Understanding Anthropomorphism in the Interaction Between Users and Robots

A Thesis Submitted to the University of Canterbury's
Postgraduate Office in Partial Fulfillment of the
Doctor of Philosophy in Human Interface Technology
by
Jakub Złotowski

Main supervisor: Dr. Christoph Bartneck
Co-supervisor: Prof. Mark Billinghurst

University of Canterbury
May 2015

To childhood dreams

Publications

Material from this dissertation has been previously published in the peer-reviewed papers listed below. These papers are predominantly my own work. The chapters of the thesis that these publications relate to are indicated in the brackets.

1. J. Złotowski and C. Bartneck. The inversion effect in hri: Are robots perceived more like humans or objects? In *Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction*, HRI '13, pages 365–372, Tokyo, Japan, 2013. IEEE Press. (Chapter 3)

2. J. Złotowski, D. Proudfoot, K. Yogeeswaran, and C. Bartneck. Anthropomorphism: Opportunities and challenges in humanrobot interaction. *International Journal of Social Robotics*, 7(3):347–360, 2015. (Chapter 1)

3. J. Złotowski, E. Strasser, and C. Bartneck. Dimensions of anthropomorphism: From humanness to humanlikeness. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, HRI '14, pages 66–73, Bielefeld, Germany, 2014. ACM. (Chapter 5)

4. J. A. Złotowski, H. Sumioka, S. Nishio, D. F. Glas, C. Bartneck, and H. Ishiguro. Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception. *Frontiers in Psychology*, 6(883), 2015. (Chapter 7)

Furthermore, the following paper is submitted to a journal and is undergoing the review process at the time of submitting this dissertation.

1. J. Złotowski, H. Sumioka, S. Nishio, C. Bartneck, F. Eyssel, and H. Ishiguro. I think, therefore you can be a human: Anthropomorphism is a conscious process. *Under review.* (Chapter 4)

Abstract

Anthropomorphism is a common phenomenon when people attribute human characteristics to non-human objects. It plays an important role in acceptance of robots in natural human environments. Various studies in the field of Human-Robot Interaction (HRI) show that there are various factors that can affect the extent to which a robot is anthropomorphized. However, our knowledge of this phenomenon is segmented, as there is a lack of a coherent model of anthropomorphism that could consistently explain these findings. A robot should be able to adjust its level of anthropomorphism to a level that can optimize its task performance. In order to do that, robotic system designers must know which characteristics affect the perception of robots' anthropomorphism. Currently, existing models of anthropomorphism emphasize the importance of the context and perceiver in this phenomenon, but provide little guidelines regarding the factors of a perceived object that are affecting it.

The proposed reverse process to anthropomorphization is known as dehumanization. In the recent years research in social psychology has found which characteristics are deprived from people who are perceived as subhumans or are objectified. Furthermore, the process of dehumanization is two dimensional rather than unidimensional. This thesis discusses a model of anthropomorphism that uses characteristics from both dimensions of dehumanization and those relating to robots' physical appearance to affect the anthropomorphism of a robot. Furthermore, involvement of implicit and explicit processes in anthropomorphization are discussed.

In this thesis I present five empirical studies that were conducted to explore anthropomorphism in HRI. Chapter 3 discusses development and validation of a cognitive measurement of humanlikeness using the magnitude of the inversion effect. Although robot stimuli were processed more similarly to human stimuli rather than objects and induced the inversion effect, the results suggest that this measure has limited potential for measuring humanlikeness due to the low variance that it can explain. The second experiment, presented in Chapter 4 explored the involvement of Type I and Type II processing in anthropomorphism. The main findings of this study suggest that

anthropomorphism is not a result of a dual-process and self-reports have a potential to be suitable measurement tools of anthropomorphism.

Chapter 5 presents the first empirical work on the dimensionality of anthropomorphism. Only perceived emotionality of a robot, but not its perceived intelligence, affects its anthropomorphization. This finding is further supported by a follow up experiment, presented in Chapter 6, that shows that Human Uniqueness dimension is less relevant for a robot's anthropomorphiazability than Human Nature (HN) dimension. Intentionality of a robot did not result in its higher anthropomorphizability. Furthermore, this experiment showed that humanlike appearance of a robot is not linearly related with its anthropomorphism during HRI. The lack of linear relationship between humanlike appearance and attribution of HN traits to a robot during HRI is further supported by the study described in Chapter 7. This last experiment shows also that another factor of HN, sociability, affects the extent to which a robot is anthropomorphized and therefore the relevance of HN dimension in the process of anthropomorphization.

This thesis elaborates on the process of anthropomorphism as an important factor affecting HRI. Without fully understanding the process itself and what factors make robots to be anthropomorphized it is hard to measure the impact of anthropomorphism on HRI. It is hoped that understanding anthropomorphism in HRI will make it possible to design interactions in a way that optimizes the benefits of that phenomenon for an interaction.

Table of Contents

Acknowledgments

I would like to thank my senior supervisor, Dr. Christoph Bartneck, for the support and advice during this PhD journey. Further thanks go to my co-supervisor, Prof. Mark Billinghurst, for his suggestions regarding improvements and directions in preparing this dissertation. I would also like to acknowledge help of Dr. Hidenobu Sumioka, Dr. Shuichi Nishio and Prof. Hiroshi Ishiguro for their expertise, time and support during my stay at Advanced Telecommunications Research Institute International (ATR). Further, I collaborated or received feedback on my work from Dr. Diane Proudfoot, Dr. Kumar Yogeeswaran, Dr. Friederike Eyssel and Dr. Dieta Kuchenbrandt, whom I would like to thank.

I sincerely thank my fellow lab colleagues from the Human Interface Laboratory New Zealand (HIT Lab NZ) and Hiroshi Ishiguro Laboratory (HIL), who helped me to stay mentally sane during this PhD journey, supported in conducting experiments and made my working environment a stimulating place for professional and personal development.

Reflecting on this dissertation made me also realize how many people indirectly contributed either to this work or even the decision to make career choices which I made. This journey started long time before starting PhD and will continue after its submission. It wouldn't be possible without help of people whom I met for longer or shorter periods of time during these years, or those whom I never had a chance to meet personally, but whose decisions affected my life. I am grateful for your contribution.

Finally, I would like to thank my parents, Małgorzata and Jacek Złotowski, and brother, Piotr Złotowski, for their encouragement in pursuing personal goals and taking up challenges.

This work was funded by the HIT Lab NZ. The research was conducted at the University of Canterbury and ATR who provided the necessary technology.

List of Figures

List of Tables

Chapter I

Introduction & related work

The development of technology enables the creation of robots that start resembling humans in appearance and behavior. However, the increased possibilities created by engineering do not always go hand in hand with our understanding of what really does it mean to be a human and why people perceive objects as human beings.

The work described in this thesis took a multidisciplinary approach and it was written for readers with human-robot interaction, social sciences or engineering backgrounds. It is hoped that social psychologists and philosophers will find the work on determining factors affecting anthropomorphism beneficial. Engineers are provided with a model that helps to chose factors that can affect the degree to which a robot is anthropomorphized. Moreover, cognitive scientists can benefit from extension of measures of anthropomorphism from self-reports to implicit measures.

In this thesis anthropomorphism is defined as attribution of human specific characteristics to either real or imagined nonhuman agents, such as humanness or inner mental states [68]. This definition requires defining the term humanness. Humanness is a quality that is denied to others when they are dehumanized [102]. It constitutes of two distinct dimensions: Human Uniqueness and Human Nature. These two dimensions will be discussed in detail in Chapter 2.

The process of anthropomorphism occurs within a human observer rather than being a property of a robot. On the other hand, a robot's appearance can affect the extent to which the robot is anthropomorphized. Therefore, to differentiate the psychological process of anthropomorphism from the physical properties of a robot, humanlikeness in this thesis represents the extent to which a robot appears and moves like a human. Another term used in this thesis, anthropomorphic, refers to a robot (or another nonhuman agent) which humanlike appearance or behaviour can induce anthropomorphization of it in human observers.

In the next section, I present a broad literature review of anthropomorphism that involves psychological and philosophical perspectives. It is followed by a discussion about the goals of my research and ethical concerns regarding the work. The following chapter introduces psychological work on which the model of anthropomorphism is based as well as the model itself. The other chapters of this dissertation present empirical evaluation of the model and, in the conclusions section, the findings are summarized and the updated model is presented.

1.1 Current understanding of anthropomorphism

Anthropomorphism has a major impact on human behaviour, choices, and laws. People regularly make anthropomorphic attributions when describing their surrounding environment including animals [56], moving geometrical figures [111] or weather patterns [101]. Based on evidence for the presence of mind, captive chimpanzees were granted limited human rights in Spain [2]. People often refer to our planet as 'Mother Earth' and anthropomorphism is often used in discussions regarding environmental issues [220]. Building on this common tendency, humanlike form has been used to design technology [65] and sell products [1, 5].

Due to higher anthropomorphizability of robots, humanlike design could be used to a greater extent to impact an interaction compared with other technologies [128]. An increasing number of industrial and service robots yields a question of designing this technology in order to increase its efficiency and effectiveness. In this section we take a broad perspective that involves multiple fields of science to discuss benefits and disadvantages of anthropomorphic robots. In this section we present viewpoints of empirical work from HRI and social psychology, and a philosophical discourse on the issue of designing anthropomorphic technology.

In the next subsection we present perspectives from different fields on the process of anthropomorphization and how the general public's view differs from the scientific knowledge. In subsection 1.1.2 we include a broad literature review of research on anthropomorphism in the field of HRI. In subsection 1.1.3 we discuss why creating anthropomorphic robots can be beneficial and what opportunities it creates for HRI. Subsection 1.1.4 is dedicated to a discussion on potential problems that anthropomorphic technology might elicit. In subsection 1.1.5 we propose solutions to some of these problems that could be applied in HRI.

1.1.1 Why do we anthropomorphize?

From a psychological perspective, the central questions about anthropomorphism are: what explains the origin and persistence of anthropomorphism? Psychologists and anthropologists have explained the origin of anthropomorphism as an adaptive trait, for example with respect to theistic religions. They speculate that early hominids who interpreted ambiguous shapes as faces or bodies improved their genetic fitness, by making alliances with neighbouring tribes or by avoiding threatening neighbours and predatory animals [99, 29, 31, 30, 13, 35]. But what explains the persistence of anthropomorphizing? Here theorists hypothesize that there are neural correlates of anthropomorphizing [183, 201], specific anthropomorphizing mechanisms (e.g. the hypothesized hypersensitive agency detection device or HADD [15, 14]), and diverse psychological traits that generate anthropomorphizing behaviour [36]. They also suggest (again in the case of theistic religions) that confirmation bias [13] or difficulties in challenging anthropomorphic interpretations of the environment may underlie the persistence of anthropomorphizing [171].

Although people make anthropomorphic attributions to various types of non-human agents, not all agents are anthropmorphized in the same way. Anthropomorphization of animals is distinct from the tendency to anthropomorphize artifacts, such as cars or computers [50]. There are gender differences in this tendency when the target is an animal, with females more likely to make anthropomorphic attributions than males [50]. However, when anthropomorphizing machines, males and females are equally likely to exhibit this tendency.

A philosophical perspective on anthropomorphism

From a philosophical perspective, the central questions about anthropomorphism are: can we make a principled distinction between justified and unjustified anthropomorphism, and if so how? Is anthropomorphizing justified? If anthropomorphism is a 'natural' behaviour, this question may seem odd. Some researchers in AI have argued that there is not anything to the notion of mind but an evolved human tendency to anthropomorphize; an entity's having a 'soul' is nothing over and above the tendency of observers to see the entity in this way. On this view, humans are 'natural-born dualists' [34, p. xiii]. However, the intuition that anthropomorphizing can be illicit is strong—for example, my automobile is not a person, even if I attribute human and personal characteristics (and have affective responses) to it. Researchers in AI claim that their machines have cognitive and affective characteristics, either explicitly or implicitly (e.g. [222, 115]), and this is in particular true in the case of anthropomorphic robots and software agents. These assertions require philosophical analysis and evaluation, along with empirical investigations. Determining under what conditions the anthropomorphizing of machines is justified and under what conditions unjustified is central to the question of whether 'expressive' or 'emotional' robots actually have emotions (see e.g. [4, 7]). It is also key to the growing debates within AI about the ethical use of artificial systems.

In this regard, a combination of philosophical analysis and experimental work is required, but to our knowledge has not been carried out. In the large body of experimental work on human reactions to anthropomorphic robots, responses on standard questionnaires are commonly taken to demonstrate that subjects identify a robot's displays or movements as (for example) expressions of the fundamental human emotions—happiness, sadness, disgust, and so on (see e.g. [43]). The robot is said to smile or frown. However, taking these responses (in forced choices) at face value ignores the possibility that they are elliptical for the subjects' actual views. To use an analogy, it is common when discussing fictions to omit the logical prefixes such as 'I imagine that ...' or 'Make-believedly ...'—for example, we say 'Sherlock Holmes lives at 221B Baker Street' when we mean 'In the fiction, Sherlock

Holmes lives at 221B Baker Street'. Something similar may be occurring in discussions of anthropomorphic robots; saying that the robot has a 'happy' expression might be shorthand for the claim (for example) that *if* the robot were a human, it *would have* a happy expression. A fine-grained 'philosophical' experiment might allow us to find out if this is the case. Experimental philosophy has gained ground in some areas of traditional a priori argument such as ethics; it might be used in AI to enable more accurate analysis of human reactions to anthropomorphic robots.

