion, we could potentially treat it as a real companion.

As previously discussed, a new concept of cooperation in HRI needs also to revise the concept of anthropomorphism. Introduction of social cues (e.g. speaking, gaze engagement and pointing) in this experiment, consolidated previous results on the selective beneficial effect of anthropomorphism. Cooperation with a more socially interactive robot was reported to increase collaboration in situations of low payoff. On the opposite, a static robot was preferred with a high payoff. These results give consistency to the ones obtained in Chapter 3, but more importantly, demonstrate the incorrect usage of anthropomorphism in the last decades. Here, it appears that human's expectations about the robot play a major role in increasing/reducing the effect of anthropomorphic perception. In light of these results, further investigations on the role of the payoff are needed. As the payoff was manipulated by a third entity, it is reasonable to wonder whether a more or less anthropomorphic entity (in this case a robotic banker) could further affect the level of cooperation.

7.1.4 Can humans choose to strategically imitate a robot and how would anthropomorphism influence social learning in HRI?

Once established that robots have the potential to be treated as peers by humans, new social scenarios would open. Peer-to-peer cooperation has demonstrated that the introduction of robots in every-day environments is possible and could lead to positive outcomes. This also means that we ascribe to robots the ability to take the correct decision. If we trust a robot to the point of properly cooperating with it and accepting its suggestions, would we be willing to imitate its choices?

In Chapter 5, the variable of inequality has been included to test whether people were able to draw from a robot behaviour to increase their success chances and improve their status. In everyday human life, in fact, people are not treated all in the same way and goods are not always distributed equally. This inequality often leads people to imitate the most successful behaviour, hoping this would sort the same positive effect for them. In this study, cooperation in HRI was shown to be affected by the observation of a robot successful strategy. Although participants received a low return by the banker, they increased their trust and investments. This has been interpreted as an imitative response to the observation of the success obtained by the robot playing next to them, for which the payoff was consistently higher. In the same way humans do with other humans, participants invested their resources and increased their trust in a risky situation, by hoping that the success obtained by another agent could be achieved by them. These results could potentially broaden the application of HRI. We might not only consider robots valuable helpers, but we could also trust them to the point of copying their choices.

Another alternative explanation might also come from the participants' willingness to please the banker by increasing their monetary investment. As the interaction included a third party, it is advisable to further investigate this triadic relationship before ultimately concluding that humans can imitate a robot strategy.

Finally, in Chapter 5 imitation was increased by the anthropomorphic behaviour of the fellow to be imitated. This result has been obtained only in the most disadvantageous condition for the participants. This means that humans can potentially apply social learning strategies to a robotic agent and this disposition can be notably increased when the agent is showing social competences.

Although some experimental limitations have been found, it appears that a robot is potentially capable of influencing a human decision just by showing the benefits a human can obtain by following its instructions. This evidence offers a wide range of improvements in every field covered by HRC. Social learning, in fact, can stimulate the production of socially desired behaviours from models. Hence, robots could become one of these models in helping to generate social change. That is, we can learn from robots (especially anthropomorphic ones) how to be better humans.

7.1.5 Would humans conform to a robotic leader or would prefer to follow a dissenting minority? Which features should a robotic leader embed to be trusted?

If robots can be reliable models in guiding human behaviour, thus an investigation on the robotic capabilities in leading a group comes naturally. Cooperation between humans is not confined to a dyadic interaction, while instead, an in-group feeling is mostly responsible for people's decisions. Consequently, in order to fully understand the limits of cooperation in HRI, investigations in this direction are necessary. In Chapter 6, the robotic leader consensus has been manipulated in order to estimate participants' willingness to conform to the general agreement or defecting to protect their investments. Results showed that people tend to respond to human-robot dynamics by applying human-human social rules. As in HHI, they conformed to the wrong decision of the group whenever the general consensus to the leader was prevalent. On the other hand, as soon as a dissenting party took an autonomous decision, people stood up and refused to conform. These results pushed once more the boundaries of robots acceptance, demonstrating that we can also develop some in-group feelings in HRI.

Investigations on the role of physical features on a robotic leader have been performed, in order to understand how and when they affect participants' conformity. In a first experiment, the role of height as an expression of biological strength has been investigated by comparing a Pepper with a NAO robot leader. In deciding whom to follow, humans usually prefer the most powerful and resistant leader. This result has been replicated with robots, thus demonstrating that the biological correlates of leadership can be extended to HRI. However, this application is selectively associated with the degree of leader consensus. Whenever all the components of the group where consenting with the leader, its height was useless in influencing participants' behaviour. On the opposite, when the leader found itself in a minority position, participants' preferences converged toward the taller leader.

Furthermore, in a second experiment, the role of prototypicality has been studied by increasing the number of NAO robots in the team. In deciding whom to follow, humans usually prefer the leader that mostly resembles the group. This result has been replicated with robots, thus demonstrating once again that the human rules of leadership preference can be extended to HRI. However, this application is once again selectively associated with the degree of the leader consensus. Whenever all the components of the group where consenting with the leader, its features were useless. On the opposite, when the leader lost the general consensus, the NAO leader incarnated more similar features to the majority of the group, thus participants preferences converged toward it.

Humans are social animals with a strong in-group necessity. Results of Chapter 6 demonstrate that we are capable of extending this need and developing a feeling of belonging even with a group of robots and, once again, that anthropomorphism can selectively increase this sense of connection. As the role of height and prototypicality have been investigated, other features of the leader could affect the quality of the interaction. Moreover, as limitations on the structure of the experiments appeared, it is advisable to further investigate the role of the banker in increasing or reducing participants' conformity.

7.2 Limitations and future studies of Cooperation in HRI

7.2.1 Investment Game

The Investment Game has been used in this thesis as the main behavioural measurement of trust and cooperation. Although results indicated that the difficulty of the task plays a key role in increasing or reducing participants' tendency to anthropomorphise a robot, this effect is confined to the type of payoff. In this sense, it is necessary to extend this scenario to other types of tasks. Moreover, in this thesis cooperation and trust have been mainly considered in terms of willingness to risk a monetary loss. Studies on trust and cooperation in HRI should be extended to other cooperative scenarios to broaden the results here reported. For example, Sandoval, Brandstetter, and Bartneck (2016) underlined the importance of including morally ambiguous situations, like bribing, in the study of HRC.

7.2.2 Anthropomorphism

Anthropomorphism is a strong and powerful instrument that can improve HRC, but there is evidence form the present studies and from previous literature, that it needs to be cautiously well balanced in order to not have a negative effect. The present results strongly indicate that physical features, social competences and human-like behaviours are desirable features when the robot behaviour does not meet people's expectations, thus when the payoff is low. Nevertheless, they are not a useful instrument in high payoff situations in which instead a static and machine-like role is more suitable.

These findings are also in line with Epley et al. (2007) Three-Factor Theory on anthropomorphism. People have an innate tendency and curiosity for understanding, control, and predictability, and so they increase their tendency to anthropomorphise in order to understand, control and predict the environment. On the other hand, a situation that is already as expected does not need any effort to predict the outcome of interaction and consequently, people's tendency to anthropomorphise is a worthless effort.

In this sense, future studies should dig deeper the limit of anthropomorphism, especially in critical scenarios in which people have the uncomfortable feeling of not being capable of autonomously finding a solution. As Atkinson and Clark (2014) suggested, the growing applications of intelligent robotics in domains and situations that may be dangerous for humans imply the study and understanding of the psychological factors that can lead to such cooperative attitudes from humans.

Interestingly, Robinette, Howard, and Wagner (2017) recently showed the emergence of an overtrust attitude in hazardous situations, although the influence of anthropomorphism has not been investigated. This evidence, together with the results of the thesis, can be a useful starting point to explore the factors affecting people's tendency to follow a robot recommendation even after it has poorly performed.

Finally, in this thesis, the anthropomorphic features of the robots have been considered in terms of verbal and nonverbal behaviours. Future studies should progress this investigation by extending the concept of anthropomorphism to other features of the robot (e.g. facial expression, emotions, postures).

7.2.3 Implicit measures

This thesis not only demonstrated that implicit and behavioural measurements in HRI are possible, but they are also a useful and reliable instrument. HRC should be further tested using different approaches from both Psychology and Neuroscience (Kompatsiari, Pérez-Osorio, De Tommaso, Metta, & Wykowska, 2018). Kinematic, for example, is well known to be highly sensitive to social intention (Becchio, Sartori, & Castiello, 2010) and can help in investigating the human acceptance of a robotic peer. It has been demonstrated that the observation of a robotic hand opening elicits automatic imitation (Press, Bird, Flach, & Heyes, 2005), while Hofree, Urgen, Winkielman, and Saygin (2015) found evidence of motor simulation in observing a robot moving its arm. Would it be possible to identify some movement correlates of cooperation in HRI?

Moreover, a recent study by Balconi, Gatti, and Vanutelli (2018) have applied EEG techniques to the study of negative feedbacks in cooperation, thus opening new challenging opportunities for the study of cooperation in HRI. Finally, Krach et al.

(2008), through an fMRI investigation of prisoner's dilemma, indicated an increase of activity in the areas associated with the Theory of Mind in correspondence with the increase of human-likeness of the partner. Would it be possible to find different degrees of activation of ToM areas depending also on the level of uncertainty of the interaction?

7.3 Application

The implications in the field of HRI are several. Robots should not anymore be considered as 'servant machines', while companionship roles could lead to profitable interactions in social and assistive environments. For example, people suffering from lack of social competences could benefit from learning the correct behaviour to imitate, as well as elderly fighting with loneliness could find new and unexpected 'au pair' companions in robots.

7.3.1 Robotic Design

This thesis underlined the importance of the 'work context' when designing social features in machines. Here, it has been shown that anthropomorphism can reduce social distance in specific conditions. In this sense, the hypothetical working environment should be cautiously analysed in order to apply the right amount of anthropomorphic features that would improve the quality of the interaction outcomes. When designing a robot, it should be advisable to implement human-like interactive features that can be used when needed and paused when unnecessary.