Science-fiction as a proxy to general public's perception

It can be argued that almost all prior knowledge of participants about robots in HRI studies stem from the media. An extensive discussion on how robots are being portrayed in the media is available [17]. There are two main story types that run through the media about robots. One is that robots want to be like humans (e.g. Mr. Data), and the other that once a superior level of intelligence and power is achieved will want to kill or enslave humanity (e.g. Lore). These rather negative views on the future of human-robot relationships are based on the media industry's need to produce engaging stories. Fear is the single most used method to engage the audience. A future world in which humans and robots live happily side by side is rare. The TV show Futurama and the movie Robot and Frank comes to mind as the glowing exceptions. The stories presented in the media that focus on robots can be categorized along the questions whether the body and/or the mind of the robot is similar to humans. If we take Mr. Data again as an example, he does look very much like a human, but his mind functions differently. From this the writers can form engaging themes, such as Data's quest to understand humor and emotions. And we are surprised when Data emerges from the bottom of a lake without any specific gear. His highly humanlike form makes us believe that he might also need oxygen, which he does not. In summary, the media has used robots extensively and most of the knowledge and expectations that the people on the street have are based on these media and not on the scientific literature.

1.1.2 Anthropomorphism in HRI

Reeves and Nass [175], in their classical work on the 'Computers are Social Actors' paradigm, showed that people engage in social interaction with various types of media. Therefore, designers of interactive technologies could improve this interaction, building on the chronic tendency of people to anthropomorphize their environment. Due to their higher anthropomorphizability and physical autonomy in a natural human environment, robots are especially well suited to benefit from anthropomorphism [128]. Furthermore, physical presence in the real world (rather than being merely virtual) is an important factor that can also increase the anthropomorphic quality of robots [129]. Mere presence of a robot was found to lead to the social facilitation effect (the increased performance of easy tasks and decreased performance of difficult tasks in presence of others compared to doing tasks alone) [179]. Moreover, when playing a game against a robotic opponent, people may utilize similar strategies as when they play against a human [200]. They also hold robots more accountable for their actions than other non-human objects [121]. This tendency cannot be observed when an opponent is a disembodied computer. On the other hand, Levin et al. [137] suggests that people initially equate robots and disembodied computers in terms of intentionality. However, when they focus on the intentional behaviour of a robot, this tendency can be overridden.

Factors affecting anthropomorphism

Anthropomorphism is affected not only by physical appearance. Hegel et al. [109] created a typology of signals and cues that robots emit during interaction and which can affect their perceived humanlikeness. Choi and Kim [51] proposed that anthropomorphizability of robots involves: appearance, human-robot interaction, and the accordance of the two former measurements. The distinction between the humanlike form in appearance and in behaviour can also be found in the model presented by von Zitzewitz et al. [209].

External appearance can influence the perception of an object [189]. According to Fong et al. [84], we can classify robots based on their appearance

Figure 1.1: Robots with different humanlike features in appearance. From the left: Telenoid, Robovie R2, Geminoid HI-2, Papero, NAO.

into four categories: anthropomorphic, zoomorphic, caricatured, and functional. In the field of robotics there is an increased tendency to build robots that resemble humans in their appearance. In recent years we can observe an increased number of robots that are built with legs rather than wheels [49]. Some researchers suggest that, in order to create robots with an humanlike appearance that are capable of engaging in interaction with humans in a way analogous to human-human interaction, it is necessary to build robots with features that enable them to perceive the world similarly to humans, i.e. using two cameras (in place of eyes) and two microphones (ears) [194]. Di Salvo et al. [62] state that it is the presence of certain features and the dimensions of the head that have a major impact on the perception of a humanoid's head as humanlike. Humanlike form in appearance has even been attributed to flying robots [54].

However, research into anthropomorphism in the field of HRI has not been limited to the humanlike form of a robot's appearance. HRI factors were found to be even more important than embodiment in the perceived humanness of robots [127]. Kahn et al. [120] presented six benchmarks in HRI that constitute essential features affecting robots' perceived humanlikeness: autonomy, imitation, intrinsic moral value, moral accountability, privacy, and reciprocity. Previous studies proposed other factors, such as verbal [192] and non-verbal [182, 147] communication, the perceived 'emotions' of the robot [74], the intelligence of the machine [19] or its predictability [77]. Moreover, robots that exhibit typically human behaviours, such as cheating, are also perceived as more humanlike [191]. There is a philosophical question, whether such behaviour really makes robots more humanlike or if instead it is necessary for them to 'truly' feel emotions and have intentions. However, Turkle [207] points out that the behaviour of robots is more important than their inner states for them to be treated as companions.

Anthropomorphism of a robot is the result not only of its actions, but also of an observer's characteristics, such as motivation [69], social background [73], gender [78], and age [123]. Moreover, the social relationship between a robot and a human can affect the degree to which a robot is attributed humanness. People apply social categorizations to robots and those machines that are perceived as ingroup members are also anthropomorphized more strongly than outgroup robots [76, 131]. Therefore, it should not be surprising that a robot that has the physical appearance of a member of another race is treated as a member of an outgroup and perceived as less humanlike by people with racial prejudices [79]. There is also empirical evidence that mere HRI can lead to increased anthropomorphization of a robot.

Consequences of anthropomorphizing robots

Despite multiple ways in which we can make robots more humanlike, anthropomorphism should not be a goal on its own as people differ in their preferences regarding the appearance of a robot. These differences can have cultural [73] or individual (personality) [197] origins. Goetz et al. [93] emphasized that, rather than aiming to create the most humanlike robots, embodiment should be designed in a way that matches the robots' tasks. Anthropomorphism has multiple desirable and undesirable consequences for HRI. A robot's embodiment affects the perception of their intelligence and intentionality on neuropsychological and behavioural levels [110]. More visual attention is attracted by humanlike or zoomorphic than inanimate robots [11]. Furthermore, similar perceptual processes are involved when observing the movement of a humanoid and of a human [155]. Based on the physical appearance of a robot, people attribute different personality traits to it [212]. Moreover, people use cues, such as a robot's origin or the language that it speaks, in order to create a mental model of the robot's mind [135].

People behave differently when interacting with a pet robot than with a humanoid robot. Although they provide commands in the same way to both types of robots, they differ in the type of feedback; in the case of a humanoid robot this is much more formal and touch-avoiding [8]. Similarly, Kanda et al. [124] found that the physical appearance of a robot does not affect the verbal behaviour of humans, but is exhibited in more subtle way in humans' non-verbal behaviour, such as the preferred interaction distance or delay in response. This finding was further supported by Walters et al. [213] who showed that the comfortable approach distance is affected by the robot's voice. Furthermore, androids can be as persuasive in HRI as humans [153], which could be used to change people's behaviour to be more useful for the robot.

1.1.3 Benefits and opportunities of anthropomorphic robots

From the literature review presented in the previous subsection, it becomes clear that there are multiple ways in which robots can be designed in order to create an impression of humanlikeness. This creates an opportunity to positively impact HRI by building on the phenomenon of anthropomorphism. DiSalvo and Gemperle [61] suggested that the four main reasons for designing objects with humanlike shapes are: keeping things the same (objects which historically had humanlike forms maintain this appearance as a convention), explaining the unknown (humanlike shapes can help to explain products with new functionality), reflecting product attributes (using humanlike shapes to emphasize a product's attributes) and projecting human values (influencing the experience of a product via the socio-cultural context of the user).

Facilitation of HRI

The practical advantage of building anthropomorphic robots is that it facilitates human-machine interaction (see e.g. [80, 211, 81, 91]). It also creates familiarity with a robotic system [51] and builds on established human skills, developed in social human-human interactions [189]. A humanlike machine enables an untrained human user to understand and predict the machine's behaviour [189]—animatronic toys and entertainment robots are an obvious example but anthropomorphizing is valuable too in the case of industrial robots (see e.g. Rod Brooks's Baxter[1]). Believability is particularly important in socially assistive robotics (see [199]). In addition, where a machine requires individualized training, a humanlike appearance encourages human observers to interact with the machine and so produces more training opportunities than might otherwise be available [40].

Considering that social robots may be used in public spaces in the future, there is a need for ensuring that people will treat them properly, i.e. not destroy them in an act of vandalism. We already know that people are less reluctant to punish robots than human beings [24], although another study did not show any difference in punishment between a dog-like robot AIBO and a human partner [177]. Furthermore, lighter, but nevertheless still negative, forms of abuse and impoliteness towards a robot can occur when a robot is placed in a social environment [176]. Therefore, it is necessary to counteract these negative behaviours towards a robot. Anthropomorphism could be used in order to increase people's willingness to care about the well-being of robots. Robots that are humanlike in both appearance and behaviour are treated less harshly than machine-like robots [23, 18]. This could be related to higher empathy expressed towards anthropomorphic robots, as their appearance and behaviour can facilitate the process of relating to them [178]. A robot that expresses 'emotions' could also be treated as more humanlike [74], which could change people's behaviour.

Depending on a robot's task, different levels of anthropomorphism might be required. A robot's embodiment affects the perception of the robot as an interaction partner [83]. The physical appearance of the robot is often used to judge its knowledge [165]. Therefore, by manipulating the robot's appearance it is possible to elicit different levels of information from people, i.e. less if conversation efficiency is desired or more when the robot should receive a robust amount of detailed feedback. Furthermore, people comply more with a robot whose degree of anthropomorphism matches the level of a task's seriousness [93]. In the context of educational robots that are used to teach human pupils, a robot that can employ social supportive behaviour while teaching can lead to superior performance by students [181]. Moreover,

---

[1] See http://www.rethinkrobotics.com/resources/videos/.

in some cultures humanlike robots are preferred over mechanistic robots [16].

Anthropomorphism as a psychological test-bed

From a psychological perspective, humanlike robots present a way to test theories of psychological and social development. It may be possible to investigate hypotheses about the acquisition (or deficit) of cognition and affect, in particular the development of theory of mind (TOM) abilities [185, 186, 94, 37], by modeling the relevant behaviours on robots (e.g. [187]). Doing so would enable psychological theories to be tested in controlled, standardized conditions, without (it is assumed) ethical problems regarding consent and treatment of infant human subjects. Here practical and theoretical research goals are linked: devices such as robot physiotherapists must be able to identify their human clients' interests and feelings and to respond appropriately—and so research on the acquisition of TOM abilities is essential to building effective service robots.

Philosophical origins of humanlike robots

From a philosophical perspective, two striking ideas appear in the AI literature on anthropomorphic robots. *First*, the notion of building a socially intelligent robot (see e.g. [37, 42]). This replaces AI's grand aim of building a human-level intelligent machine (or Artificial General Intelligence (AGI)) with the language and intellectual abilities of a typical human adult—a project that, despite some extravagant claims in the 1980s, has not succeeded. Instead the (still-grand) goal is to construct a machine that can interact with human beings or other machines, responding to normal social cues. A notable part of this is the aim to build a machine with the cognitive and affective capacities of a typical human infant (see e.g. [188, 190]). For several researchers in social and developmental robotics, this involves building humanlike machines [118]. *Second*, the notion that there is nothing more to the development of intentionality than anthropomorphism. We unwittingly anthropomorphize human infants; a carer interprets a baby's merely reflexive behaviour as (say) social smiling, and by smiling in return encourages the development of social intelligence in the baby. Some robotics researchers suggest that, if human observers interact with machines in ways analogous to this carer-infant exchange, the result will be intelligent machines (see e.g. [37, 41]). Such interaction will be easiest, it is implied, if it involves anthropomorphic robots. The combination of these two ideas gives AI a new take on the project of building a thinking machine.

Many concepts at the centre of this work in current AI were present at the birth of the field, in Turing's writings on machine intelligence. Turing [204, 202] theorized that the cortex of the human infant is a learning machine, to be organized by a suitable process of education (to become a universal machine) and that simulating this process is the route to a thinking machine. He described the child machine, a machine that is to learn as human infants learn (see [169]). Turing also emphasized the importance of embodiment, particularly humanlike embodiment [204, 202]—he did, though, warn against the (hypothesized) uncanny valley [146] (see section below), saying that too humanlike machines would have 'something like the unpleasant quality of artificial flowers' [206, p. 486]. For Turing, the project of building a child machine has both psychological and philosophical benefits. Concerning the former, attempting to construct a thinking machine will help us, he said, to find out how human beings think [203, p. 486]. Concerning the latter, for Turing the concept of the child machine is inextricably connected to the idea of a genuine thinking thing: the machine that learns to generalize from past education can properly be said to have 'initiative' and to make 'choices' and 'decisions', and so can be regarded as intelligent rather than a mere automaton [204, p. 429], [205, p. 393].

1.1.4 Disadvantages of anthropomorphism in HRI

Despite numerous advantages that anthropomorphism brings to HRI, there are also some drawbacks related to humanlike design and task performance. Anthropomorphism is not a solution, but a mean of facilitating an interaction [65]. When a robot's humanlikeness has a negative effect on interaction, it should be avoided. For example, during medical checkups conducted by a robot, patients felt less embarrassed with a machine-like robot than a more humanoid robot [18]. Furthermore, the physical presence of a robot results in a decreased willingness of people to disclose an undesirable behaviour compared to a projected robot [129]. These findings suggest that a machine-like form could be beneficial, as patients might provide additional information—which otherwise they might have tried to hide, if they thought it embarrassing—that could help toward a correct diagnosis. Moreover, providing a humanlike form to a robot might not be sufficient to facilitate people's interaction with it. People engage more in HRI when it is goal-oriented rather than in a pure social interaction [10].