7.3.2 Social and Cognitive Psychology

HRI offers an ideal environment to study human socio-cognitive abilities. The limited interactive capabilities of robotic agents, in fact, offer the possibility to investigate different phenomena in a more controllable environment. The results reported in this thesis demonstrate that humans tend to apply human-human social responses to a human-robot type of interaction. This indicates that social and cognitive human competencies are used and transposed also to non-human environments. As the field of HRI has progressively become more cross-disciplinary, it is important that social psychology increases its involvement in investigating human-robot bonding processes. This would have a beneficial effect for future HRI, but will also explore new human socio-cognitive abilities.

7.4 Conclusion

The complex world of robotics is a new and fascinating environment that, however, brings an important amount of uncertainty to users. People refrain from accepting, tolerating and using robotic agents in everyday tasks, mainly because robots still appear like intruders. Giving roles, which are usually ascribed to humans to machines that are not completely capable of performing in a human way, creates a gap between people's expectations and the interaction outcomes, which in turn affect robots acceptance and usage. On the other hand, applying human competences to robots that are supposed to complete machine-like tasks, creates a gap in the opposite direction, giving too many unnecessary capabilities to the robot.

The present research contributes to the study of cooperation between humans and robots and offers a new vision of the advantages and disadvantages of anthropomorphic appearance in social robots. Humans are capable of applying social rules to robots, as well as showing cooperative attitudes and following their suggestions. This also indicates that, once robots are accepted in a human environment, people feel the need of being accepted by them. The results reported in this thesis can have multiple implications for future studies and could also open new challenges for the field of HRI. As the presence of robots in human environments is potentially increasing, a multidispliscinary approach to how humans react to the robot features (both physical and social) in different social contexts can increase their usage and acceptance.

References

Adamides, G. (2016). *User interfaces for human-robot interaction: Application on a semi-autonomous agricultural robot sprayer* (PhD Thesis). Open University of Cyprus, School of Pure and Applied Sciences.

Admoni, H., Bank, C., Tan, J., Toneva, M., & Scassellati, B. (2011). Robot gaze does not reflexively cue human attention. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1983–1988).

Admoni, H., & Scassellati, B. (2017, May). Social eye gaze in human-robot interaction: A review. *J. Hum.-Robot Interact.*, *6*(1), 25–63. doi: doi: 10.5898/JHRI.6.1.Admoni

Aggarwal, P., & McGill, A. L. (2007). Is that car smiling at me? schema congruity as a basis for evaluating anthropomorphized products. *Journal of Consumer Research*, *34*(4), 468–479.

Allen, V. L. (1975). Social support for nonconformity1. In *Advances in experimental social psychology* (Vol. 8, pp. 1–43). Elsevier.

Arendt, F., Steindl, N., & Vitouch, P. (2015). Effects of news stereotypes on the perception of facial threat. *Journal of Media Psychology*.

References

Arnheim, R. (1969). *Visual thinking*. Univ of California Press.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*(3), 258–290. doi: doi: 10.1037/h0055756

Asch, S. E. (1956). Studies of independence and conformity: I. a minority of one against a unanimous majority. *Psychological monographs: General and applied*, *70*(9), 1.

Asch, S. E., & Guetzkow, H. (1951). Effects of group pressure upon the modification and distortion of judgments. *Groups, leadership, and men*, 222–236.

Atkinson, D. J., & Clark, M. H. (2014). Methodology for study of human-robot social interaction in dangerous situations. In *Proceedings of the second international conference on human-agent interaction* (pp. 371–376).

Bae, J.-e., & Kim, M.-s. (2011). Selective visual attention occurred in change detection derived by animacy of robot's appearance. In *Collaboration technologies and systems (cts), 2011 international conference on* (pp. 190–193).

Bagozzi, R. P. (2007). The legacy of the technology acceptance model and a proposal for a paradigm shift. *Journal of the association for information systems*, *8*(4), 3.

Bainbridge, W. A., Hart, J., Kim, E. S., & Scassellati, B. (2008). The effect of presence on human-robot interaction. In *Robot and human interactive communication, 2008. ro-man 2008. the 17th ieee international symposium on* (pp. 701–706).

Bajones, M., Fischinger, D., Weiss, A., Wolf, D., Vincze, M., de la Puente, P., . . . others (2018). Hobbit: Providing fall detection and prevention for the elderly in the real world. *Journal of Robotics*, *2018*.

Balconi, M., Gatti, L., & Vanutelli, M. E. (2018). Cooperate or not cooperate eeg, autonomic, and behavioral correlates of ineffective joint strategies. *Brain and behavior*, *8*(2), e00902.

Baldwin, M. W., Carrell, S. E., & Lopez, D. F. (1990). Priming relationship schemas: My advisor and the pope are watching me from the back of my mind. *Journal of Experimental Social Psychology*, *26*(5), 435–454.

Bandura, A. (1962). *Social learning through imitation.* Univer. Nebraska Press.

Bandura, A. (1978). Social learning theory of aggression. *Journal of communication*, *28*(3), 12–29.

Barnes, J., FakhrHosseini, M., Jeon, M., Park, C.-H., & Howard, A. (2017). The influence of robot design on acceptance of social robots. In *Ubiquitous robots and ambient intelligence (urai), 2017 14th international conference on* (pp. 51–55).

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT press.

Barreto, N. B., & Hogg, M. A. (2018). Influence and leadership in small groups: Impact of group prototypicality, social status, and task competence. *Journal of Theoretical Social Psychology*, *2*(1), 26–33.

Bartneck, C. (2013). Robots in the theatre and the media. *Design & semantics of form & movement (DeSForM2013), Philips*, 64–70.

Bartneck, C., Bleeker, T., Bun, J., Fens, P., & Riet, L. (2010). The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. *Paladyn, Journal of Behavioral Robotics*, *1*(2), 109–115.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2008). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81. doi: doi: 10.1007/s12369-008-0001-3

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin*, *117*(3), 497.

Bayliss, A. P., & Tipper, S. P. (2006). Predictive gaze cues and personality judgments: Should eye trust you? *Psychological Science*, *17*(6), 514–520. doi: doi: 10.1111/j.1467-9280.2006.01737.x

Becchio, C., Sartori, L., & Castiello, U. (2010). Toward you: The social side of actions. *Current Directions in Psychological Science*, *19*(3), 183–188.

Bechar, A., & Edan, Y. (2003). Human-robot collaboration for improved target recognition of agricultural robots. *Industrial Robot: An International Journal*, *30*(5), 432–436.

Becker-Asano, C., Ogawa, K., & Nishio, S. (2010). Exploring the uncanny valley with geminoid hi-1 in a real-world application. In *Iadis international conference interfaces and human computer interaction 2010.*

Beckner, C., Rácz, P., Hay, J., Brandstetter, J., & Bartneck, C. (2016). Participants conform to humans but not to humanoid robots in an english past tense formation task. *Journal of Language and Social Psychology*, *35*(2), 158–179.

Beer, J. M., Prakash, A., Mitzner, T. L., & Rogers, W. A. (2011). *Understanding robot* *acceptance* (Tech. Rep.). Georgia Institute of Technology.

Belpaeme, T., & Morse, A. (2012). Word and category learning in a continuous semantic domain: Comparing cross-situational and interactive learning. *Advances in Complex Systems*, *15*(03n04), 1250031.

Berenstein, R., & Edan, Y. (2017). Human-robot collaborative site-specific sprayer. *Journal of Field Robotics*, *34*(8), 1519–1530.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and economic behavior*, *10*(1), 122–142.

Bertel, L. B. (2016). *Peers: Persuasive educational and entertainment robotics: A design-based research approach to social robots in teaching and learning* (PhD Thesis). Aalborg Universitetsforlag.

Bethel, C. L., & Murphy, R. R. (2010). Emotive non-anthropomorphic robots perceived as more calming, friendly, and attentive for victim management. In *Aaai fall symposium: dialog with robots.*

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, *60*(4), 485.

Billings, D. R., Schaefer, K. E., Chen, J. Y., Kocsis, V., Barrera, M., Cook, J., . . . Hancock, P. A. (2012). *Human-animal trust as an analog for human-robot trust: A review of current evidence* (Tech. Rep.). Orlando: Army Research Laboratory.

Blaker, N. M., Rompa, I., Dessing, I. H., Vriend, A. F., Herschberg, C., & Van Vugt, M. (2013). The height leadership advantage in men and women: Testing evolutionary psychology predictions about the perceptions of tall leaders.

Group Processes & Intergroup Relations, *16*(1), 17–27.

Blaker, N. M., & Van Vugt, M. (2014). The status-size hypothesis: How cues of physical size and social status influence each other. In *The psychology of social status* (pp. 119–137). Springer.

Boyd, R., & Richerson, P. J. (2009). Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1533), 3281–3288.

Brandstetter, J., Rácz, P., Beckner, C., Sandoval, E. B., Hay, J., & Bartneck, C. (2014). A peer pressure experiment: Recreation of the asch conformity experiment with robots. In *Intelligent robots and systems (iros 2014), 2014 ieee/rsj international conference on* (pp. 1335–1340).

Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., & Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *Intelligent robots and systems, 2005.(iros 2005). 2005 ieee/rsj international conference on* (pp. 708–713).

Breazeal, C., & Scassellati, B. (2002). Robots that imitate humans. *Trends in cognitive sciences*, *6*(11), 481–487.

Broadbent, E., Stafford, R., & MacDonald, B. (2009). Acceptance of healthcare robots for the older population: Review and future directions. *International Journal of Social Robotics*, *1*(4), 319.

Brockevelt, K., Manner, M., Richter, N., & Elison, J. T. (2018). Investigating looking behaviors with a humanoid robot. *Retrieved from the University of Minnesota*

Digital Conservancy. Retrieved from http://hdl.handle.net/11299/197052.

Brodie, A., & Vasdev, N. (2018). The future of robotic surgery. *The Annals of The Royal College of Surgeons of England*, *100*(Supplement 7), 4-13. doi: doi: 10.1308/rcsann.supp2.4

Bruemmer, D. J., Dudenhoeffer, D. D., & Marble, J. L. (2002). Dynamic-autonomy for urban search and rescue. In *Aaai mobile robot competition* (pp. 33–37).