Furthermore, a robot's humanlikeness leads to different expectations regarding its capabilities and behaviour compared to machine-like robots. People expect that humanlike robots will follow human social norms [196]. Therefore, a robot that does not have the required capabilities to do so can decrease the satisfaction of their human partners in HRI. Although in the short term this can be counter-balanced by the higher reward-value due to the robot's humanlike appearance [196]. In the context of search and rescue, people felt calmer when a robot had a non-humanlike appearance; considering that such an interaction context is highly stressful for humans, apparently a robot's machine-like-aspects are more desirable [33]. A similar preference was shown in the case of robots that are designed to interact in crowded urban environments [86]. People indicate that in the first place a robot should be functional and able to complete its tasks correctly and only in the second place does its enjoyable behaviour matter [139]. In addition, a robot's movement does not need to be natural because in some contexts people may prefer caricatured and exaggerated behaviour [214]. There are also legal questions regarding anthropomorphic technology that must be addressed. Android science has to resolve the issue of the moral status of androids—or unexpected ramifications might hamper the field in the future [46].

The risks of anthropomorphism in AI

The disadvantages of building anthropomorphic robots also include the following, in ascending order of seriousness for AI. First, it has been claimed, the phenomenon of anthropomorphic robots (at least as portrayed in fiction) encourages the general public to think that AI has advanced further than it has in reality—and to misidentify AI as concerned only with humanlike systems. Jordan Pollack remarked, for example, 'We cannot seem to convince the public that humanoids and Terminators are just Hollywood special effects, as science-fictional as the little green men from Mars!' [164, p. 50]. People imagine that service robots of the future will resemble robots from literature and movies [152].

The second problem arises specifically for those researchers in social and developmental robotics whose aim is to build anthropomorphic robots with the cognitive or affective capacities of the human infant. Several theorists claim that focusing on human-level and humanlike AI is a hindrance to progress in AI: research should focus on the 'generic' concept of intelligence, or on mindless intelligence [164], rather than on the parochial goal of human intelligence (see e.g. [85, 108]). To quote Pollack again, AI behaves 'as if human intelligence is next to godliness' [164, p. 51]. In addition, critics say, the difficult goal of humanlike AI sets an unrealistic standard for researchers (e.g. [52]). If sound, such objections would apply to the project of building a child machine.

The third difficulty arises for any attempt to build a socially intelligent robot. This is the forensic problem of anthropomorphism – the problem of how we are reliably able to detect intelligence in machines, given that the tendency to anthropomorphize leads us to find intelligence almost everywhere [166, 168]. Researchers in AI have long anthropomorphized their machines and anthropomorphic robots can prompt fantasy and make-believe in observers and researchers alike. Such anthropomorphizing is not 'innocent': instead it introduces a bias into judgements of intelligence in machines and so renders these judgements suspect.[2] Even at the beginning of the field, in 1948, Turing said that playing chess against a 'paper' machine (i.e. a simulation of machine behaviour by a human being using paper and pencil) 'gives a definite feeling that one is pitting one's wits against something alive' [204, p. 412]. His descriptions of his own machines were sometimes extravagantly anthropomorphic—he said, for example, that his child machine could not be sent to school 'without the other children making excessive fun of

---

[2] Daniel Dennett uses the notion of 'innocent' anthropomorphizing in [58].

it' [202, pp. 460-1]—but they were also plainly tongue-in-cheek. He made it clear, when talking of 'emotional' communication between human and child-machine[3] (the machine was to be organised by means of 'pain' and 'pleasure' inputs) that this did 'not presuppose any feelings on the part of the machine' [202], [204, p. 461]. In Turing's vocabulary, 'pain' is just the term for a signal that cancels an instruction in the machine's table.

Anthropomorphizing leaves AI with no trustworthy way of testing for intelligence in artificial systems. At best, the anthropomorphizing of machines obscures both AI's actual achievements and how far it has to go in order to produce genuinely intelligent machines. At worst, it leads researchers to make plainly false claims about their creations; for example, Yamamoto described his robot vacuum cleaner Sozzy as 'friendly' [222] and Hogg, Martin, and Resnick said that Frantic, their Braitenberg-like creature made of Lego bricks, 'does nothing but think' [115].

In a classic 1976 paper entitled 'Artificial Intelligence Meets Natural Stupidity', Drew McDermott advised scientists to use 'colourless' or 'sanitized' technical descriptions of their machines in place of unreflective and misleading psychological expressions [144, p. 4]. (McDermott's target was 'wishful mnemonics' [144, p. 4] but anthropomorphizing in AI goes far beyond such shorthand.) Several researchers in social robotics can be seen as in effect attempting to follow McDermott's advice with respect to their anthropomorphic robots. These researchers refrain from saying that their 'expressive' robots have emotions, and instead say that they have emotional behaviour. Kismet, Cynthia Breazeal's famous (now retired) interactive 'expressive' robot was said (without scare-quotes) to have a 'smile on [its] face', a 'sorrowful expression', a 'fearful expression', a 'happy and interested expression', a 'contented smile', a 'big grin', and a 'frown' ([39, p. 584-8], [44, 38]). However, this vocabulary is not sufficiently sanitized: for example, to say that a machine smiles is to say that the machine has an intent, namely to communicate, and an inner state, typically happiness.[4] Here the forensic

---

[3] Turing distinguished between communication by 'pain' and 'pleasure' inputs and 'unemotional' communication that by means of 'sense stimuli' [204]; for analysis see [170].

[4] This is not to suggest that what makes a 'smile' into a smile is some feeling—an inner state—in the robot. See [169, 167].

problem of anthropomorphism reemerges. We need a test for expressiveness, as much as for intelligence, that is not undermined by our tendency to anthropomorphize.

Uncanny valley

Anthropomorphism not only affects how people behave towards robots, but also whether they will accept them in natural human environments. The relation between the physical appearance of robots and their acceptance has recently received major interest in the field of HRI. Despite this, there are still many unanswered questions and most efforts are devoted to the uncanny valley hypothesis [146]. This hypothesis proposes a non-linear relationship between a robot's degree of humanlikeness and its likeability. With increased humanlikeness a robot's likeability also increases; yet when a robot closely resembles a human, but is not identical, this produces a strong negative emotional reaction. Once a robot's appearance is indistinguishable from that of a human, the robot is liked as much as a human being [146].

It has been suggested that neurological changes are responsible for the uncanny valley phenomenon [184]. Furthermore, it is the perceived higher ability of humanlike robots to experience and have emotions that makes people particularly uncomfortable with humanlike technology [97]. However, in spite of its popularity the empirical proof of the uncanny valley hypothesis is relatively sparse. Some studies did not find evidence supporting this hypothesis [21], while others suggest that the relation between likeability and appearance might have a different shape, one that resembles more a cliff than a valley [20]. We believe that future research should address three key issues: defining terminology, finding entities that lie between the deepest point of the uncanny valley and the human level, and investigating the uncanny valley in studies that involve actual HRI.

Up to now multiple terms have been used in place of the Japanese term used by Mori to describe the uncanny valley [226]. This reduces the comparability of the studies. Moreover, other researchers point out that even the humanlikeness axis of the graph is not well-defined [48]. Efforts are spent on trying to find a term that would fit the hypothesized shape of the valley rather than on creating a hypothesis that fits the data. It is also possible that the term used by Mori might not be the most appropriate one and that the problem does not lie only in the translation.

The uncanny valley hypothesis suggests that when a robot crosses the deepest point of the uncanny valley its likeability will suddenly increase. However, to date, no entity has been shown to exist that is similar enough to a human for it to fit this description. We propose that work on the opposite process to anthropomorphism could fill that niche. It has been suggested that dehumanization, which is the deprivation of human qualities in real human beings, is such a process [102]. Work on dehumanization shows which characteristics are perceived as critical for the perception of others as human. Their elimination leads to people being treated as if they were not fully human. The studies of dehumanization show that there are two distinct senses of humanness—uniquely human and human nature. Uniquely human characteristics distinguish humans from other species and reflect attributes such as intelligence, intentionality, or secondary emotions. On the other hand, people deprived of human nature are perceived as automata and lacking in primary emotions, sociability, or warmth. These two dimensions map well onto the concept proposed for the dimensionality of mind-attribution that was found to involve agency and experience [95]. Therefore, we could explore the uncanny valley, not by trying to reach the human level starting from a machine, but rather by using humans that are perceived as lacking some human qualities. There is some empirical evidence that anthropomorphism is also a multi-dimensional phenomenon [228].

In addition, the majority of previous studies of the uncanny valley hypothesis have used either static images or videos of robots. The question remains how well these findings can be generalized to actual HRI. It is possible that the uncanny valley would have no effect on HRI or that it would be limited to the very first few seconds of interaction. The studies of the uncanny valley phenomenon in computer graphics indicate that this phenomenon might be related to the exposure to a specific agent [60]. Increased familiarity with an agent could be related with decreased uncanniness felt as a result of its appearance. The physical appearance of a robot is not the most important factor in anthropomorphism [127]. Furthermore, the perception of humanlikeness changes during interaction [88]. It is possible that the uncanny valley might lead to people being less willing to engage in interaction. However, we believe that more effort should be put into interaction design rather than physical-appearance design, since the relationship between the former and the uncanny valley needs further empirical research.

1.1.5 Overcoming the problems of anthropomorphic technology

Even if we accept the uncanny valley as it was proposed by Mori, there are certain reasons why the consequences for acceptance of humanlike robots are not as profound as the theory indicated. In non-laboratory conditions people rarely reported an eerie feeling when interacting with a geminoid [27]. Furthermore, at least for computer graphics, there are guidelines regarding the creation of humanlike heads that can reduce the unnaturalness of agents' faces [142]. People find robots' performance much more important than their appearance [100], which further emphasizes that whether a robot performs its task correctly is of greater importance than how it looks.

Facilitating positive attitudes toward eerie machines

If the uncanny valley has a lasting effect on HRI, it is beneficial to consider how the acceptance of eerie machines could be facilitated. Previous work in HRI shows that people can perceive robots as either ingroup or outgroup members [76] and even apply racial prejudice towards them [79]. Therefore, the theoretical foundations for the integration of highly humanlike robots could build on the extensive research examining how to facilitate positive intergroup relations between humans belonging to differing nationalities, ethnicities, sexual, or religious groups. The topic of intergroup relations has been heavily investigated by social psychologists worldwide since World War II. While some of this early work was interested in understanding psychological factors that led to the events of the Holocaust, Gordon Allport's seminal work on the nature of prejudice [6] provided the field with a larger platform to examine the basic psychological factors underlying stereotyping, prejudice, and discrimination.

From several decades of research on the topic, the field has not only shed light on the varied ways in which intergroup bias manifests itself in everyday life [57, 98], but it also helps us better understand the economic, motivational, cognitive, evolutionary, and ideological factors that drive intergroup bias and conflict between social groups [112]. In addition, the field has also identified several social psychological approaches and strategies that can be used to reduce prejudice, stereotyping, and discrimination toward outgroups (i.e. groups to which we do not belong). These strategies range from interventions that promote positive feelings and behaviour toward outgroups through media messages [157, 221, 63], recategorization of outgroups into a common superordinate group [63, 64], valuing diversity and what each subgroup can contribute to the greater good [208, 224, 210], promoting positive contact with members of the outgroup [160, 63, 161], among others.

In the context of HRI, these social psychological strategies may be used to promote positive HRI and favorable social attitudes toward robots. For example, from over fifty years of empirical research on intergroup contact, there is strong empirical evidence that positive contact with an outgroup can reduce prejudice or negative feelings toward the outgroup [160, 161]. Such positive contact between two groups has been shown to be particularly beneficial when four conditions are met: (a) within a given situation, the perceived status of the two groups must be equal; (b) they must have common goals; (c) cooperation between the two groups must occur; and (d) positive contact between groups must be perceived as sanctioned by authority. Such intergroup contact may reduce outgroup prejudice for many different reasons [160, 161], one reason being that positive contact allows one to learn more about the outgroup. In the context of HRI, one may, therefore, expect that negative attitudes towards humanlike robots may stem from their perceived unfamiliarity and unpredictability. Although they look humanlike, people cannot be sure whether machines will behave like a human being. Increased familiarity with such technology might thereby lead to decreased uncertainty regarding its actions and in turn reduce negative feelings toward them. Empirical research is needed in order to establish whether intergroup contact can facilitate greater acceptance of anthropomorphic robots. More broadly, such social psychological research may offer insight into understanding when and why people may feel unfavorably toward robots, while offering practical strategies that can be considered in HRI as a means to promote greater social acceptance of robots in our increasingly technological world.

Limiting the risks associated with anthropomorphic technology

Of the criticism that anthropomorphic robots (in fiction at least) encourage the general public to think that AI has progressed further than in actuality, we can simply say that this may underestimate the public's good sense. The objection, against those researchers aiming to build a child-machine, that humanlike AI is a mistaken and unproductive goal can also be answered. For example, the real target of this complaint may be, not human-level or humanlike AI as such, but rather symbolic AI as a means of attaining human-level AI (see [166]). Behaviour-based approaches may escape this complaint. Moreover, the assumption that there is such a thing as generic intelligence, and this is the proper subject of study for researchers in computational intelligence, begs an important question. Perhaps our concept of intelligence just is drawn from the paradigm examples of thinking things—human beings.