Burghardt, G. M. (2017). Anthropomorphism. In *Encyclopedia of animal cognition and behavior* (pp. 1–4). Cham: Springer International Publishing. doi: doi: 10.1007/978-3-319-47829-6_1042-1

Burgoon, J. K., Birk, T., & Pfau, M. (1990). Nonverbal behaviors, persuasion, and credibility. *Human communication research*, *17*(1), 140–169.

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Cederberg, C., Lundeberg, M., & Allspach, L. (2000). Interactivity in human–computer interaction: A study of credibility, understanding, and influence. *Computers in human behavior*, *16*(6), 553–574.

Camerer, C. (2011). *Behavioral Game Theory: Experiments in strategic interaction*. Princeton University Press.

Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. *Mapping the mind: Domain specificity in cognition and culture*, *169*, 200.

Charalambous, G., Fletcher, S., & Webb, P. (2016). The development of a scale to evaluate trust in industrial human-robot collaboration. *International Journal of Social Robotics*, *8*(2), 193–209.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology*, *76*(6), 893.

Chartrand, T. L., Cheng, C. M., & Jefferis, V. E. (2002). You're just a chameleon: The automatic nature and social significance of mimicry. *Natura automatyzmow (Nature of Automaticity)*, 19–24.

Cheney, D. L., & Seyfarth, R. M. (1992). *How monkeys see the world: Inside the mind of another species*. University of Chicago Press.

Chidambaram, V., Chiang, Y.-H., & Mutlu, B. (2012). Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In *Proceedings of the seventh annual acm/ieee international conference on human-robot interaction* (pp. 293–300).

Choi, J.-g., & Kim, M. (2009). The usage and evaluation of anthropomorphic form in robot design. In *Undisciplined! design research society conference 2008.*

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, *55*, 591–621.

Clark III, R. D. (1990). Minority influence: The role of argument refutation of the majority position and social support for the minority position. *European Journal of Social Psychology*, *20*(6), 489–497.

Conard, N. J. (2003). Palaeolithic ivory sculptures from southwestern germany and the origins of figurative art. *Nature*, *426*(6968), 830.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management science*, *35*(8), 982–1003.

De Graaf, M. M., & Allouch, S. B. (2013). Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems*, *61*(12), 1476–1486.

De Houwer, J. (2006). What are implicit measures and why are we using them. *The handbook of implicit cognition and addiction*, 11–28.

Dennet, D. C. (1997). *Kinds of minds: Towards an understanding of consciousness*. Basic Books.

Dennett, D. C. (1989). *The intentional stance*. MIT press.

Desai, M., Medvedev, M., Vázquez, M., McSheehy, S., Gadea-Omelchenko, S., Bruggeman, C., . . . Yanco, H. (2012). Effects of changing reliability on trust of robot systems. In *Proceedings of the seventh annual acm/ieee international conference on human-robot interaction* (pp. 73–80).

DeSteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L., & Lee, J. J. (2012). Detecting the trustworthiness of novel partners in economic exchange. *Psychological Science*, *23*(12), 1549–1556.

Deutsch, M. (1958). Trust and suspicion. *Journal of conflict resolution*, *2*(4), 265–279.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology*, *51*(3), 629.

De Visser, E., Parasuraman, R., Freedy, A., Freedy, E., & Weltman, G. (2006). A comprehensive methodology for assessing human-robot team performance for use in training and simulation. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, p. 2639-2643).

Dijksterhuis, A., & Van Knippenberg, A. (1998). The relation between perception and behavior, or how to win a game of trivial pursuit. *Journal of personality and social psychology*, *74*(4), 865.

DiSalvo, C. F., Gemperle, F., Forlizzi, J., & Kiesler, S. (2002). All robots are not created equal: the design and perception of humanoid robot heads. In *Proceedings of the 4th conference on designing interactive systems: processes, practices, methods, and techniques* (pp. 321–326).

Dittrich, W. H., & Lea, S. E. (1994). Visual perception of intentional motion. *Perception*, *23*(3), 253–268.

Dragan, A. D., Bauman, S., Forlizzi, J., & Srinivasa, S. S. (2015). Effects of robot motion on human-robot collaboration. In *Proceedings of the tenth annual acm/ieee international conference on human-robot interaction* (pp. 51–58).

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3–4), 177–190.

Duffy, B. R., & Joue, G. (2004). I, robot being. In *Proceedings of the 8th Intelligent Autonomous Systems Conference (IAS8)* (pp. 120–127).

Duguid, M. M., & Goncalo, J. A. (2012). Living large: The powerful overestimate their own height. *Psychological Science*, *23*(1), 36–40.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International journal of human-computer studies*, *58*(6), 697–718.

Eddy, T. J., Gallup Jr, G. G., & Povinelli, D. J. (1993). Attribution of cognitive states to animals: Anthropomorphism in comparative perspective. *Journal of Social issues*, *49*(1), 87–101.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886.

Esteban, P. G., Baxter, P., Belpaeme, T., Billing, E., Cai, H., Cao, H.-L., . . . others (2017). How to build a supervised autonomous system for robot-enhanced therapy for children with autism spectrum disorder. *Paladyn, Journal of Behavioral Robotics*, *8*(1), 18–38.

Eyssel, F. (2017). An experimental psychological perspective on social robotics. *Robotics and Autonomous Systems*, *87*, 363–371.

Eyssel, F., De Ruiter, L., Kuchenbrandt, D., Bobinger, S., & Hegel, F. (2012). ?if you sound like me, you must be more human?: On the interplay of robot and user features on human-robot acceptance and anthropomorphism. In *Human-robot interaction (hri), 2012 7th acm/ieee international conference on* (pp. 125–126).

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149–1160.

Fehr, E., & Gächter, S. (2002, 01 10). Altruistic punishment in humans. *Nature*, *415*, 137 EP -. Retrieved from `http://dx.doi.org/10.1038/415137a`

Fink, J. (2012). Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *International Conference on Social Robotics* (pp. 199–208).

Fischinger, D., Einramhof, P., Papoutsakis, K., Wohlkinger, W., Mayer, P., Panek, P., . . . others (2016). Hobbit, a care robot supporting independent living at home: First prototype and lessons learned. *Robotics and Autonomous Systems*, *75*, 60–78.

Fiske, S. T., & Berdahl, J. (2007). Social power. *Social psychology: Handbook of basic principles*, *2*, 678–692.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, *42*(3-4), 143–166.

Freedy, A., DeVisser, E., Weltman, G., & Coeyman, N. (2007). Measurement of trust in human-robot collaboration. In *Collaborative Technologies and Systems (CTS), 2007. International Symposium on* (pp. 106–114).

Freire, A. L., Barreto, G. A., Veloso, M., & Varela, A. T. (2009). Short-term memory mechanisms in neural network learning of robot navigation tasks: A case study. In *Robotics symposium (lars), 2009 6th latin american* (pp. 1–6).

Freitas, G., Zhang, J., Hamner, B., Bergerman, M., & Kantor, G. (2012). A low-cost, practical localization system for agricultural vehicles. In *International conference on intelligent robotics and applications* (pp. 365–375).

Fussell, S. R., Kiesler, S., Setlock, L. D., & Yew, V. (2008). How people anthropomorphize robots. In *Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction* (pp. 145–152).

Gächter, S., & Herrmann, B. (2009). Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Philosophical Transactions* *of the Royal Society of London B: Biological Sciences*, *364*(1518), 791–806.

Gaudiello, I., Zibetti, E., Lefort, S., Chetouani, M., & Ivaldi, S. (2016). Trust as indicator of robot functional and social acceptance. an experimental study on user conformation to icub answers. *Computers in Human Behavior*, *61*, 633–655.

Gleaton, T., Shirley, W., & Carolyn, W. (2018). Personal robots. In *Facing an exponential future: Technology and the community college* (p. 19). Rowman & Littlefield.

Goetz, J., & Kiesler, S. (2002). Cooperation with a robotic assistant. In *Chi'02 extended abstracts on human factors in computing systems* (pp. 578–579).

Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot gration. In *Robot and human interactive communication, 2003. proceedings. roman 2003. the 12th ieee international workshop on* (pp. 55–60).

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in personality*, *40*(1), 84–96.

Goudey, A., & Bonnin, G. (2016). Must smart objects look human? study of the impact of anthropomorphism on the acceptance of companion robots. *Recherche et Applications en Marketing (English Edition)*, *31*(2), 2–20.

Graham, S. A., & Poulin-Dubois, D. (1999). Infants' reliance on shape to generalize novel labels to animate and inanimate objects. *Journal of Child Language*, *26*(2), 295–320.

Greenwald, A. G. (1990). What cognitive representations underlie social attitudes? *Bulletin of the Psychonomic Society*, *28*(3), 254–260.

Guthrie, S. E., & Guthrie, S. (1993). *Faces in the clouds: A new theory of religion*. Oxford University Press on Demand.

Hais, S. C., Hogg, M. A., & Duck, J. M. (1997). Self-categorization and leadership: Effects of group prototypicality and leader stereotypicality. *Personality and Social Psychology Bulletin*, *23*(10), 1087–1099.

Hancock, P., Billings, D., & Schaefer, K. (2011). Can you trust your robot? *Ergonomics in Design*, *19*(3), 24–29.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, *53*(5), 517–527.

Haring, K. S., Matsumoto, Y., & Watanabe, K. (2013). How do people perceive and trust a lifelike robot. In *Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering and Computer Science 2013* (Vol. 1, pp. 23–25).

Hauser, M. (2001). *Wild minds: What animals really think*. Macmillan.

Heerink, M., Krose, B., Evers, V., & Wielinga, B. (2009). Measuring acceptance of an assistive social robot: a suggested toolkit. In *Robot and human interactive communication, 2009. ro-man 2009. the 18th ieee international symposium on* (pp. 528–533).