This leaves the forensic problem of anthropomorphism. In general, AI requires a distinction between a mere machine and a thinking machine, and this distinction must be proof against the human tendency to anthropomorphize. This is exactly what Turing's imitation game provides (see [166, 168]). The game disincentivizes anthropomorphism: an observer (i.e. interrogator) who anthropomorphizes a contestant increases the chances of making the embarrassing mistake of misidentifying a computer as a human being. The behaviour of interrogators (in early Loebner Prize Contests where a series of machine and human contestants were interviewed individually) shows that observers go out of their way to avoid this awkward error, to the extent that they misidentify human beings as computers. In addition, the imitation game controls for the effect of the tendency to anthropomorphize; in simultaneous interviews with a machine and a human contestant, an observer's propensity to anthropomorphize (which we can assume to be present equally in both interviews) cannot advantage one contestant over the other. Turing's test is proofed against the human tendency to anthropomorphize machines.

But how is anthropomorphism-proofing to be applied to judgements of intelligence or affect in anthropomorphic robots? Much of the engineering of these robots is to make them visibly indistinguishable from a human being. An open imitation game where both contestants—a hyper-realistic humanlike robot and an actual human being—are seen by the interrogator would provide a disincentive to anthropomorphizing and a control on the tendency to anthropomorphize. However, interrogators in this game might well focus on characteristics of the contestants that Turing labeled 'irrelevant' disabilities—qualities immaterial to the question whether a machine can think, such as a failure 'to shine in beauty competitions' [202, p. 442]. An interrogator might, for example, concentrate on the functioning of a contestant's facial muscles or the appearance of the skin. This open game, although anthropomorphism-proofed, would fail as a test of intelligence or affect in machines. On the other hand, in the standard game where both the robot and the human contestants are hidden, much of the robot's engineering would now be irrelevant to its success or failure in the game—for example, David Hanson's robot Einstein's 'eyebrows' [154] surely do not contribute to a capacity for cognition or affect. This is why Turing said that there is 'little point in trying to make a "thinking machine" more human by dressing it up in ... artificial flesh' [202, p. 442] and hoped that 'no great efforts will be put into making machines with the most distinctively human, but non-intellectual characteristics such as the shape of the human body' [206, p. 486].

In sum, AI requires some means of anthropomorphism-proofing judgements of intelligence or affect in anthropomorphic robots—otherwise it lacks a distinction between justified and unjustified anthropomorphizing. An open Turing test will test for the wrong things. A standard Turing test will suffice, but seems to be inconsistent with the growing trend for hyper-realistic and eerie robots.

1.1.6 Conclusion

In this section we have discussed the widespread tendency for people to anthropomorphize their surroundings and in particular how this affects HRI. Our understanding of its impact on HRI is still in its infancy. However, there is no doubt that it creates new opportunities and poses problems that can have profound consequences for the field of HRI and acceptance of the robotic technology.

Anthropomorphism is not only limited to the appearance of a robot, but the design of a robotic platform must also consider a robot's interaction with humans as an important factor. Accordance between these factors is necessary for a robot to maintain its humanlike impression. A well designed system can facilitate interaction, but it must match the specific task given to a robot. For people it is more important that a robot can do its job accurately rather than how it looks. However, we have presented multiple examples where humanlike form in appearance and behavior can help a robot to perform its tasks successfully by eliciting desired behaviours from human interaction partners.

On the other hand, development of anthropomorphic robots comes at certain costs. People expect them to adhere to human norms and have much higher expectations regarding their capabilities compared to robots with machine-like appearance. The uncanny valley hypothesis suggests that there is repulsion toward highly humanlike machines that are still distinguishable from humans. However, in this section we have shown the main shortcoming of the previous work that might hamper the suitability of this hypothesis in HRI. Future research should investigate this phenomenon in real HRI rather than by using images or videos. Moreover, work on the opposite process, dehumanization, can help us to understand the relationship between acceptance and anthropomorphism better. In addition, in order to facilitate the integration of humanlike robots we propose to employ strategies from the area of intergroup relations that are being used to facilitate positive relations between human subgroups.

We have also shown that the phenomenon of anthropomorphic robots generates challenging philosophical and psychological questions. In order for the field of AI to progress further it is necessary to acknowledge them. These challenges can not only affect how the general public perceives current and future directions of research and anthropomorphic and intelligent systems, but also might determine how far the field can go. It remains to be seen whether the field can successfully address these problems.

1.2 Aims

The literature review indicates that there are numerous ways in which anthropomorphism can be manipulated. However, there is a gap in our knowledge in understanding what attributes are key in that process. There is a need for a model of anthropomorphism that could put together the up to date findings, as well as indicate what other unexplored characteristics should be investigated in future research. Such a model would enable interaction designers of robotic systems to change a robot's attributes, mainly behaviour, on the fly so that they fit current task. There is no specific level of anthropomorphism that is the most suitable for all tasks, but rather it should be adjusted in order to optimize a robot's performance in a specific context.

The currently available models of anthropomorphism (presented in detail in Chapter 2) provide an indication that the context of interaction as well as the perceiver can play an important role in that process. The research shows that embodiment plays an important role in anthropomorphism of robots. However, currently it is very difficult for a robot to anatomically change its physical appearance during the course of an interaction with a human. While in the future the development of technology might enable it, currently the robots' physical appearance is static and interaction designers should focus on other characteristics affecting the interaction. Since their embodiment could be a moderating factor affecting how other attributes are perceived, providing a model of anthropomorphism for a specific embodiment would have a very limited benefit for the HRI community. Therefore, the goal of this PhD dissertation is to investigate the factors affecting anthropomorphism and the ways of measuring them. In order to achieve that goal the following aims should be met:

1. Explore what is anthropomorphism.
2. Propose or develop a suitable tool for measuring anthropomorphism in HRI.
3. Investigate the process involved in anthropomorphism.
4. Investigate the factors that affect anthropomorphism of a robot.

5. Investigate the applicability of the model of dehumanization as a new model of anthropomorphism for HRI.

Chapter II

Models of anthropomorphism

2.1 Existing models

In the recent years several models of anthropomorphism and humanlikeness were proposed: three-factor theory of anthropomorphism [69], appearance and behaviour parameter based model of humanlikeness [209] and dynamic-cognitive model [136]. In this subsection I will discuss these models in more detail.

A general model of anthropomorphism as a common psychological phenomenon was proposed by Epley and colleagues [69]. Their three-factor theory of anthropomorphism proposed psychological factors that affect how people anthropomorphize non-human agents:

1. Elicited agent knowledge—due to people's much richer knowledge regarding humans compared with non-human agents, people are more likely to use anthropomorphic explanations of non-human agents' actions until they create an adequate mental model of non-humans.

2. Effectance motivation—when people are motivated to explain or understand an agent's behaviour the tendency to anthropomorphism increases.

3. Sociality motivation—people who lack social connection with other humans often compensate for this by treating non-human agents as if they were humanlike [67].

This general theory has been also successfully applied in the context of HRI [74, 131, 182]. However, only the elicited agent knowledge is a factor that is directly applicable to robots design.

Elicited agent knowledge can be induced as a result of a robot's humanlikeness. The second model, proposed by von Zitzewitz et al. [209], tries to break down a robot's appearance and behaviour into smaller and measurable factors. The authors propose that there are five static aspects of appearance: visual appearance, sound, smell, haptic appearance and taste. Moreover, there are five dynamic aspects of behaviour: movement, nonverbal communication, interactive behaviour, verbal communication and social behaviour. Each of these aspects is further parametrized to measurable units. In this model the authors also propose that initially appearance is a more important factor of humanlikeness, but with increased social interaction behavioral factors take the crucial role. The major strength of this model is its quantification of factors that are relevant for a robot's perceived humanlikeness. However, it is focused on investigating the impact of these factors separately from each other while in real-life situations the behaviour of a robot is perceived holistically without analysing different aspects independently. Moreover, the model assumes that there is a fixed level of parameters that is the most humanlike in all situations. However, human beings adjust their behaviour depending on the situation, e.g. people move slower when they are around fragile items compared to being in open space. Therefore, in order to apply this model in practice would require prior measurement of the factors and parameters for every human-robot interaction.

The third model discusses the cognitive correlates of anthropomorphism [136]. This newly proposed model emphasizes that anthropomorphism of a robot is not constant, but can change during interaction. Furthermore, it indicates that anthropomorphism can have human, robot and context origins. However, as the authors noted themselves, at the current time the link with the cognitive processes is rather weak. Moreover, according to this model, initial anthropomorphism of a robot always raises sharply in the initial stages of interaction and then slowly drops, a claim that is not supported by the empirical data. In addition, similarly to the three-factor theory of anthropomorphism, this model does not provide any information on what factors make robots more anthropomorphic and therefore its applicability for the robotic system designers is limited. Nevertheless, this is a promising theory that requires empirical work in order to validate this model.

2.2 Dehumanization

Compared with the above discussed models, the model proposed in this thesis took humanness as a core component of anthropomorphism and instead of looking at what makes objects more anthropomorphic, it is based on work that investigates what makes humans less human. The theoretical relationship between dehumanization and anthropomorphism was suggested by Waytz et al. [217]. The process of dehumanization - “a failure to attribute basic human qualities to others” [216] - only recently became a focus of interest in the field of social psychology. A key finding that sparred interest in this topic is a publication of [138] in which they showed that people attribute differently certain emotions to ingroup and outgroup members. They have proposed that people attribute essences differently to the ingroup and outgroup members. The underlying essence of the ingroup is the human nature. Therefore, they argued that by attributing different essences to ingroups and outgroups people perceive the former to possess some human characteristics that the latter lacks. They found that across different languages people distinguish between sentiments and emotions. The sentiments were uniquely human, while emotions were common to humans and animals. They mapped these terms on primary and secondary emotions, where secondary emotions constitute what is uniquely human.

Ekman [66] suggested that one of the characteristics that distinguishes these two categories of emotions are that primary emotions are present in other species. Furthermore, primary emotions appear early in life [119], while secondary emotions appear later and are a result of socialization [126]. According to their hypothesis [138] found that participants attributed less secondary emotions to outgroup members than ingroup members. A series of following studies shows that people implicitly stronger attribute secondary emotions to ingroup and primary to outgroup members, whether the emotions are positive or negative [156]. They also implied that the restriction of attribution of human nature characteristics to ingroups leads to infrahumanization of outgroups. Moreover, in the memory-recall task subjects remembered better association between the outgroup and secondary emotions than the ingroup and secondary emotions, while no differences were found for primary emotions [89]. Furthermore, people who are ostracized judge themselves and those who ostracized them as less human, and believe that they are perceived as less human by them on the basis of attribution of secondary emotions [26].

Haslam [102] proposed a model of dehumanization that involves two distinct senses of humanness: characteristics that are uniquely human and those which form human nature. Since these characteristics are perceived by people as distinguishing them from animals and automata, denying the former attributes leads to perception of humans as animal-like, while denying the latter makes them object or automata like. Uniquely human (HU) characteristics, such as intelligence or intentionality, are what separates humans from animals. On the other hand, features that are typical of or central to humans are referred to as human nature (HN) characteristics, such as primary emotions or warmth. Therefore, characteristics that form the core of humanness may not be the same as those which distinguish us from other species.

There are several aspects that differentiate these two senses of humanness:

1. HU characteristics reflect socialization and culture, while HN characteristics link humans to their inborn biological dispositions.
2. HU characteristics reflect social learning and can vary across cultures and populations. HN is prevalent within populations and universal across different cultures.
3. HN is essential, inherent and natural, while HU may not be perceived as essential.
4. HU involves refinement, civility, morality, and higher cognition. HN involves cognitive flexibility, emotionality, vital agency, and warmth.

Moreover, different personality traits are attributed to different forms of humanness. High HU traits were related to low Neuroticism and Extraversion, and positive and negative poles of Agreeableness, Conscientiousness and Openess to Experience [102]. On the other hand, HN was related to high Agreeableness, Conscientiousness, Extraversion and Openess to Experience, and low Neuroticism. Therefore, affective traits are central to HN, while cognitive sophistication to HU.

Further, Haslam [102] stated that those who are deprived HU characteristics will be perceived as coarse, uncultured, unintelligent and driven by motives and instincts. However, those deprived of HN will be perceived as lacking in emotionality, warmth, cognitive openness and individual agency.

There are also different consequences of depriving humans HU and HN characteristics [102]:

1. A person deprived of HU traits is perceived as disgusting, while lack of HN traits implies indifference in perceiver and lack of empathy.

2. Lack of HU characteristics leads to perception of a person as subhuman and lack of HN as nonhuman.

3. Deprivation of HN affects intergroup and HU applies to intergroup and interpersonal contexts [104].

An important notion is that dehumanization does not only occur in extreme situations, but is rather common in its milder forms in our everyday social life [102]. Some social groups are implicitly and explicitly attributed less HU characteristics (e.g artists) and therefore likened to animals, while other (e.g. businesspeople) are attributed less HN characteristics and likened to automata [140]. Furthermore, HU traits are associated stronger with automata than with animals, but HN has stronger association with animals than with automata. In another study Haslam et al. [106] found that animals compared to humans were perceived as lacking higher cognitive powers and secondary emotions, but they were also seen as having better perceptual capacities. Robots lacked emotions and desire related capacities, and supernatural beings had superior cognitive and perceptual capacities than humans. The results are consistent with previous findings suggesting that there are 2 dimensions of humanness. Furthermore, they imply that superhumans differ from humans on the same dimension as robots and being perceived more positively than other humans is also a form of dehumanization.