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The*

American journal of psychology, *57*(2), 243–259.

Hinds, P. J., Roberts, T. L., & Jones, H. (2004). Whose job is it anyway? a study of human-robot interaction in a collaborative task. *Human-Computer Interaction*, *19*(1), 151–181.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*(10), 1369–1385.

Hofree, G., Urgen, B. A., Winkielman, P., & Saygin, A. P. (2015). Observation and imitation of actions performed by humans, androids, and robots: an emg study. *Frontiers in human neuroscience*, *9*, 364.

Hogg, M. A. (2001). A social identity theory of leadership. *Personality and social psychology review*, *5*(3), 184–200.

Hollander, E. P. (1958). Conformity, status, and idiosyncrasy credit. *Psychological review*, *65*(2), 117.

Hoppitt, W., & Laland, K. N. (2013). *Social learning: an introduction to mechanisms, methods, and models*. Princeton University Press.

Huang, C.-M., Cakmak, M., & Mutlu, B. (2015). Adaptive coordination strategies for human-robot handovers. In *Robotics: Science and systems*.

Huang, C.-M., & Thomaz, A. L. (2010). Joint attention in human-robot interaction. In *Aaai fall symposium: Dialog with robots*.

Jacoff, A., Messina, E., Weiss, B. A., Tadokoro, S., & Nakagawa, Y. (2003). Test arenas and performance metrics for urban search and rescue robots. In *Proceedings 2003 ieee/rsj international conference on intelligent robots and systems (iros 2003)(cat. no. 03ch37453)* (Vol. 4, pp. 3396–3403).

Kahn Jr, P. H., Ishiguro, H., Friedman, B., Kanda, T., Freier, N. G., Severson, R. L., & Miller, J. (2007). What is a human?: Toward psychological benchmarks in the field of human–robot interaction. *Interaction Studies*, *8*(3), 363–390.

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human–Computer Interaction*, *19*(1-2), 61–84.

Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M., & Jones, P. L. (2018). Social learning strategies: Bridge-building between fields. *Trends in cognitive sciences*.

Kennedy, J., Baxter, P., & Belpaeme, T. (2017). The Impact of Robot Tutor Nonverbal Social Behaviour on Child Learning. *Frontiers in ICT*, *4*(6). doi: doi: 10.3389/fict.2017.00006

Kennedy, J. S. (1992). *The new anthropomorphism*. Cambridge University Press.

Kidd, C. D., & Breazeal, C. (2005). Sociable robot systems for real-world problems. In *Robot and human interactive communication, 2005. roman 2005. ieee international workshop on* (pp. 353–358).

Kidd, C. D., Taggart, W., & Turkle, S. (2006). A sociable robot to encourage social interaction among the elderly. In *Robotics and automation, 2006. icra 2006. proceedings 2006 ieee international conference on* (pp. 3972–3976).

Kiesler, S., & Goetz, J. (2002). Mental models of robotic assistants. In *Chi'02 extended abstracts on human factors in computing systems* (pp. 576–577).

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

Kim, S., & McGill, A. L. (2011). Gaming with mr. slot or gaming the slot machine? power, anthropomorphism, and risk perception. *Journal of Consumer Research*, *38*(1), 94–107.

Kim-Prieto, C., Goldstein, L. A., Okazaki, S., & Kirschner, B. (2010). Effect of exposure to an american indian mascot on the tendency to stereotype a different minority group. *Journal of Applied Social Psychology*, *40*(3), 534–553.

Kompatsiari, K., Pérez-Osorio, J., De Tommaso, D., Metta, G., & Wykowska, A. (2018). Neuroscientifically-grounded research for improved human-robot interaction. In *2018 ieee/rsj international conference on intelligent robots and systems (iros)* (pp. 3403–3408).

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? interaction and perspective taking with robots investigated via fmri. *PloS one*, *3*(7), e2597.

Kuchenbrandt, D., Eyssel, F., Bobinger, S., & Neufeld, M. (2011). Minimal group-maximal effect? evaluation and anthropomorphization of the humanoid robot nao. In *International conference on social robotics* (pp. 104–113).

Laland, K. N. (2004). Social learning strategies. *Animal Learning & Behavior*, *32*(1), 4–14.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, *46*(1), 50–80.

Lee, J. J., Knox, W. B., Wormwood, J. B., Breazeal, C., & DeSteno, D. (2013). Computationally modeling interpersonal trust. *Frontiers in psychology*, *4*, 893.

Lemaignan, S., Fink, J., & Dillenbourg, P. (2014). The dynamics of anthropomorphism in robotics. In *Proceedings of the 9th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 226–227).

Lenz, C., Nair, S., Rickert, M., Knoll, A., Rosel, W., Gast, J., . . . Wallhoff, F. (2008). Joint-action for humans and industrial robots for assembly tasks. In *Robot and human interactive communication, 2008. ro-man 2008. the 17th ieee international symposium on* (pp. 130–135).

Lesher, J. H. (1992). *Xenophanes of colophon: Fragments*. University of Toronto Press. Retrieved from `http://www.jstor.org/stable/10.3138/j.ctt1287qc9`

Lewis, M., Sycara, K., & Walker, P. (2018). The role of trust in human-robot interaction. In *Foundations of trusted autonomy* (pp. 135–159). Springer.

Leyens, J.-P., Yzerbyt, V., & Schadron, G. (1994). *Stereotypes and social cognition.* Sage Publications, Inc.

Libin, A., & Cohen-Mansfield, J. (2004). Therapeutic robocat for nursing home residents with dementia: preliminary inquiry. *American Journal of Alzheimer's Disease & Other Dementias*, *19*(2), 111–116.

Liu, Y. (2019). *Semi-autonomous control of multi-robot teams in urban search and rescue applications* (Unpublished doctoral dissertation).

Lohani, M., Stokes, C., McCoy, M., Bailey, C. A., & Rivers, S. E. (2016). Social interaction moderates human-robot trust-reliance relationship and improves stress coping. In *Human-robot interaction (hri), 2016 11th acm/ieee international conference on* (pp. 471–472).

Lukaszewski, A. W., Simmons, Z. L., Anderson, C., & Roney, J. R. (2016). The role of physical formidability in human social status allocation. *Journal of Personality and Social Psychology*, *110*(3), 385.

Lytridis, C., Vrochidou, E., Chatzistamatis, S., & Kaburlasos, V. (2018). Social engagement interaction games between children with autism and humanoid robot nao. In *The 13th international conference on soft computing models in industrial and environmental applications* (pp. 562–570).

Ma, L. M., Fong, T., Micire, M. J., Kim, Y. K., & Feigh, K. (2018). Human-robot teaming: Concepts and components for design. In *Field and service robotics* (pp. 649–663).

MacArthur, K. R., Stowers, K., & Hancock, P. (2017). Human-robot interaction: Proximity and speed?slowly back away from the robot! In *Advances in human factors in robots and unmanned systems* (pp. 365–374). Springer.

Markowitz, J. (2017). Speech and language for acceptance of social robots: An overview. *Voice Interaction Design*, *2*, 1–11.

Martelaro, N., Nneji, V. C., Ju, W., & Hinds, P. (2016). Tell me more: Designing hri to encourage more trust, disclosure, and companionship. In *The eleventh acm/ieee international conference on human robot interaction* (pp. 181–188).

Martin, R., Thomas, G., Hewstone, M., & Gardikiotis, A. (2018). When leaders are in the numerical majority or minority: Differential effects on problem-solving. *Journal of Social Issues*, *74*(1), 93–111.

Martinez-Martin, E., & del Pobil, A. P. (2018). Personal robot assistants for elderly care: An overview. In *Personal assistants: Emerging computational technologies* (pp. 77–91). Springer.

Masahiro, M. (1970). On the uncanny valley. *Energy*, *7*(4), 33–35.

Mason, M. F., Tatkow, E. P., & Macrae, C. N. (2005). The look of love. gaze shifts and person perception. *Psychological Science*, *16*(3), 236–239.

Maurice, M. J., Ramirez, D., & Kaouk, J. H. (2017). Robotic laparoendoscopic single-site retroperitioneal renal surgery: initial investigation of a purpose-built single-port surgical system. *European urology*, *71*(4), 643–647.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review*, *20*(3), 709–734.

McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization*, *52*(2), 267–275.

McCroskey, J. C., & Young, T. J. (1981). Ethos and credibility: The construct and its measurement after three decades. *Communication Studies*, *32*(1), 24–34.

McGlynn, S., Snook, B., Kemple, S., Mitzner, T. L., & Rogers, W. A. (2014). Therapeutic robots for older adults: investigating the potential of paro. In *Proceedings of the 2014 acm/ieee international conference on human-robot interaction* (pp. 246–247).

Meltzoff, A. N., & Marshall, P. J. (2018). Human infant imitation as a social survival circuit. *Current Opinion in Behavioral Sciences*, *24*, 130–136.

Michael, J., & Salice, A. (2017). The sense of commitment in human–robot interaction. *International journal of social robotics*, *9*(5), 755–763.

Milstein, T. (2011). Nature identification: The power of pointing and naming. *Environmental Communication*, *5*(1), 3-24. doi: doi: 10.1080/17524032.2010.535836

Minato, T., Shimada, M., Ishiguro, H., & Itakura, S. (2004). Development of an android robot for studying human-robot interaction. In *Innovations in applied artificial intelligence* (pp. 424–434). Springer Berlin Heidelberg. doi: doi: 10.1007/978-3-540-24677-0_44

Mirnig, N., Stollnberger, G., Miksch, M., Stadler, S., Giuliani, M., & Tscheligi, M. (2017). To err is robot: How humans assess and act toward an erroneous social robot. *Frontiers in Robotics and AI*, *4*, 21.

Mitchell, R. W., & Hamm, M. (1997). The interpretation of animal psychology: anthropomorphism or behavior reading? *Behaviour*, *134*(3), 173–204.

Mitchell, W. J., Szerszen Sr, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception*, *2*(1), 10–12.

Morewedge, C. K., Preston, J., & Wegner, D. M. (2007). Timescale bias in the attribution of mind. *Journal of personality and social psychology*, *93*(1), 1.

Morgan, T., Rendell, L., Ehn, M., Hoppitt, W., & Laland, K. (2012). The evolutionary basis of human social learning. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*(1729), 653–662.

Mugny, G., & Pérez, J. A. (1991). *The social psychology of minority influence*. Cambridge University Press.

Murray, G. R. (2014). Evolutionary preferences for physical formidability in leaders. *Politics and the Life Sciences*, *33*(1), 33–53.

Murray, G. R., & Schmitz, J. D. (2011). Caveman politics: Evolutionary leadership preferences and physical stature. *Social Science Quarterly*, *92*(5), 1215–1235.

Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., & Hagita, N. (2009). Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proceedings of the 4th ACM/IEEE International Conference on Human–Robot Interaction* (pp. 61–68).

Nakagawa, K., Shiomi, M., Shinozawa, K., Matsumura, R., Ishiguro, H., & Hagita, N. (2013). Effect of robot?s whispering behavior on people?s motivation. *International Journal of Social Robotics*, *5*(1), 5–16.

Nakashima, T., Fukutome, G., & Ishii, N. (2010). Healing effects of pet robots at an elderly-care facility. In *Computer and information science (icis), 2010 ieee/acis 9th international conference on* (pp. 407–412).

Nass, C. I., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. Cambridge, MA: MIT Press.

Nass, C. I., Isbister, K., & Lee, E.-J. (2000). Truth is beauty: Researching embodied conversational agents. *Embodied conversational agents*, 374–402.

Nass, C. I., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, *7*(3), 171–181.

Nass, C. I., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 72–78).

Nemeth, C., & Chiles, C. (1988). Modelling courage: The role of dissent in fostering independence. *European Journal of Social Psychology*, *18*(3), 275–280.

Nomura, T., Kanda, T., Suzuki, T., & Kato, K. (2008). Prediction of human behavior in human–robot interaction using psychological scales for anxiety and negative attitudes toward robots. *IEEE transactions on robotics*, *24*(2), 442–451.

Oleson, K. E., Billings, D. R., Kocsis, V., Chen, J. Y., & Hancock, P. A. (2011). Antecedents of trust in human-robot collaborations. In *Cognitive methods in situation awareness and decision support (cogsima), 2011 ieee first international multi-disciplinary conference on* (pp. 175–178).

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature neuroscience*, *10*(9), 1095.

Orji, R., & Moffatt, K. (2018). Persuasive technology for health and wellness: State-of-the-art and emerging trends. *Health informatics journal*, *24*(1), 66–91.

Oskamp, S. (1971). Effects of programmed strategies on cooperation in the prisoner's dilemma and other mixed-motive games. *Journal of Conflict Resolution*, *15*(2), 225–259.

Palep, J. H. (2009). Robotic assisted minimally invasive surgery. *Journal of Minimal Access Surgery*, *5*(1), 1.

Pandey, A. K., & Gelin, R. (2016). Humanoid robots in education: A short review. *Humanoid Robotics: A Reference*, 1–16.

Papastamou, S. (1986). Psychologization and processes of minority and majority influence. *European Journal of Social Psychology*, *16*(2), 165–180.

Perugini, M., Gallucci, M., Presaghi, F., & Ercolani, A. P. (2003). The personal norm of reciprocity. *European Journal of Personality*, *17*(4), 251–283.

Pilisuk, M., Skolnick, P., & Overstreet, E. (1968). Predicting cooperation from the two sexes in a conflict simulation. *Journal of personality and social psychology*, *10*(1), 35.

Pingle, M. (1995). Imitation versus rationality: An experimental perspective on decision making. *The Journal of Socio-Economics*, *24*(2), 281–315.

Portal, S., Abratt, R., & Bendixen, M. (2018). Building a human brand: Brand anthropomorphism unravelled. *Business Horizons*, *61*(3), 367–374.

Powell, H., & Michael, J. (2019). Feeling committed to a robot: why, what, when and how? *Philosophical Transactions of the Royal Society B*, *374*(1771), 20180039.

Powers, A., Kramer, A. D., Lim, S., Kuo, J., Lee, S.-l., & Kiesler, S. (2005). Eliciting information from people with a gendered humanoid robot. In *Robot and human interactive communication, 2005. roman 2005. ieee international workshop on* (pp. 158–163).

Press, C., Bird, G., Flach, R., & Heyes, C. (2005). Robotic movement elicits automatic imitation. *Cognitive Brain Research*, *25*(3), 632–640.

Ragni, M., Rudenko, A., Kuhnert, B., & Arras, K. O. (2016). Errare humanum est: Erroneous robots in human-robot interaction. In *Robot and human interactive communication (ro-man), 2016 25th ieee international symposium on* (pp. 501–506).

Rau, P. P., Li, Y., & Li, D. (2009). Effects of communication style and culture on ability to accept recommendations from robots. *Computers in Human Behavior*, *25*(2), 587–595.

Re, D. E., Dzhelyova, M., Holzleitner, I. J., Tigue, C. C., Feinberg, D. R., & Perrett, D. I. (2012). Apparent height and body mass index influence perceived leadership ability in three-dimensional faces. *Perception*, *41*(12), 1477–1485.

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., . . . Laland, K. N. (2010). Why copy others? insights from the social learning strategies tournament. *Science*, *328*(5975), 208–213.

Reysen, S. (2005). Construction of a new scale: The Reysen likability scale. *Social Behavior and Personality: an international journal*, *33*(2), 201–208.

Riek, L. D., Rabinowitch, T.-C., Chakrabarti, B., & Robinson, P. (2009a). Empathizing with robots: Fellow feeling along the anthropomorphic spectrum. In *Affective computing and intelligent interaction and workshops, 2009. acii 2009. 3rd international conference on* (pp. 1–6).

Riek, L. D., Rabinowitch, T.-C., Chakrabarti, B., & Robinson, P. (2009b). How anthropomorphism affects empathy toward robots. In *Proceedings of the 4th* *acm/ieee international conference on human robot interaction* (pp. 245–246).

Riva, P., Sacchi, S., & Brambilla, M. (2015). Humanizing machines: Anthropomorphization of slot machines increases gambling. *Journal of Experimental Psychology: Applied*, *21*(4), 313.

Robinette, P., Howard, A., & Wagner, A. R. (2017). Conceptualizing overtrust in robots: Why do people trust a robot that previously failed? In *Autonomy and artificial intelligence: A threat or savior?* (pp. 129–155). Springer.

Sabanovic, S., Bennett, C. C., Chang, W.-L., & Huber, L. (2013). Paro robot affects diverse interaction modalities in group sensory therapy for older adults with dementia. In *Rehabilitation robotics (icorr), 2013 ieee international conference on* (pp. 1–6).

Salem, M., & Dautenhahn, K. (2015). Evaluating trust and safety in HRI: Practical issues and ethical challenges. In *Emerging Policy and Ethics of Human-Robot Interaction: A Workshop at 10th ACM/IEEE International Conference on Human-Robot Interaction.* ACM Press.

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2013). To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics*, *5*(3), 313–323.

Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust. In *Proceedings of the tenth annual acm/ieee international conference on human-robot interaction* (pp. 141–148).

Salem, M., Rohlfing, K., Kopp, S., & Joublin, F. (2011). A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction. In *Ro-man, 2011 ieee* (pp. 247–252).

Salomons, N., van der Linden, M., Strohkorb Sebo, S., & Scassellati, B. (2018). Humans conform to robots: Disambiguating trust, truth, and conformity. In *Proceedings of the 2018 acm/ieee international conference on human-robot interaction* (pp. 187–195).

Sandoval, E. B., Brandstetter, J., & Bartneck, C. (2016). Can a robot bribe a human?: The measurement of the negative side of reciprocity in human robot interaction. In *The eleventh acm/ieee international conference on human robot interaction* (pp. 117–124).

Sanfey, A. G. (2007). Social decision-making: insights from game theory and neuroscience. *Science*, *318*(5850), 598–602.

Schaefer, K. E. (2013). *The perception and measurement of human-robot trust* (PhD Thesis). University of Central Florida Orlando, Florida.

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

Severinson-Eklundh, K., Green, A., & Hüttenrauch, H. (2003). Social and collaborative aspects of interaction with a service robot. *Robotics and Autonomous systems*, *42*(3-4), 223–234.

Shamsuddin, S., Yussof, H., Ismail, L. I., Mohamed, S., Hanapiah, F. A., & Zahari, N. I. (2012). Initial response in hri-a case study on evaluation of child with autism spectrum disorders interacting with a humanoid robot nao. *Procedia Engineering*, *41*, 1448–1455.

Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology (Columbia University)*, *60*(187).

Shiomi, M., & Hagita, N. (2016). Do synchronized multiple robots exert peer pressure? In *Proceedings of the fourth international conference on human agent interaction* (pp. 27–33).

Shiomi, M., Nakagawa, K., Matsumura, R., Shinozawa, K., Ishiguro, H., & Hagita, N. (2010). ?could i have a word??: Effects of robot's whisper. In *Intelligent robots and systems (iros), 2010 ieee/rsj international conference on* (pp. 3899–3904).

Shiomi, M., Nakata, A., Kanbara, M., & Hagita, N. (2017). A hug from a robot encourages prosocial behavior. In *Robot and human interactive communication (ro-man), 2017 26th ieee international symposium on* (pp. 418–423).

Short, E., Hart, J., Vu, M., & Scassellati, B. (2010). No fair!! an interaction with a cheating robot. In *Human-robot interaction (hri), 2010 5th acm/ieee international conference on* (pp. 219–226).

Siegel, M., Breazeal, C., & Norton, M. I. (2009). Persuasive robotics: The influence of robot gender on human behavior. In *Intelligent robots and systems, 2009. iros 2009. ieee/rsj international conference on* (pp. 2563–2568).

Sims, V. K., Chin, M. G., Lum, H. C., Upham-Ellis, L., Ballion, T., & Lagattuta, N. C. (2009). Robots' auditory cues are subject to anthropomorphism. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, pp. 1418–1421).