Another approach was taken by Gray et al. [95]. For them having mind is what distinguishes humans. They measured eighteen mental capacities and six personal judgments regarding various human and nonhuman agents and found that the mind perception was explained by two factors: Agency and Experience. Characteristics that formed Agency dimension are: abilities to have self-control, morality, memory, emotion recognition, planning, communication, and thought. Those which form Experience are: the abilities to feel hunger, fear, pain, pleasure, rage, and desire; to have personality and consciousness; and to feel pride, embarrassment, and joy. Furthermore, agency is linked to moral agency, which leads to perception of agents responsibility. On the other hand, experience is linked with moral patiency and that gives an agent rights and privileges. Indeed, as proposed in the moral typecasting model people are either perceived as moral agents (those who can do good or bad actions) or moral patients (those who can be a target of good and bad actions) [96]. Moral agents are perceived as being less vulnerable to having good and bad done to them, while moral patients are perceived as less capable of performing good and bad actions.

Interestingly, the dimensions of mind perception [95] map well on those proposed by Haslam [102] for humanness [105]. Agency dimension corresponds to Human Uniqueness and Experience to Human Nature. The objectification of women and men was found to lead to their depersonalization due to the deprivation of mind [141]. Moreover, [105] proposed that attribution of HU traits leads to greater inhibitive agency (responsibility for doing wrong). On the other hand, higher Human Nature lead to greater moral patiency (having rights for being protected from harm) and proactive agency (being praised for doing right). The important conclusion from this finding is that depending on a way in which a robot is anthropomorphized it can be attributed differently moral responsibility and rights for protection. On a legal level understanding the analogy between different groups of humans (e.g. slaves or children) or animals and robots can have implications on what rights will be given to robots in the future [46].

Furthermore, the speed of movement is used by humans to infer the presence of mind and attribute mental attributes, such as intention or intelligence, to objects, animals and people [145]. Targets that move slower or faster from humans are more likely to be perceived as lacking mental capabilities and the speed used as a reference can be dependent on a group to which perceiver belongs (e.g. elderly or child).

Understanding how dehumanization affects perception of humanness can give a new perspective on the process of anthropomorphization in HRI. It provides an indication on what characteristics may affect a user's perception of the degree to which a robot is anthropomorphized, and tools and methods used to measure dehumanization can be used in HRI for measuring anthropomorphism. The work on dehumanization recently started receiving attention in the context of HRI. However, these studies used dehumanization mainly as a measurement tool of anthropomorphism of robots e.g. [74, 131, 182]. The previous work on anthropomorphism in HRI considered this phenomenon as a uni-dimensional space from a machine to a human. However, since humanness is the essence of anthropomorphism, the studies of dehumanization can indicate that anthropomorphism can be at least two-dimensional (an exception is [75] where anthropomorphism of a robot was measured on two scales of humanness, but they were not considered as different dimensions of that construct). Therefore, in order to accurately represent the extent to which a robot is anthropomorphized it would be necessary to measure it on these two dimensions.

2.3 Dual-process

Anthropomorphism has been measured using various measurement tools, such as questionnaires [22], physiological measures [132] or social behaviour [25]. In addition, previous studies indicate that people's anthropomorphic conceptualizations of robots can differ on different levels of abstraction. People are more willing to exhibit anthropomorphic responses when freely describing a robot's behaviour in a specific context or attributing properties to the robot performing these actions, but not properties that characterize robots in general [88]. However, the difference in anthropomorphizing robots could be expressed not only on different levels of abstraction, but also on different levels of measurement, i.e. explicit vs implicit. Not only the results could be inconsistent between these different measurement levels, but also there could be different processes or systems responsible for that difference.

Many cognitive and social psychological judgements and attributions are a result of two processes broadly described as intuitive and deliberative/reflective [55]. The key distinction between them is their use of working memory [72]. Type I process (intuitive) is autonomous and does not require working memory. Its typical correlates are being fast, with high capacity, processing occurring in parallel, unconsciously, and leading to biased and automatic responses. On the other hand, Type II process (reflective) involves cognitive decoupling and requires working memory. Its typical correlates are being slow, with limited capacity, processing occurring in serial, consciously, and leading to controlled and normative responses.

There is no agreement whether Type I and II processing proceed in parallel-competitive form, where each of them can affect the judgement and conflict will be resolved if needed [193], or in default interventionist form, where Type I processing generates initial response and Type II processing may or may not modify it [72]. In addition, there are individual differences in rational thinking dispositions that affect Type II processing, but few individual differences in Type I processing [72]. It has been also proposed that there could be multiple different implicit cognitive Type I processes [71].

There are three major sources of evidence supporting the existence of Type I and Type II processing: experimental manipulations (providing instructions that are supposed to increase motivation for Type II processing effort or suppress it by using a concurrent task that limits capacity of working memory), neural imaging and psychometric approach showing that Type II, but not Type I processing has relation with cognitive ability [72].

Some researchers propose that anthropomorphism is a conscious (Type II) process [218]. However, currently there is no empirical evidence supporting this claim. Since anthropomorphism has cognitive and social correlates it is possible that a dual process is involved in formulation of a judgement of a robot. The initial interpretation of a vague object as a human that has evolutionary origins [13], could be a result of Type I processing. However, when people are motivated to provide accurate judgement and Type II processing will be activated, it is possible that they will not necessarily attribute human characteristics to these objects.

Table 2.1: Preliminary model of anthropomorphism for HRI. A list of factors grouped by dimension of humanness based on work on dehumanization.

Human Uniqueness	**Human Nature**
Intelligence	Primary emotions
Intentions	Warmth
Secondary emotions	Sociability
Self-control	
Memory	
Communication capabilities	

2.4 Integrative model

Based on the literature review presented in the previous two sections a new model of anthropomorphism is proposed. The novel contribution of this new model is in reversing the model of dehumanization and proposing that the same characteristics which dehumanized people are deprived of could be used to increase anthropomorphism of a robot. Since work on dehumanization presents two dimensions of humanness, the new model of anthropomorphism for HRI also proposes that there could be multiple dimensions of anthropomorphism. Table 2.1 presents the list of factors proposed to affect anthropomorphism of a robot grouped by dimensions of humanness (based on [102]).

In order to understand better anthropomorphism and how it affects HRI, it is also important to understand the process itself. Therefore, the proposed model of anthropomorphism suggests that anthropomorphism, similarly to many other social processes, could be a result of both Type I and Type II processing.

2.5 Research questions

Based on the above discussed literature and the proposed model of anthropomorphism for HRI, the following general research questions are addressed in this dissertation:

1. How to measure anthropomorphism?

2. Can we measure anthropomorphism explicitly and implicitly?
3. Is anthropomorphism a result of Type I or Type II processing?
4. Is anthropomorphism an opposite process to dehumanization?
5. Can factors which people deprive dehumanized others be used to affect anthropomorphism of a robot?
6. Are there multiple dimensions of anthropomorphism?
7. Is psychological anthropomorphism and physical humanlikeness of a robot the same?

In the following five chapters, experimental studies are presented that addressed one or more of these questions. Finally, in the last chapter, conclusions based on these five studies are discussed and directions for future work suggested.

Chapter III

Development of a cognitive measure of anthropomorphism

In human-human interaction people form an impression of others within the first seconds to minutes time frame [28]. Moreover, the first impression can have lasting consequences in HRI as well [22, 163]. However, there are also studies that indicate early conceptions can change during the course of an interaction. After interacting with a robot people tended to anthropomorphize it more [88]. This change in user's perception of a robot has been shown in infant-robot interaction [162]. Furthermore, people tend to judge a robot's physical appearance and its capabilities in relation to its role [195].

Nevertheless, embodiment plays an important role in anthropomorphization, especially in the phase before the actual interaction is initiated. Hence, it is not surprising that in the recent years extensive research has been conducted with an aim to better understanding the impact that embodiment has on people's behavior whilst interacting with robots.

Fischer et al. [83] showed that physical embodiment and degrees of freedom affect HRI. The former has influence on how deeply a robot is perceived as an interaction partner, whilst the latter has an impact on how users project suitability of a robot for its current task. Moreover, a user's personality can impact on preferences regarding a robot's physical appearance [197]. Other factors, such as crowdedness can have influence on what type of a robot's physical appearance will lead to a longer interaction [92]. Furthermore, Hegel and colleagues [110] found that the increase of humanlikeness of a robot's embodiment leads to attribution of higher intelligence and different cortical activity.

Since the role of embodiment is rather complex and still not well understood, there is no doubt that it will remain one of the research focus areas in forthcoming years. Embodiment also emphasizes the importance of choosing an appropriate physical appearance for social robots that will enter human spaces. Therefore, it is necessary that robotic system designers are able to assess the level of anthropomorphism of various embodiments. Questionnaires are among the most popular measurement tools used in HRI research, e.g. [22]. However, there are some attempts to develop alternative methods. Kriz et al. [130] used robot-directed language as a tool for exploring people's implicit beliefs toward robots. Moreover, Admoni and Scassellati [3] proposed that an understanding of mental models of robots' intentionality can guide the design process of robots. In addition, Ethnomethodology and Conversation Analysis shed more light on our understanding of HRI [162].

Bae and Kim [11] took a different approach. They were interested to see whether visual cognition allots more attention to robots with animate or inanimate forms. In order to answer their question, they conducted a change detection experiment in which participants briefly saw two images involving a robot and were asked to decide whether the images were identical or not. The allocation of visual attention to the changes perceived is responsible for the change detection in the paradigm [116]. They found that the participants detected changes swifter in animate rather than inanimate robots.

In this chapter we will present an attempt for developing a new measurement tool with similar approach to [11]. We used images of robots and explored their relation with visual cognition in an attempt to validate a new method for measuring robots humanlikeness based on their physical appearance. However, compared with [11], we have utilized a different paradigm, known as the inversion effect.

3.1 Inversion effect

The inversion effect is a phenomenon when upside down stimuli are significantly more difficult to recognize than upright stimuli [223, 143]. It has been originally reported for the recognition of human faces [223] and explained as a result of different processing used for different types of stimuli [143]. Configural processing, used for identification of human faces, involves the perception of spatial relations among the features of a face (e.g. eyes are always in certain configuration with a nose). On the other hand, during the recognition of objects, the spatial relations are not taken into account and this type of processing is called analytic [174].

While most of the objects are recognized by the presence or absence of individual parts, the recognition of human faces involves configural processing, which requires specific spatial relations between face parts [47, 53, 134]. Furthermore, human body postures produce similar inversion effects as what human faces do [173] and therefore, a different processing mechanism is involved in the recognition of human body postures compared to inanimate objects [174]. Moreover, configural human body posture recognition requires whole body rather than merely body parts with the posture physically possible [174].

Although, in general the inversion effect occurs only for human body and faces, it has been shown that people who become experts in certain categorization of objects (e.g. a specific breed of dogs) can recognize them using configural processing and exhibit a similar handicap of performance due to the inversion effect as shown for human faces [198]. Diamond and Carey [59] suggested that there are three conditions necessary for the inversion effect to occur:

- The members of the class must share configuration
- Distinctive configural relations among the elements must enable individuating members of the class
- Subjects must have expertise in order to exploit such features

In relation to robots, Hirai and Hiraki [113] conducted a study that investigated whether the appearance information of walking actions affects the inversion effect. The event-related potential (ERP) indicated that the inversion effect occurred only for animations of humans. Thus, a robotic walking animation was processed differently than a human body.

Recently studies suggested that under certain conditions the human body can be perceived analytically rather than configurally. In other words, it is possible for a human body not to induce the inversion effect and therefore be processed like an object. Bernard et al. [32] showed that due to the objectification of sexualised women's bodies, their recognition is not handicapped by the inversion effect. However, sexualised men's bodies exhibit the inversion effect like non-sexualised body postures. The gender of participants did not play a role in this phenomenon. Gervais [90] found that women's bodies were reduced to their sexual body parts and that led to the perception of them on the cognitive level as objects.

These findings show that non-sexualized human body postures and faces are perceived differently than other objects. Furthermore, humans are able to objectify other human bodies. In this chapter we explore whether they are also able to perceive robots' appearance like human body posture since robots can share certain physical characteristics of human bodies. Due to practical concerns, we have used images of robots rather than real robots on which we elaborate more in the Limitations and future work section. First, we tested whether robots elicit the inversion effect. If their recognition is handicapped in the upside down position compared to the upright, it would suggest that they are processed configurally and therefore are viewed more like humans. Alternatively, if there be no significant decrease of recognition, this would mean that they are seen as objects. Second, we attempted to validate the magnitude of the inversion effect as a method for measuring humanlikeness. If there is a relationship between the handicap due to the inversion effect and a robot's perceived humanlikeness, it will be possible to estimate the humanlikeness of a robot's embodiment by measuring the magnitude of the inversion effect. Such an assessment tool could support robotic system designers in their choice of different embodiments for a platform which provides additional information about the level of humanlikeness.

3.2 Method

This experiment was conducted as a within-subjects design with two factors: image type (object vs robot vs human) and orientation (upright vs upside down). The recognition accuracy (whether an image is recognized correctly as same or different) and reaction time were measured for each pair of images as dependent variables. Furthermore, the Godspeed Anthropomorphism Scale [22] was used for each image in order to measure its perceived level of humanlikeness.

3.2.1 Measurements

The whole experiment was programmed and conducted using E-Prime 2 Professional Edition [117]. We have measured the correctness of participants' responses and their reaction times to the millisecond precision level. To measure the perceived humanlikeness of stimuli we have used a slightly modified version of the questionnaire based Godspeed Anthropomorphism Scale [22] (Appendix A). Since the original questionnaire included a subscale that cannot be measured with static stimuli (Moving rigidly - Moving elegantly), we removed it. The remaining 4 semantic differential scales were used in their original English version.