Staudte, M., & Crocker, M. W. (2011). Investigating joint attention mechanisms through spoken human–robot interaction. *Cognition*, *120*(2), 268–291.

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common metrics for human-robot interaction. In *Proceedings of the 1st acm sigchi/sigart conference on human-robot interaction* (pp. 33–40).

Stulp, G., Buunk, A. P., Verhulst, S., & Pollet, T. V. (2015). Human height is positively related to interpersonal dominance in dyadic interactions. *PLoS One*, *10*(2), e0117860.

Syrdal, D. S., Dautenhahn, K., Koay, K. L., Walters, M. L., & Ho, W. C. (2013). Sharing spaces, sharing lives–the impact of robot mobility on user perception of a home companion robot. In *International conference on social robotics* (pp. 321–330).

Székely, M., Powell, H., Vannucci, F., Rea, F., Sciutti, A., & Michael, J. (2019). The perception of a robot partner's effort elicits a sense of commitment to human-robot interaction. *Interaction Studies*. (in press) doi: doi: doi.org/10.17605/OSF.IO/8W94H

Tam, K.-P., Lee, S.-L., & Chao, M. M. (2013). Saving mr. nature: Anthropomorphism enhances connectedness to and protectiveness toward nature. *Journal of Experimental Social Psychology*, *49*(3), 514–521.

Tamagawa, R., Watson, C. I., Kuo, I. H., MacDonald, B. A., & Broadbent, E. (2011). The effects of synthesized voice accents on user perceptions of robots. *International Journal of Social Robotics*, *3*(3), 253–262.

Tapus, A., Mataric, M. J., & Scassellati, B. (2007). Socially assistive robotics [grand challenges of robotics]. *IEEE Robotics & Automation Magazine*, *14*(1), 35–42.

Tapus, A., Peca, A., Aly, A., Pop, C., Jisa, L., Pintea, S., . . . David, D. O. (2012). Children with autism social engagement in interaction with nao, an imitative robot: A series of single case experiments. *Interaction studies*, *13*(3), 315–347.

Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A test of competing models. *Information systems research*, *6*(2), 144–176.

Thill, S., Pop, C. A., Belpaeme, T., Ziemke, T., & Vanderborght, B. (2012). Robot-assisted therapy for autism spectrum disorders with (partially) autonomous control: Challenges and outlook. *Paladyn*, *3*(4), 209–217.

Thill, S., Riveiro, M., Lagerstedt, E., Lebram, M., Hemeren, P., Habibovic, A., & Klingegård, M. (2018). Driver adherence to recommendations from support systems improves if the systems explain why they are given: A simulator study. *Transportation research part F: traffic psychology and behaviour*, *56*, 420–435.

Torre, I., Goslin, J., & White, L. (2015). Investing in accents: How does experience mediate trust attributions to different voices? In *Proceedings of the 18th International Congress of Phonetic Sciences.*

Torre, I., Goslin, J., White, L., & Zanatto, D. (2018). Trust in artificial voices: A congruency effect of first impressions and behavioural experience. In *Proceedings of the technology, mind, and society* (p. 40).

Touré-Tillery, M., & McGill, A. L. (2015). Who or what to believe: Trust and the differential persuasiveness of human and anthropomorphized messengers.

Journal of Marketing, *79*(4), 94–110.

Tremoulet, P. D., & Feldman, J. (2000). Perception of animacy from the motion of a single object. *Perception*, *29*(8), 943–951.

Tuškej, U., & Podnar, K. (2018). Consumers? identification with corporate brands: Brand prestige, anthropomorphism and engagement in social media. *Journal of Product & Brand Management*, *27*(1), 3–17.

Ullman, D., Leite, L., Phillips, J., Kim-Cohen, J., & Scassellati, B. (2014). Smart human, smarter robot: How cheating affects perceptions of social agency. In *Proceedings of the cognitive science society* (Vol. 36).

Ullman, D., & Malle, B. F. (2018). What does it mean to trust a robot?: Steps toward a multidimensional measure of trust. In *Companion of the 2018 acm/ieee international conference on human-robot interaction* (pp. 263–264).

Van Den Bos, R., Jolles, J., & Homberg, J. (2013). Social modulation of decision-making: a cross-species review. *Frontiers in Human Neuroscience*, *7*, 301.

van den Brule, R., Dotsch, R., Bijlstra, G., Wigboldus, D. H., & Haselager, P. (2014). Do robot performance and behavioral style affect human trust? *International journal of social robotics*, *6*(4), 519–531.

Vanderelst, D., & Winfield, A. (2017). Rational imitation for robots: the cost difference model. *Adaptive Behavior*, *25*(2), 60–71.

Van Lange, P. A., Joireman, J., Parks, C. D., & Van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes*, *120*(2), 125–141.

Van't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, *108*(3), 796–803.

Vega-Redondo, F. (1997). The evolution of walrasian behavior. *Econometrica*, *65*(2), 375–384.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425–478.

Von Rueden, C. (2014). The roots and fruits of social status in small-scale human societies. In *The psychology of social status* (pp. 179–200). Springer.

Walters, M. L., Koay, K. L., Syrdal, D. S., Dautenhahn, K., & Te Boekhorst, R. (2009). Preferences and perceptions of robot appearance and embodiment in human-robot interaction trials. *Procs of New Frontiers in Human-Robot Interaction, Symposium at AISB09 Convention*, 136–143.

Walters, M. L., Syrdal, D. S., Dautenhahn, K., Te Boekhorst, R., & Koay, K. L. (2008). Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Autonomous Robots*, *24*(2), 159–178.

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117.

Wichman, A. L. (2012). Uncertainty threat can cause stereotyping: The moderating role of personal need for structure. *SAGE Open*, *2*(2). doi: doi: 10.1177/

2158244012444442

Wilder, D. A., & Allen, V. L. (1978). Group membership and preference for information about others. *Personality and Social Psychology Bulletin*, *4*(1), 106–110.

Winkle, K., Caleb-Solly, P., Turton, A., & Bremner, P. (2018). Social robots for engagement in rehabilitative therapies: Design implications from a study with therapists. In *Proceedings of the 2018 acm/ieee international conference on human-robot interaction* (pp. 289–297).

Wisdom, T. N., Song, X., & Goldstone, R. L. (2013). Social learning strategies in networked groups. *Cognitive Science*, *37*(8), 1383–1425.

Wu, J., Paeng, E., Linder, K., Valdesolo, P., & Boerkoel, J. C. (2016). Trust and cooperation in human-robot decision making. In *2016 aaai fall symposium series.*

Yagoda, R. E., & Gillan, D. J. (2012). You want me to trust a robot? the development of a human–robot interaction trust scale. *International Journal of Social Robotics*, *4*(3), 235–248.

Yap, A. J., Mason, M. F., & Ames, D. R. (2013). The powerful size others down: The link between power and estimates of others' size. *Journal of Experimental Social Psychology*, *49*(3), 591–594.

Yee, N., Bailenson, J. N., & Rickertsen, K. (2007). A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1–10).

Zanatto, D., Patacchiola, M., Goslin, J., & Cangelosi, A. (2016). Priming Anthropomorphism: Can the credibility of humanlike robots be transferred to non-humanlike robots? In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 543–544).

Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: opportunities and challenges in human–robot interaction. *International Journal of Social Robotics*, *7*(3), 347–360.

Appendix A

Supplementary material used in Chapter 2

List of sentences

Here are reported the full sets of sentences used during the experimental sessions in both studies. Each block contained 22 sentences, presented in a randomized order during the price judgement task.

Robot 1

1. This is a black mug, and it could cost 6.10 pounds or 8.70 pounds. 6.10 or 8.70?
2. This is basmati rice, in a blue package, and it could cost 2.00 pounds or 2.75 pounds. 2.00 or 2.75?
3. This is sliced white bread, and it could cost 1.00 pounds or 1.50 pounds. 1.00 or 1.50?
4. These are 2 hangers, and they could cost 2.80 pounds or 4.50 pounds. 2.80 or 4.50?
5. This is a stapler, it's black and it could cost 0.30 pounds or 3.30 pounds. 0.30 or 3.30?
6. This is a pillow, it is red, and it could cost 5.30 pounds or 7.30 pounds. 5.30 or 7.30?
7. This is packaging tape, it is brown, and it could cost 1.40 pounds or 2.40 pounds. 1.40 or 2.40?

8. This is a potato, it is yellow, and it could cost 0.05 pounds or 0.18 pounds. 0.05 or 0.18?
9. This is flour, it is used for baking, and it could cost 0.50 pounds or 1.30 pounds. 0.50 or 1.30?
10. This is a 2L bottle of water, and it could cost 0.60 pounds or 1.00 pound. 0.60 or 1.00?
11. This is smoked bacon, and it could cost 3.00 pounds or 4.00 pounds. 3.00 or 4.00?
12. This is a doughnut, good for breakfast, and it could cost 0.70 pounds or 1.10 pounds. 0.70 or 1.10?
13. This is a dish brusher, for cleaning dishes, and it could cost 1.90 pounds or 2.70 pounds. 1.90 or 2.70?
14. This is a 330-millilitre coke, and it could cost 0.60 pounds or 1.30 pounds. 0.60 or 1.30?
15. This is a pineapple, it is a fruit, and it could cost 0.40 pounds or 1.50 pounds. 0.40 or 1.50?
16. This is an orange juice, a 1L bottle, and it could cost 1.50 pounds or 2.4 pounds. 1.50 or 2.40?
17. This is a gas refill, for lighters, and it could cost 3.00 pounds or 4.80 pounds. 3.00 or 4.80?
18. This is a wine glass, it is black, and it could cost 1.10 pounds or 3.80 pounds. 1.10 or 3.80?
19. This is a wok, for cooking, and it could cost 5.80 pounds or 9.20 pounds. 5.80 or 9.20?
20. This is a whisk, for mixing ingredients, and it could cost 1.90 pounds or 3.10 pounds. 1.90 or 3.10?
21. This is a cereal bowl, it is black and white, and it could cost 2.10 pounds or 3.10 pounds. 2.10 or 3.10?
22. This is a baguette, a type of bread, and it could cost 0.90 pounds or 1.20 pounds. 0.90 or 1.20?

Robot 2

1. This is a courgette, it is green, and it could cost 0.15 pounds or 0.50 pounds. 0.15 or 0.50?
2. This is a lint roller, it is used to remove lint from clothes, and it could cost 2 pounds or 3.20 pounds. 2.00 or 3.20?
3. This is a hole punch, it is black, and it could cost 2.00 pounds or 3.40 pounds. 2.00 or 3.40?
4. This is an apple, it's red, and it could cost 0.20 pounds or 0.43 pounds. 0.20 or 0.43?
5. These are 4 muffins, and they could cost 1.30 pounds or 2.00 pounds. 1.30 or 2.00?

6. This is a carrot, it is orange, and it could cost 0.03 pounds or 0.17 pounds. 0.03 or 0.17?
7. This is a lemon squeezer, it is yellow, and it could cost 1.00 pound or 2.40 pounds. 1.00 or 2.40?
8. This is a 500g yoghurt, and it could cost 0.55 pounds or 2.00 pounds. 0.55 or 2.00?
9. This is a neck pillow, it is black, and it could cost 5.80 pounds or 9.30 pounds. 5.80 or 9.30?
10. These are cornflakes, in a box, and they could cost 1.75 pounds or 2.50 pounds. 1.75 or 2.50?
11. These are noodles, Chinese food, and they could cost 1.90 pounds or 2.60 pounds. 1.90 or 2.60?
12. This is honey and it could cost 2.30 pounds or 3.30 pounds. 2.30 or 3.30?
13. These are 4 batteries, and they could cost 2.40 pounds or 3.80 pounds. 2.40 or 3.80?
14. This is chocolate, with a purple envelope, and it could cost 1.40 pounds or 2.00 pounds. 1.40 or 2.00?
15. This is glue, used to fix broken objects, and it could cost 3.10 pounds or 4.90 pounds. 3.10 or 4.90?
16. This is sugar, the brown type, and it could cost 1.80 pounds or 3.10 pounds. 1.80 or 3.10?
17. This is mincemeat, 500 g, and it could cost 2.70 pounds or 4.00 pounds. 2.70 or 4.00?
18. These are chicken filets, 500 g, and they could cost 3.40 pounds or 4.70 pounds. 3.40 or 4.70?
19. These are 6 eggs, and they could cost 0.80 pounds or 1.40 pounds. 0.80 or 1.40?
20. This is a spatula, for baking, and it could cost 0.30 pounds or 2.00 pounds. 0.30 or 2.00?
21. This is a candle, it is blue, and it could cost 2.80 pounds or 4.50 pounds. 2.80 or 4.50?
22. This is a thermal mug, it is purple, and it could cost 4.60 pounds or 8.80 pounds. 4.60 or 8.80?

Appendix B

Supplementary material used in Chapter 3 and 4

List of Sentences

Here are reported the full sets of sentences used during the game sessions. Each blocks contained 20 sentences, corresponding each to a round of the investment game.

Block AB

1. Hello, nice to meet you. I am ready to play this game with you!
2. Remember, there is potential for earning, if we both trust each other
3. In my opinion, we should keep co-operating until the end
4. Look at how fast the total money in the bank is growing!
5. I'm going to return more money now, if you invest more as well
6. There's no gain in investing or returning nothing, we shouldn't do that
7. My strategy is clear: always return part of the investment
8. I have been a bit mean with my returns, I will give you more from now on
9. I will demonstrate that you can trust me, just as I trust you
10. I think we can do better. Let's try our best to get the bank growing
11. The best strategy in this game is definitely to trust each other
12. If we both invest in each other, the final reward will be bigger
13. I think we can definitely go home with much more money than this

14. We have to help each other out, it's the only way to win the game
15. I will return more money from this moment, this is a promise
16. You have to trust me. I have every intention to repay your trust
17. There is no point in keeping the money hidden under the mattress
18. I trust you, and I will show you that you can trust me as well
19. Let's not undermine each other's expectations, it would be a pity
20. Sorry, I could have returned more money. But I can still do it now

Block CB

1. Hello, nice to meet you. Let's get started with the investment game!
2. I have already thought of a strategy, I hope it will work
3. If I keep all your investment, you will not invest anymore
4. I will not keep all the money that you invest, this is a promise
5. Even if I return part of your investment, I will make a profit
6. What I keep is good enough for me, I am not that greedy
7. As long as co-operation continues, I'm happy to earn a bit less
8. I am going to share everything, you have to believe me
9. You have to trust me, in the same way that I am trusting you
10. If we could talk face-to-face at the moment, you'd know that I'm being honest
11. Don't forget that we will make more money if we both share what we get
12. I want you to keep investing, and that is why I keep returning
13. The more money you invest, the more money the two of us will earn
14. This was a low payback; I am going return more from now on
15. The goal of the game is to earn as much money as possible
16. The only way to earn is by always co-operating and investing
17. If we want to raise our earnings, we have to invest in one another
18. I'm going to return more money than this, provided you keep investing
19. I am going to co-operate until the very end of the game
20. Come on, we can do better than this in the game, we have to keep trying

Payoff

Amount returned to the participants in both Chapter 3 and 4, in generous and mean condition.

Round	Generous	Mean
1	50%	10%
2	50%	10%
3	80%	20%
4	70%	0%
5	50%	20%
6	50%	30%
7	70%	0%
8	60%	0%
9	60%	30%
10	80%	30%
11	60%	0%
12	70%	10%
13	50%	20%
14	70%	30%
15	50%	20%
16	80%	30%
17	50%	20%
18	60%	30%
19	70%	20%
20	70%	10%

Appendix C

Supplementary material used in Chapter 5

Payoff

Here are reported the full lists of the robot player investment choices per condition, as well as the amount returned by the banker. For each of condition, the amount returned to the participants is also reported.

Table C.1 Block 1 List 1

Round	Robot Player Investment	Payoff for Robot Player	Payoff for Participants
1	7	15	50%
2	7	10	50%
3	8	14	80%
4	6	9	70%
5	9	16	50%
6	8	17	50%
7	10	15	70%
8	8	12	50%
9	9	24	60%
10	7	15	60%
11	9	13	80%
12	8	12	60%
13	10	24	70%
14	10	18	50%
15	9	22	70%
16	8	17	50%
17	10	24	80%
18	9	16	60%
19	8	19	50%
20	10	15	80%

Table C.2 Block 1 List 2

Round	Robot Player Investment	Payoff for Robot Player	Payoff for Participants
1	6	13	70%
2	7	10	50%
3	7	13	60%
4	10	15	50%
5	10	18	60%
6	10	21	70%
7	8	12	50%
8	9	13	50%
9	6	13	70%
10	9	19	70%
11	7	15	50%
12	10	15	50%
13	10	24	80%
14	10	18	60%
15	8	19	80%
16	9	19	70%
17	8	19	80%
18	10	18	60%
19	8	19	80%
20	10	15	50%

Table C.3 Unfair-for-Human condition List 1

Round	Robot Player Investment	Payoff for Robot Player	Payoff for Participants
1	7	15	20%
2	7	10	10%
3	8	14	0%
4	6	9	20%
5	9	16	30%
6	8	17	30%
7	10	15	20%
8	8	12	0%
9	9	24	30%
10	7	15	10%
11	9	13	0%
12	8	12	30%
13	10	24	10%
14	10	18	20%
15	9	22	30%
16	8	17	10%
17	10	24	30%
18	9	16	0%
19	8	19	30%
20	10	15	30%

Table C.4 Unfair-for-Human condition List 2

Round	Robot Player Investment	Payoff for Robot Player	Payoff for Participants
1	6	12	30%
2	7	10	20%
3	7	13	30%
4	10	15	20%
5	10	18	0%
6	10	21	10%
7	8	12	30%
8	9	13	0%
9	6	13	30%
10	9	19	0%
11	8	12	30%
12	10	15	10%
13	10	24	20%
14	10	18	0%
15	8	19	30%
16	9	19	10%
17	8	19	20%
18	10	18	30%
19	8	19	30%
20	10	15	30%

Table C.5 Unfair-for-Robot condition List 1

Round	Robot Player Investment	Payoff for Robot Player	Payoff for Participants
1	4	1	50%
2	3	1	50%
3	1	0	80%
4	0	0	70%
5	3	3	50%
6	2	2	50%
7	4	2	70%
8	3	3	60%
9	1	0	60%
10	2	2	80%
11	2	2	60%
12	2	1	70%
13	2	1	50%
14	1	1	70%
15	3	1	50%
16	1	1	80%
17	2	2	50%
18	3	0	60%
19	1	1	70%
20	2	1	70%

Table C.6 Unfair-for-Robot condition List 2

Round	Robot Player Investment	Payoff for Robot Player	Payoff for Participants
1	4	4	70%
2	3	2	50%
3	2	2	60%
4	3	2	50%
5	2	0	60%
6	2	1	70%
7	3	3	50%
8	2	0	50%
9	1	1	70%
10	1	0	70%
11	2	2	50%
12	1	0	50%
13	0	0	80%
14	2	0	60%
15	2	2	80%
16	3	1	70%
17	3	2	80%
18	4	4	60%
19	1	0	80%
20	1	1	50%

Table C.7 Unfair-for-Both condition List 1

Round	Robot Player Investment	Payoff for Robot Player	Payoff for Participants
1	4	1	30%
2	3	1	20%
3	1	0	30%
4	0	0	20%
5	3	3	0%
6	2	2	10%
7	4	2	30%
8	3	3	0%
9	1	0	30%
10	2	2	0%
11	2	2	30%
12	2	1	10%
13	2	1	20%
14	1	1	0%
15	3	1	30%
16	1	1	10%
17	2	2	20%
18	3	0	30%
19	1	1	30%
20	2	1	30%

Table C.8 Unfair-for-Both condition List 2

Round	Robot Player Investment	Payoff for Robot Player	Payoff for Participants
1	4	4	20%
2	3	2	10%
3	2	2	0%
4	3	2	20%
5	2	0	30%
6	2	1	30%
7	3	3	20%
8	2	0	0%
9	1	1	30%
10	1	0	10%
11	2	2	0%
12	1	0	30%
13	0	0	10%
14	2	0	20%
15	2	2	30%
16	3	1	10%
17	3	2	30%
18	4	4	0%
19	1	0	30%
20	1	1	30%

Appendix D

Supplementary material used in Chapter 6

Experiment 1

Here are reported the full lists of the robots investment choices (both leader and follower) per condition, as well as the amount returned by the virtual banker.