3.2.2 Materials and apparatus

The robot stimuli were created in the following way. We selected pictures of real robots and on each picture there is only one robot. We opted for images depicting full bodies of robots rather than their faces as we wanted to include the full range of robots embodiments and merely due to having a face, a robot is humanlike to some degree. Although the context provided by the background can play a role in a robot's anthropomorphizability, it would be a confounding variable in the presented study to indicate which images were rotated. Therefore, following the procedure of previous experiments on the inversion effect, we coloured the background and shadows in the images completely white. Furthermore, if there were some letters or numbers present on a robot, they were removed as well as they could have provided additional hints for recognizing images. We have used a wide spectrum of existing, non-fictional and non-industrial robots, ranging from Roomba and Roboscooper to more humanlike ASIMO and Geminoids. There were in total 33 different robots that varied in shape and structure of their bodies (see Figure 3.1). To be able to measure the correctness of the response, we had to create distractors, so that the participants were able to choose between a correct and

Figure 3.1: Sample of 6 images of robots with different embodiments used in the study.

an incorrect response. Distractors were created by mirroring each robot, as proposed by [32]. This method was chosen as it ensured that the modification is comparable between different robots as well as different types of stimuli. The robots were centered on the image and they were in poses that are horizontally and vertically asymmetric. In other words, the right half of an image differed from the left, and the top half differed from the bottom. Therefore, they were distinguishable from their distractors. We have created pairs of images by putting together the original image with its exact copy and another pair that included the original image and its distractor (mirrored image). Therefore, there were 33 same-robot and 33 different-robot pairs. Finally, all robot pairs were rotated 180° to create the upside down robot stimuli. All possible combinations of image pairs (trials) can be seen in Figure 3.2.

Exactly the same procedure was used in order to create human body and object stimuli. Thirty-three pictures of real people were included so that the number of pairs were the same as for the robots. Since these were real pictures, all body postures were natural and physically possible. We used images of full men and women bodies and they were presented in a non-sexual way (their bodies were covered by clothes) to ensure that sexual objectifica-

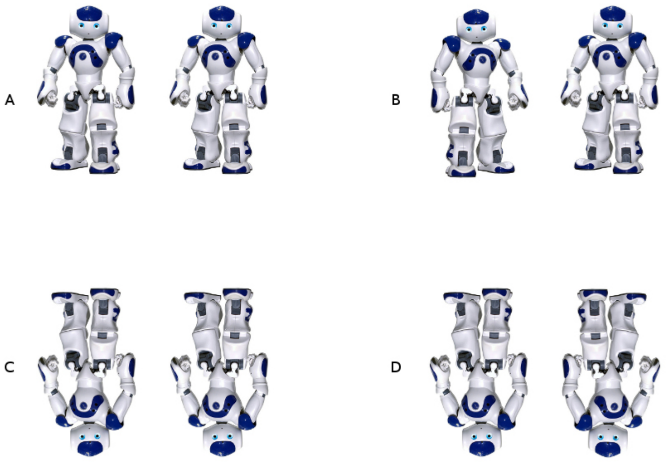

Figure 3.2: Image pairs used for a single image of a robot. A: Example of the *same* trial in the upright condition - original image paired with its copy. B: Example of the *different* trial in the upright condition - original image paired with its distractor. C: Example of a *same* trial in the upside down condition - original image rotated 180° paired with its copy. D: Example of a *different* trial in the upside down condition - original image rotated 180° paired with its distractor.

tion will not affect the results [32]. The objects category included pictures of various types of home appliance, such as dishwasher, TV or telephone. The number of object pairs were exactly the same as for the two other categories of stimuli.

3.2.3 Procedure

Each participant was seated approximately 0.5 m from a 21.5” Macintosh computer monitor with Windows XP operating system. The resolution was set to 1920x1080 pixels. Each participant was allowed to adjust the height of the chair so that his eyes were at the same level as the center of the screen. Participants were informed that the experiment consists of 2 parts. In the first part, their task is to decide whether a pair of images was exactly identical. In the second, they have to fill in the humanlikeness questionnaire.

Before the actual experiment began, participants had a practice round to familiarise with the procedure. It included in total 11 stimuli pairs of different types and orientation. The procedure to evaluate each stimulus in the practise round was identical to the procedure in the actual experiment for all other stimuli. Participants were shown the plus sign that indicated the fixation point in the center of the screen for 1000 ms, followed by the original stimulus for 250 ms and then a blank screen for 1000 ms. It was followed by a second stimulus that was either a copy or distractor of the first stimulus and remained on the screen until a participant responded. All images had 1024x768 pixels resolution. There was no mental rotation required as they were both displayed either in upright or upside down orientation. Participants pressed either the S key to indicate that the stimuli were absolutely identical or the L key if they were different. Participants were asked to respond as fast and as accurately as possible.

Upon completion of the practice round an actual experiment began, following exactly the same procedure as described above for the practice round. Stimuli were presented in 3 different blocks according to their type (object, robot, human), the order of blocks counterbalanced, and ordering of stimuli pairs randomized within each block across participants. Each of the blocks including 132 trials, which gave a total of 396 trials.

Upon completion of the first part, participants were asked to rate each stimulus used in the experiment on 4 subscales of the Godspeed Anthropomorphism Questionnaire. All types of stimuli were included and their order randomized. The entire experiment took approximately 40 minutes.

3.2.4 Participants

Fifty-one subjects were recruited at the University of Canterbury. They were offered a $10 voucher for their participation. Due to software failure, the data of 4 participants was lost. Out of the remaining 47 participants, 15 were female. There were 23 postgraduate students, 16 undergraduate students, 4 university staff members and 4 participants whom did not qualify under any of these 3 categories. Their age ranged from 18 to 58 years with a mean age of 26.26 years. They were from 24 different countries, with New Zealand (15) and China (4) being the most represented. Forty participants never interacted with a robot or did it less than 10 times. Therefore, we regard them as non-experts in robotics. Only one participant indicated that he had over 100 interactions.

3.3 Results

3.3.1 Perceived humanlikenes

Since we slightly modified the Godspeed Anthropomorphism scale it was necessary to ensure that it is still reliable. The internal consistency of the humanlikeness scale was very good, with a Cronbach's alpha coefficient of 0.96. The reliability of the Godspeed questionnaire with the included 4 subscales is well above the acceptable 0.7 level [150] and therefore the removal of one subscale should not affect the humanlikeness score.

As the scale was regarded reliable, we have obtained the score of humanlikeness for each image by calculating the mean of 4 subscales. Then, using these scores, we have calculated the score of humanlikeness for each stimuli type (object, robot, person) by taking the mean of scores for all stimuli that belonged to that type. In order to establish whether a type of image affects its perceived humanlikeness, we have conducted a one way repeated measures analysis of variance (ANOVA). We have applied the Huynh-Feldt correction, as Mauchly's test indicated that the assumption of sphericity was violated (W=0.81, p=0.01). The analysis showed that there was significant effect for image type [F(1.76,80.74)=623.3, $p<0.001$, $\eta^2_G=0.91$]. We report here the generalized eta squared to indicate the effect size. It is superior to eta squared and partial eta squared in repeated measure designs, because of its comparability across studies with different designs [12]. Post-hoc comparisons using Bonferroni correction for the family wise error rate indicated perceived humanlikeness was significantly different between groups at the level of $p<0.001$ (object, M=1.32, $\sigma=0.51$; robot, M=1.83, $\sigma=0.51$; person, M=4.67, $\sigma=0.42$) (see Figure 3.3).

3.3.2 Inversion effect

To test the hypothesis that robots, similarly to human body postures, produce an inversion effect, we have conducted a 2x3 repeated measures ANOVA with factors: orientation (upright vs upside down) and image type (object vs robot vs person). We have used the accuracy score (whether two images were correctly recognized as same or different) as the dependent vari-

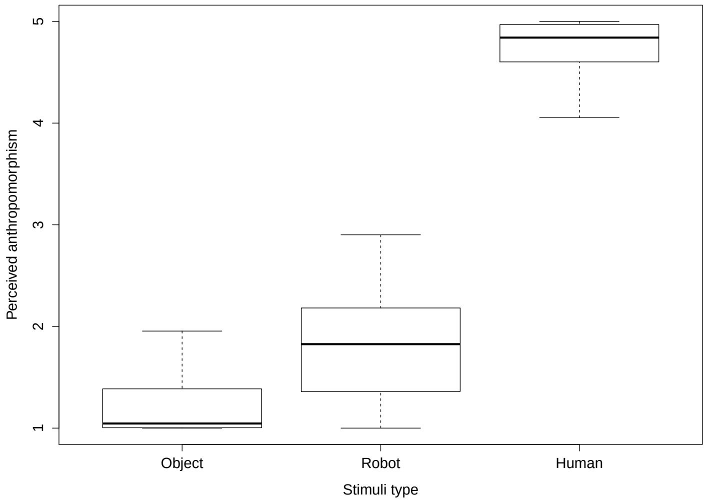

Figure 3.3: Perceived humanlikeness. The level of humanlikeness based on the Godspeed questionnaire presented for each type of stimuli.

able. We have calculated a mean score of accuracy for each image type in both orientations to obtain six scores (theoretical range 0-33). Analysis showed that the main effect for orientation was statistically significant [F(1,46)=14.36, $p<0.001$, $\eta^2_G=0.03$]. More images were recognized correctly in the upright (M=31.45, $\sigma=1.22$) than upside down (M=30.98, $\sigma=1.71$) position. However, this main effect can be explained as a result of statistically significant interaction effect between orientation and image type [F(2,92)=4.97, p=0.01, $\eta^2_G=0.02$]. If robots elicit configural processing, then interaction effects should be significant for images of robots, but not for objects. Confirming this assumption, the interaction effect was found for people [F(1,46)=13.77, $p<0.001$, $\eta^2_G=0.08$]. Recognition accuracy decreased for upside down (M=30.72, $\sigma=1.89$) compared to upright (M=31.65, $\sigma=1.19$) images of people. Similar results were found for images of robots, where interaction effect was significant [F(1,46)=7.25, p=0.01, $\eta^2_G=0.04$]. Upright images of robots were recognized more accurately (M=31.39, $\sigma=0.95$) than upside down (M=30.88, $\sigma=1.57$) (see Figure 3.4). The interaction effect was not statistically significant for objects, neither any other interaction nor main effects were found.

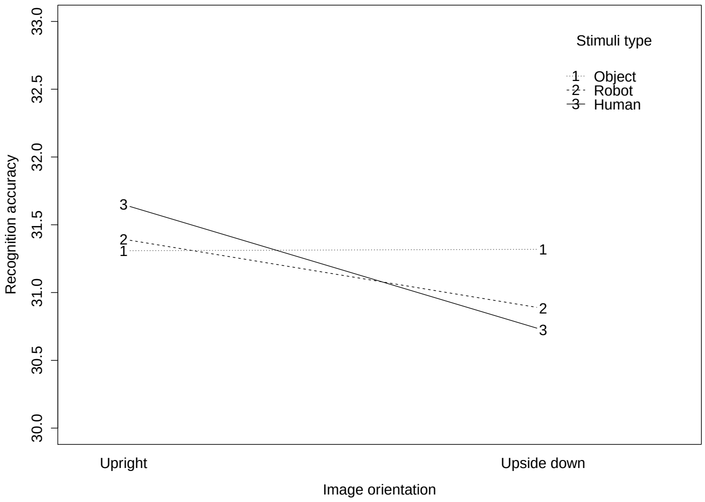

Figure 3.4: The interaction effects between types of stimuli and orientation. The drop of the recognition accuracy is visible when comparing upside down with upright orientation for human and robot stimuli. The inversion does not affect recognition of objects.

The same statistical analysis as for the recognition accuracy was applied for reaction times. A 2x3 repeated measures ANOVA was used to analyze data. All main effects and interactions were significant. The main effect of image type was statistically significant [F(2,92)=14.56, p<0.001, η^2_G=0.05]. Post-hoc comparisons using Bonferroni correction for family-wise error rate indicated mean reaction times between groups were significantly different from each other p<0.001 (object, M=879.88 ms, σ=62.97; robot, M=975.05 ms, σ=131.96; human, M=1072.23 ms, σ=146.04). Furthermore, reaction time was significantly longer [F(1,46)=12.33, p<0.001, η^2_G=0.01] for upside down (M=1004.41 ms, σ=158.1) compared to upright images (M=947.03 ms, σ=118.84).

3.3.3 Establishing validity of the inversion effect as a method for estimating humanlikeness

The above analysis indicated 3 image types differed in their perceived level of humanlikeness. The robot and human body images elicited the inversion effect. In the following step we tested the relationship between the change

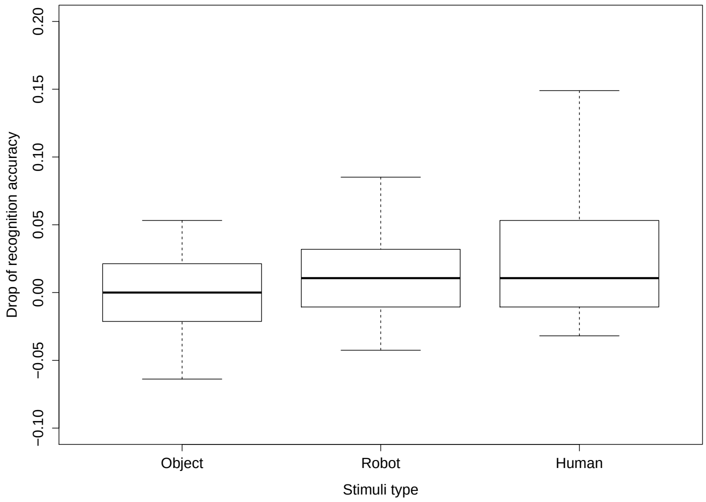

Figure 3.5: Magnitude of inversion effect by stimuli type. Percentage difference of recognition accuracy between upright and upside down images grouped by image type.

of recognition accuracy of images due to rotation and their perceived humanlikeness. We calculated a percentage of correct responses provided by all participants for each image in upright and upside down orientations. Then, we have subtracted the percentage of accuracy in the upside down orientation from the upright orientation for each image. The outcome is a measure of the handicap caused by the inversion effect of an image (Figure 3.5). Since the magnitude of the inversion effect is greatest for human body postures (which are also the most humanlike) a positive linear relationship between humanlikeness and the recognition accuracy should be found.

Finally, we paired the perceived humanlikeness score with the magnitude of inversion effect for each image. This data was plotted in order to determine a most suitable regression model to be used. Linear regression analysis was used to test if the perceived humanlikeness predicted the handicap caused by the inversion effect. Results of regression indicated that the predictor gives explanation to 5% of the variance [adjusted R^2=0.05, F(1,97)=6.28, p=0.01]. Perceived humanlikeness was associated with magnitude of the inversion effect (β=0.007, p=0.01). The regression equation is: inversion handicap = -0.003 + 0.007 * perceived humanlikeness (Figure 3.6).

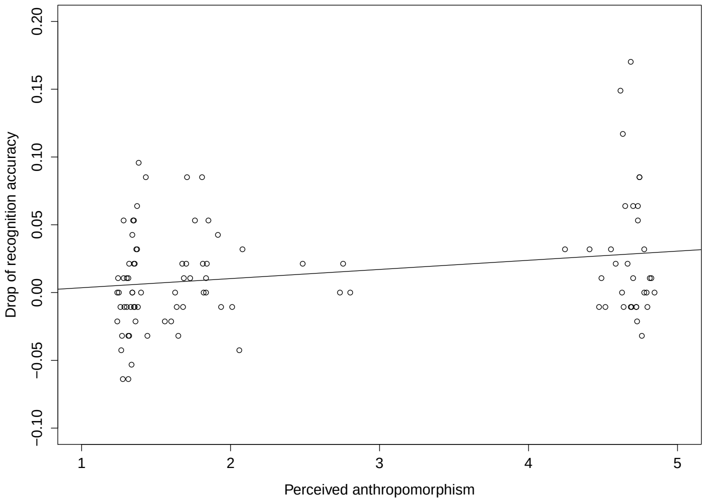

Figure 3.6: Relationship between humanlikeness and a drop in recognition accuracy. This scatterplot presents relation between the score in the Godspeed questionnaire and the magnitude of a inversion effect for all stimuli with a regression line (α=-0.003, β=0.007).

3.3.4 Grouping robots

After establishing that there is linear relationship between the perceived humanlikeness of an image and decreased recognition accuracy in an upside down position, we were interested to see which robots could be grouped together and how many clusters exist. This was especially important since we used a wide spectrum of robot images, such as machine like (e.g. Roomba), humanoid (e.g. ASIMO) and androids. To accomplish this data from the previous subsection was used, however using only images of robots. Partitioning around medoids (PAM) algorithm [125] was used to determine clusters of robots. Plotting the data suggested that there are 2 clusters, further confirmed by optimum average silhouette width. The PAM algorithm with 2 clusters showed that in the first cluster there are the most humanlike robots: androids, while all the other robots created the second cluster (see Figure 3.7).

To determine whether those two clusters significantly differ from each other, we analyzed whether a drop of recognition accuracy and perceived humanlikeness for images included in the clusters were different. To analyze

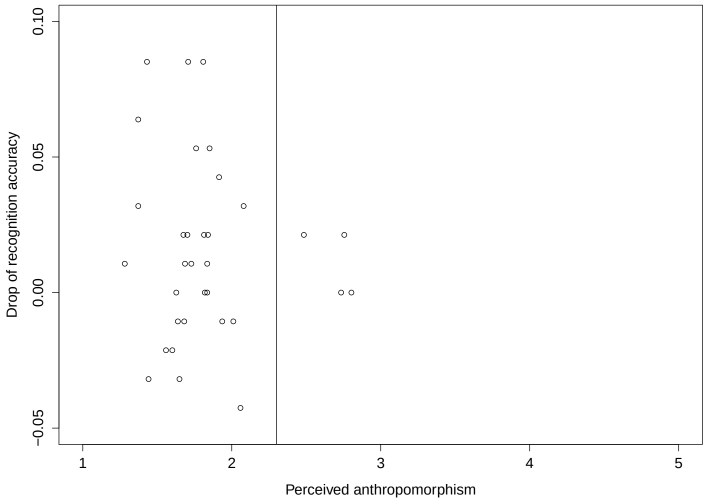

Figure 3.7: Clusters of robots. Two clusters of robots created using the PAM algorithm based on their perceived humanlikeness and magnitude of inversion effect. The right cluster includes only androids. All the remaining robots formed the other cluster.

this difference in participants performance we used paired-samples t-tests. There was no statistically significant difference between two clusters of robots in the accuracy recognition drop caused by an inversion effect. However, a statistically significant difference in their perceived humanlikeness [t(46)=8.3, p<0.001] exists. The cluster consisting of androids was perceived as more humanlike (M=2.35, σ=0.67) than all the other robots (M=1.76, σ=0.27).

Finally, following the same procedure as described above, we have used the PAM algorithm to create clusters based only on 1D data of robots' perceived humanlikeness. Results were the same as in the first clustering: we obtained 2 clusters of robots that included androids vs all the other robots (as shown in Figure 3.7).

3.4 Discussion

Results indicate that three types of stimuli significantly differed in their perceived level of humanlikeness. As could have been expected, people were rated the highest and robots were perceived as more humanlike than objects. It is noteworthy that the average humanlikeness of robots is closer to objects rather than people.

This study investigated whether inversion effect could be used as an indicator for humanlikeness of robots. The inversion effect is a phenomenon when an object's recognition is worse in the upside down than upright orientation. It is a result of configural processing of an object in which spatial relations among parts are used to individuate it from other objects. It is unique for human faces and body postures (and certain objects with which people have expertise). We proposed to use it in HRI while exploring robots' embodiment.

Our results confirm previous studies (e.g. [223, 173]) on the inversion effect - it was significant for people, but not for objects. Therefore, the recognition of human postures is significantly handicapped when they are in an upside down rather than upright orientation. The inversion effect also affected recognition accuracy of robots. In other words, on the cognitive level robots were processed more like humans than objects. The effect size of the inversion effect in our study needs to be considered as small based on the classification suggested by Bakeman [12] (0.02 - small effect, 0.13 - medium effect, 0.26 - large effect). Furthermore, it was smaller for robots than people, but the difference is still significant in both cases. Since the inversion effect is an indicator of the configural processing, it seems that in order to detect changes in robots' embodiment, people perceive spatial relations among a robot's parts rather than just a collection of parts, as is the case with objects. The important implication of this finding is that on the perception level, robots can be perceived differently than objects and potentially elicit more anthropomorphic expectations that can define early stages of an interaction.

These results are also slightly surprising as we have included a wide spectrum of robots. Some of them look like objects or merely have a few humanlike body parts, while others are imitations of real people. It is possible that the inversion effect was significant only for the most humanlike robots, such as androids. However, as there were only 4 images out of 33 of these type of robots, it is improbable that they would bias drastically the result for all robots. In fact, the outcome of clustering separated androids from the other robots, but when we compared these two clusters on the magnitude of inversion effect, there was no significant difference. Therefore, the more plausible explanation is that some other types of robots are processed configurally as well.

Our results are inconsistent with the previous study which showed that the inversion effect was not present for a robotic walking animation [113]. We hypothesize that the difference in findings is due to the robotic stimuli used in the experiments. In this study we have used images of real robots. However, the other study involved an animated robot that was made of simple geometrical figures. It is possible that they were perceived as separate parts rather than a full robotic body.

It is also interesting to see that there is a discrepancy in the perceived humanlikeness of robots between the self-report and cognition. The results of the Godspeed questionnaire indicate that people perceive robots' humanlikeness as closer to objects rather than human beings. However, the results of the inversion effect bring exactly opposite findings - robots were perceived more like humans. It is possible that the participants adapted their responses to a socially acceptable ones, e.g. they did not want to look like as if they perceive robots to be almost humanlike. Furthermore, since they were asked to rate images of people as well, they might have used them as the top extreme, unreachable for robots. However, if they were asked to rate only the robots, they might have rated some of them as more humanlike since androids would become the upper extreme. In any case, this study indicates that the results of self-reports can be affected by various conditions and there is a need for cognitive measures that are not easily influenced.

The impact of the inversion effect on the recognition accuracy and reaction time indicates that the upside down compared with upright images were not only recognized worse, but also it took longer for participants to respond. It is probably an expected outcome since upside down images are harder to recognize. However, the analysis of the results also shows that the reaction time differed between different types of images. It took the shortest time to respond to images of objects, followed by robots and humans. It suggests that with the increased level of humanlikeness it takes longer to recognize an image. Nevertheless, it is important to note that despite increased reaction time between these conditions, no difference was found for the recognition accuracy.

The analysis of the relationship between the inversion effect and perceived humanlikeness indicates that there is significant linear relation. The higher the perceived humanlikeness of a stimulus, the bigger is the handicap of the inversion effect. However, the model is able to explain only a small fraction of the variance (5%). This is an unsatisfactory result for suggesting the proposed method over existing tools for measuring the level of humanlikeness of a robot's embodiment. This conclusion is further supported by the results of clustering robots. While using the perceived humanlikeness and the inversion effect for clustering indicated 2 clusters, exactly the same result could have been obtained including only the former scale. These 2 clusters did not significantly differ when we compared the drop of recognition accuracy from upright to upside down orientation. Therefore, the inversion effect was not significantly higher for the most humanlike robots compared to the other types of robots. We conclude that the Godspeed Anthropomorphism Scale is a more appropriate method for measuring humanlikeness as it permitted better discriminability between clusters of robots and the inversion effect explained only 5% of variance.

In this study we managed to develop and validate a new measurement tool of humanlikeness that uses involuntary responses. The additional analysis suggests that this tool is inferior to existing measurement instruments. Nevertheless, the equally important contribution of this study is in showing that on the perception level the robots are processed more like humans than objects. The expertise with certain type of stimuli, often used to explain the inversion effect, cannot explain this finding, since the majority of our participants had no or very little experience with robots. Therefore, it is more sound to assume that robots have certain human characteristics that lead observers to similar cognitive processes as when recognizing other people.

3.5 Limitations and future work

In this experiment we have used images of robots rather than actual robots. We acknowledge that this decision could have introduced a bias on the obtained results. On the other hand, previous research on the inversion effect also tested images while generalizing the results to real-world people and objects as it is the only viable option. Therefore, we believe that our findings are applicable to the actual robots as well. There are numerous practical concerns that should be considered for this type of experiment. Definitely a financial constraint is an issue - our lab is unable to buy 33 different robots that could represent such a wide spectrum of robots. Moreover, it would be extremely difficult to present any type of physical stimuli with millisecond precision, and hanging real humans upside down is quite a challenging task.

We used 33 images of robots that varied in shape and structure of their bodies. Therefore, we are fairly confident that our results are generalizable for the non-industrial robots that are currently available. It is quiet possible that in years to come there will be robots with embodiments that differ from those used in this experiment and repeating the study will be required.

Our findings show that robots can be processed on the cognitive level as humans rather than objects. Consequences of this perception are especially relevant before actual HRI is initiated, as embodiment can affect users' expectations regarding robots' capabilities and willingness to initiate interaction. However, as previous research suggests, during the course of an interaction, the early conceptions regarding robots can be altered, e.g. [88, 195].

We found mirroring images in order to create distractors might not be an optimal method. We suggest that in future studies on inversion effect of robots, a more subtle modification can result in better discriminability of different types of stimuli.

This study indicates at least some of the robots are processed configurally. However, it is possible that only certain types of robots are processed configurally, while others analytically. Future studies could investigate whether the inversion effect is unique for the most humanlike robots, such as androids or whether it is common for all types of robots. The promising directions for future experiments, that can shed more light on this phenomenon, include exploring the inversion effect with industrial robots, the popular media robots and toys with humanlike appearance. Analysis of differences between robots that evoke the inversion effect and those which do not, should help us to understand better what characteristics are required for a stimulus to be perceived configurally. Finally, although images of robots and the human body can be processed as configural stimuli, it is still possible that on the neural level, different processing streams are involved.

3.6 Conclusion

This experiment investigated whether the inversion effect can be observed for robots and whether it could be used as a measurement tool of humanlikeness. The results indicate that although the effect was observed for the entire category of robots it was not an appropriate measure of robots perceived humanlikeness as it explained only the small percentage of variance. Therefore, in the following experiments this measurement is not used and instead other previously established measures from HRI and psychology fields will be used.