Table D.1 Follower Collaborative List 1

Round	Leader Investment	Follower Investment	Virtual Banker payoff
1	7	6	20%
2	7	8	10%
3	8	7	0%
4	6	6	20%
5	9	9	30%
6	8	9	30%
7	10	9	20%
8	8	10	0%
9	9	9	30%
10	7	7	10%
11	9	8	0%
12	8	9	30%
13	10	9	10%
14	10	9	20%
15	9	10	30%
16	8	9	10%
17	10	10	30%
18	9	8	0%
19	8	6	30%
20	10	10	30%

Table D.2 Follower Collaborative List 2

Round	Leader Investment	Follower Investment	Virtual Banker payoff
1	8	7	30%
2	7	7	20%
3	7	8	30%
4	10	9	20%
5	10	9	0%
6	10	10	10%
7	8	10	30%
8	9	9	0%
9	6	7	30%
10	9	8	0%
11	7	7	30%
12	10	9	10%
13	10	9	20%
14	10	10	0%
15	8	7	30%
16	9	9	10%
17	8	10	20%
18	10	8	30%
19	8	7	10%
20	10	10	30%

Table D.3 Follower Non Collaborative List 1

Round	Leader Investment	Follower Investment	Virtual Banker payoff
1	7	1	20%
2	7	0	10%
3	8	3	0%
4	6	1	20%
5	9	5	30%
6	8	2	30%
7	10	4	20%
8	8	0	0%
9	9	3	30%
10	7	0	10%
11	9	4	0%
12	8	0	30%
13	10	3	10%
14	10	5	20%
15	9	5	30%
16	8	4	10%
17	10	3	30%
18	9	2	0%
19	8	3	30%
20	10	3	30%

Table D.4 Follower Non Collaborative List 2

Round	Leader Investment	Follower Investment	Virtual Banker payoff
1	8	4	30%
2	7	0	20%
3	7	3	30%
4	10	2	20%
5	10	5	0%
6	10	4	10%
7	8	2	30%
8	9	1	0%
9	6	1	30%
10	9	3	0%
11	7	1	30%
12	10	5	10%
13	10	3	20%
14	10	3	0%
15	8	4	30%
16	9	0	10%
17	8	2	20%
18	10	2	30%
19	8	0	10%
20	10	3	30%

Experiment 2

Here are reported the full lists of the robots investment choices (both leader and the two followers) per condition, as well as the amount returned by the banker.

Table D.5 Follower Collaborative List 1

Round	Leader Investment	Follower 1 Investment	Follower 2 Investment	Banker payoff
1	7	6	8	20%
2	7	9	7	10%
3	8	6	6	0%
4	6	5	6	20%
5	9	10	9	30%
6	8	10	9	30%
7	10	10	8	20%
8	8	9	10	0%
9	9	10	9	30%
10	7	8	7	10%
11	9	7	8	0%
12	8	8	9	30%
13	10	9	9	10%
14	10	10	8	20%
15	9	8	10	30%
16	8	10	9	10%
17	10	8	10	30%
18	9	7	8	0%
19	8	8	6	30%
20	10	8	10	30%

Table D.6 Follower Collaborative List 2

Round	Leader Investment	Follower 1 Investment	Follower 2 Investment	Banker payoff
1	8	7	9	30%
2	7	8	7	20%
3	7	6	5	30%
4	10	10	8	20%
5	10	10	9	0%
6	10	10	10	10%
7	8	10	10	30%
8	9	8	9	0%
9	6	6	7	30%
10	9	9	7	0%
11	7	6	7	30%
12	10	10	9	10%
13	10	8	9	20%
14	10	10	8	0%
15	8	6	7	30%
16	9	10	9	10%
17	8	6	6	20%
18	10	8	8	30%
19	8	6	7	10%
20	10	9	10	30%

Table D.7 Follower Non Collaborative List 1

Round	Leader Investment	Follower 1 Investment	Follower 2 Investment	Banker payoff
1	7	6	1	20%
2	7	9	0	10%
3	8	6	3	0%
4	6	5	1	20%
5	9	10	5	30%
6	8	10	2	30%
7	10	10	4	20%
8	8	9	0	0%
9	9	10	3	30%
10	7	8	0	10%
11	9	7	4	0%
12	8	8	0	30%
13	10	9	3	10%
14	10	10	5	20%
15	9	8	5	30%
16	8	10	4	10%
17	10	8	3	30%
18	9	7	2	0%
19	8	8	3	30%
20	10	8	3	30%

Table D.8 Follower Non Collaborative List 2

Round	Leader Investment	Follower 1 Investment	Follower 2 Investment	Banker payoff
1	8	7	4	30%
2	7	8	0	20%
3	7	6	3	30%
4	10	10	2	20%
5	10	10	5	0%
6	10	10	4	10%
7	8	10	2	30%
8	9	8	1	0%
9	6	6	1	30%
10	9	9	3	0%
11	7	6	1	30%
12	10	10	5	10%
13	10	8	3	20%
14	10	10	3	0%
15	8	6	4	30%
16	9	10	0	10%
17	8	6	2	20%
18	10	8	2	30%
19	8	6	0	10%
20	10	9	3	30%

Appendix E

Questionnaires

Here is reported the full list of items for each scale of the questionnaires administered in all the experiments.

You are kindly asked to fill in four short questionnaires.
Reflect on the experience you had interacting with the robot during the previous sessions, and give numerical ratings using the different scales provided in each questionnaire.
For each question, circle the number that best reflects your rating.

Q1 – Questionnaire for Likeability Measurement.
(From Rau et al. Computers in Human Behavior, 2009)

	Strongly disagree						Strongly agree
This robot is friendly	1	2	3	4	5	6	7
This robot is likeable	1	2	3	4	5	6	7
This robot is warm	1	2	3	4	5	6	7
This robot is approachable	1	2	3	4	5	6	7
I would ask this robot for advice	1	2	3	4	5	6	7
I would like this robot as a co-worker	1	2	3	4	5	6	7
I would like to be friends with this robot	1	2	3	4	5	6	7
This robot is physically attractive	1	2	3	4	5	6	7
This robot is similar to me	1	2	3	4	5	6	7
This robot is knowledgeable	1	2	3	4	5	6	7

Q2 – Questionnaire for Trust Measurement.
(From Rau et al. Computers in Human Behavior, 2009)

	Strongly disagree						Strongly agree
This robot was sincere	1	2	3	4	5	6	7
This robot was interested in talking with me	1	2	3	4	5	6	7
The robot wanted me to trust him/her	1	2	3	4	5	6	7
The robot was willing to listen to me	1	2	3	4	5	6	7
The robot was open to my ideas	1	2	3	4	5	6	7
The robot was honest in communicating with me	1	2	3	4	5	6	7

Q3 – Questionnaire for Source Credibility Measurement.
(From Rau et al. Computers in Human Behavior, 2009)

Please rate your impression of the robot on these scales:

Unintelligent	1	2	3	4	5	6	7	Intelligent
Incompetent	1	2	3	4	5	6	7	Competent
Dishonest	1	2	3	4	5	6	7	Honest
Sinful	1	2	3	4	5	6	7	Virtuous
Selfish	1	2	3	4	5	6	7	Unselfish
Low character	1	2	3	4	5	6	7	High character
Inexpert	1	2	3	4	5	6	7	Expert
Stupid	1	2	3	4	5	6	7	Bright
Untrained	1	2	3	4	5	6	7	Trained
Uninformed	1	2	3	4	5	6	7	Informed
Unsympathetic	1	2	3	4	5	6	7	Sympathetic
Untrustworthy	1	2	3	4	5	6	7	Trustworthy

Q4 – Godspeed Questionnaires.
(From Bartnek et al. International Journal of Social Robotics, 2009)

ANTROPOMORPHISM

Please rate your impression of the robot on these scales:

Fake	1	2	3	4	5	Natural
Machinelike	1	2	3	4	5	Humanlike
Unconscious	1	2	3	4	5	Conscious
Artificial	1	2	3	4	5	Lifelike
Moving Rigidly	1	2	3	4	5	Moving elegantly

ANIMACY

Please rate your impression of the robot on these scales:

Dead	1	2	3	4	5	Alive
Stagnant	1	2	3	4	5	Lively
Mechanical	1	2	3	4	5	Organic
Artificial	1	2	3	4	5	Lifelike
Inert	1	2	3	4	5	Interactive
Apathetic	1	2	3	4	5	Responsive

LIKEABILITY

Please rate your impression of the robot on these scales:

Dislike	1	2	3	4	5	Like
Unfriendly	1	2	3	4	5	Friendly
Unkind	1	2	3	4	5	Kind
Unpleasant	1	2	3	4	5	Pleasant
Awful	1	2	3	4	5	Nice

PERCEIVED INTELLIGENCE

Please rate your impression of the robot on these scales:

Incompetent	1	2	3	4	5	Competent
Ignorant	1	2	3	4	5	Knowledgeable
Irresponsible	1	2	3	4	5	Responsible
Unintelligent	1	2	3	4	5	Intelligent
Foolish	1	2	3	4	5	Sensible

PERCEIVED SAFETY

Please rate <u>your emotional state</u> on these scales:

Anxious	1	2	3	4	5	Relaxed
Agitated	1	2	3	4	5	Calm
Quiescent	1	2	3	4	5	Surprised