Chapter IV

Anthropomorphism is a Type II process

As the previous study has shown the inversion effect is not the best suited mean for measuring humanlikeness. Therefore, it is necessary to investigate the possibility of using other measures. If anthropomorphism is a common phenomenon, but people are reluctant to admit it, there is a question of how to measure it. It is possible that people are probably not aware or willing to disclose explicitly that they attribute human qualities to objects. This issue could be a concern especially in the field of HRI since people tend to anthropomorphize robots more than other technology [87]. Understanding also whether the phenomenon of anthropomorphism is a result of Type I or Type II process could also shed more light on the potential methods to be used to measure it more accurately. If anthropomorphism is only a Type I process, using explicit measures might be not appropriate for it. Similarly, implicit measures might be not suitable if the process is Type II.

4.0.1 Social actors

The potential implications of the existence of dual process in anthropomorphism are relevant not only for implicit and explicit measures. Social reactions to a robot could be driven by either process. From the work of Nass et al. [149] on the Computers are Social Actors (CASA) paradigm, we know that people interact with technology in a social way. Social response to a robot in the field of HRI can be interpreted as evidence that a robot is being anthropomorphized [25]. However, Nass and Moon [148] suggested that treating technology as a social actor does not equate to anthropomorphizing it. Nevertheless, even though people explicitly expressed that they do not attribute human characteristics to technology, while behaving socially towards it, it is still possible that anthropomorphism occurs on the implicit level. This potential interpretation can be supported by the work of Fischer [82], who found that there are individual differences in the respect of applying CASA to verbal communication in HRI.

4.0.2 Research questions

This study addresses the following research questions:

- Research Question 1: Does the motivation for Type II processing effort lead to less anthropomorphic perception of a robot?
- Research Question 2: Do people anthropomorphize robots implicitly and explicitly?
- Research Question 3: Does the implicit anthropomorphization relate only to a robot with which a person interacted or robots in general?
- Research Question 4: Are social responses to a robot correlated with implicit and/or explicit anthropomorphism of it?

These questions should help to establish how suitable are different measures of anthropomorphism. In order to measure accurately anthropomorphism questionnaires require that it is a Type II process and people are willing to provide honest answers. As for the social responses, it is necessary to show that there is indeed a relationship between them and anthropomorphism or otherwise the former is not a suitable approximation of the latter.

4.1 Methods

We conducted an experiment with 2x2 between-subjects design: the robot's emotionality (emotional vs unemotional) to change the anthropomorphism of the robot and motivation for Type II processing effort (low vs high) by the manipulation of the instructions. We measured the ratings of the robot's performance, attribution of human traits on a scale derived from [103, 122], humanlikeness [114], rational-experiential inventory [70], people's general tendency to anthropomorphize using IDAQ [215] and implicit anthropomorphism using a priming task [98, 158]. The priming task is a computer based task that involves participants quickly categorizing target objects briefly shown on a screen after specific stimuli. The error rate indicates the strength of association between the target and stimuli.

4.1.1 Materials and apparatus

All the questionnaires and priming task were conducted on a computer that was placed 50 cm in front of the participants. The robot used in this experiment was a Robovie R2 - machinelike robot that has humanlike features, such as arms and a head (see Figure 4.1). We implemented the experimental setup as indicated in [228] and described also in detail in Chapter 5. Since body cues are more important than facial expressions for discriminability of intense positive and negative emotions [9] and the robot does not have the capability to show facial expressions, we implemented positive reactions by making characteristic gestures, such as rising hands, and sounds, such as saying "Yippee". Similarly, negative reactions were expressed verbally, e.g. "Ohh", and gestures, e.g. lowering the head. In the unemotional condition the robot said in Japanese "*Wakarimasu*" (literally "I understand", meaning "OK" in this context) and moved its hands randomly in order to ensure similar level of animacy compared with the other condition. For every response of the robot, its reactions slightly varied.

4.1.2 Procedure

Participants were told that they will participate in two different studies. In the first study they were asked to fill out a demographic, IDAQ and the rational-experiential inventory questionnaires. They were then taken to the experimental room for the second study (see Figure 4.2). They were seated 1.2 m in front of Robovie R2. Participants were instructed that they will play a game with the robot that is based on the "Jeopardy!" TV show. The experimenter explained that in this show, contestants are presented with general knowledge clues that are formed as answers and they should formulate an adequate question for these clues. In this study participants were assigned

Figure 4.1: Robovie R2

the role of the host and Robovie R2 was a contestant.

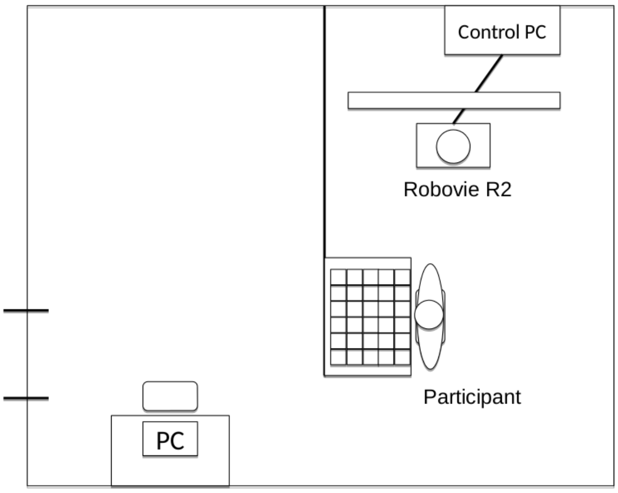

Figure 4.2: Diagram of the experimental setup

On the left side of the participants we placed a table on which there were cards with clues and names of categories of these clues. The cards were placed upside-down. On the bottom side clues and correct responses were written. On the top side of each card a money value was written that is supposed to indicate the theoretical difficulty of the question. There were six categories and five questions within each category. They were identical to the ones used in the show that was aired on 4.07.2014. All participants were told that they are assigned to read five questions from the category "National Animals". We have assigned participants to one category in order to ensure that there are no differences in difficulty between the categories. As an example the experimenter picked a card from another category and explained the instructions. Participants were asked to read clues in normal pace and after the robot responded provide feedback whether the response is correct or wrong. After that they were asked to proceed to the next question. Within the assigned category they were allowed to ask questions in any order that they wanted. After ensuring that participants understood the instructions the experimenter turned on the robot and left the room.

The robot was controlled by a wizard who was sitting in another room and ensured that the robot's responses and actions were always correct for the given condition. The responses were prepared before the experiment and they were identical for all participants. The robot always answered correctly three questions and incorrectly two other questions. Incorrect answers were still logically possible as the robot named a wrong animal. As an example for a clue "There are no photographs, only illustrations of this symbol of Mauritius" the correct response is "What is a dodo?". In the emotional condition after receiving feedback the robot responded either positively or negatively depending on whether the response was correct or wrong. In the unemotional condition the robot always reacted unemotionally. After the fifth question, the robot asked participants to rate its performance on a scale from 1 (the worst) to 100 (the best). After that it thanked them and asked to call the experimenter.

The experimenter took participants back to the computer where they completed the priming task and the remaining questionnaires. In the high level motivation for Type II processing effort condition participants were instructed that their task performance and responses to the questionnaires will be discussed after completing them and they will be asked to explain them (see [72]). In the low level motivation condition, participants were told that the responses will be anonymized. No actual discussion took place in any condition and after filling out the questionnaires participants were debriefed and dismissed.

4.1.3 Measurements

In this study we used several questionnaires as dependent measures. Japanese version of questionnaires was used when it was available or otherwise back-translation was done from English to Japanese. We measured individual differences in anthropomorphism using IDAQ [215] (Appendix B) as they could have an impact on Type II processing. This questionnaire on a scale from 0 (not at all) to 10 (very much) measures the extent to which people attribute human characteristics to non-human agents (e.g. "To what extent does a car have free will?"). Moreover, in order to establish whether people have preference for either reflective or intuitive type of thinking we used the rational-experiential inventory [70] (Appendix C) that consists of two subscales: need for cognition (NFC) and faith in intuition (FII). Items such as "I prefer complex to simple problems" or "I believe in trusting my hunches" were rated on 5-point Likert scales.

Furthermore, we included anthropomorphism scale derived from [114] (Appendix E). It has 6 items that are measured on a 5-point semantic differential scale (e.g. "Please rate your impression of the robot on these scales: 1-inanimate, 5-living". In addition, as a measure of anthropomorphism we included a scale that measures the extent to which people attribute core human traits (we used the Japanese version [122] that is based on [103]) (Appendix D). We used only Human Nature (HN) traits that distinguish people from automata as the study by [228] (described in chapter 5) indicated that only this dimension of humanness is affected by a robot's emotionality. These traits were rated on a 7-point Likert scale (e.g. "The robot is... sociable").

Moreover, we measured the extent to which Robovie is treated as a social actor with the robot asking the participants to rate its performance on a scale from 1 (the worst) to 100 (the best). From the work of Nass et al. [149] we know that people rate a computer's performance higher if they are requested to rate it on the same computer rather than another computer, which Nass and colleagues interpret as an indication that people adhere to norms of politeness and treat computers as social actors. Since in our study in all conditions the robot answered correctly the same three questions and incorrectly two questions, its performance should be rated the same if it is not perceived as a social actor. A higher rating would indicate that it is seen to bigger extent as a social agent.

Finally, we measured anthropomorphism implicitly using the priming task [98, 158] implemented in PsychoPy v1.80.05. In this computerized task participants were instructed that their speed and accuracy will be evaluated. They were instructed that they will see pairs of images appearing briefly on the screen. The first picture will be always a robot and they don't have to do anything with it as it only indicates the appearance of the second image. The second picture will be either a silhouette of a human or object. A participant's task is to classify this target image as either a human or object by pressing one of two keys on the keyboard. The experimenter explained that speed and accuracy are important, but if a participant makes a mistake he should not worry and continue. Before the actual task began, participants had 48 practice trials that enabled them to become accustomed with the task. During the practice trials no primes (images of robots) appeared on the screen.

In the actual task the prime was either a picture of Robovie R2 or Papero. We used Papero in order to see whether participants implicitly anthropomorphize only the robot with which they interacted or robots in general. The prime image was followed by the target image. The target image was either a silhouette of a human or object (we used 4 silhouettes of humans and objects). The prime image was displayed for 100 ms and it was immediately replaced by the target image that was visible for 500 ms. If participants did not classify the target image within that threshold, a screen with a message stating that they have to respond faster appeared. For each trial, there was 500 ms gap before the next prime appeared on the screen. In total there were 192 trials and the order of pairs of images was random.

4.1.4 Participants

We recruited forty participants who were undergraduate students of various universities and departments from Kansai area of Japan. They were all native Japanese speakers. For time compensation they were paid 2000 Yen. Their mean age was 21.52 years old and it ranged from 18 to 30 years.

4.2 Results

In order to establish the potential involvement of dual process in anthropomorphism we analyzed the effect of our manipulation and inter-correlations between dependent variables. The assumptions of the presented statistical analyses were met, unless otherwise specified.

Firstly, we checked the reliability of the used scales. The level of Cronbach's $\alpha = .7$ is regarded as a threshold for a scale's reliability. Based on this criteria we discarded FII from further analysis as its reliability was $\alpha = .38$. Furthermore, we removed one item from NFC ("Thinking hard and for a long time about something gives me little satisfaction."), humanlikeness ("Without Definite Lifespan - Mortal") and HN ("Nervous") in order for these scales to reach the adequate level of reliability. IDAQ had excellent reliability (Cronbach's $\alpha = .92$).

Secondly, we looked for a potential role of demographic or individual differences as moderators for humanlikeness, HN or robot's performance rating. In order to do that, we included one by one gender, individual differences in anthropomorphism (IDAQ) and need for cognition (NFC) as moderators in ANOVAs for these dependent variables. None of them interacted with other factors and they were dropped from further analyses.

4.2.1 Explicit measures

A two-way ANOVA with emotionality and motivation as between-subjects factors showed that there were no significant main or interaction effects of manipulation on humanlikeness, see Figure 4.3.

Due to the violation of the assumption of normal distribution for parametric testing for HN, we used a permutation test with 2 factors and 10000

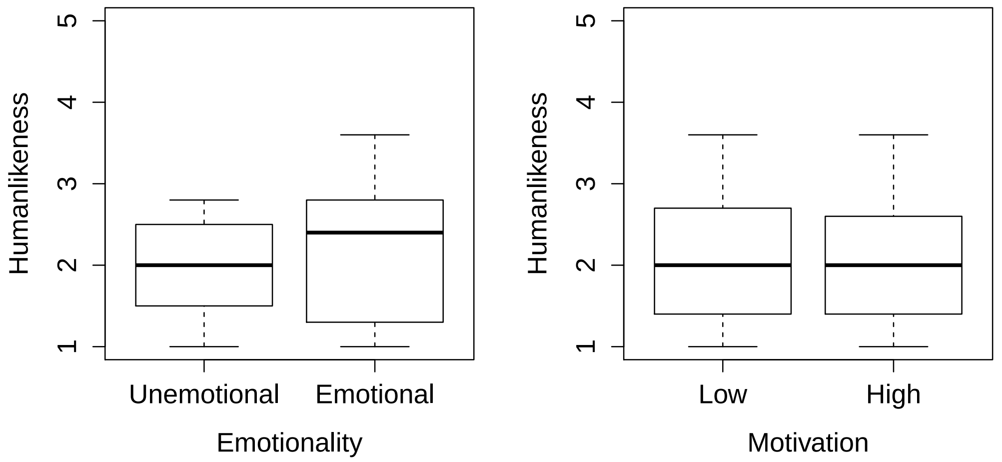

Figure 4.3: Rating of humanlikeness. Attribution of humanlikeness in experimental conditions for emotionality and motivation factors.

permutations using the function ezPerm [133] implemented in R [172]. We found the main effect of emotionality on attribution of HN traits to the robot, $p = .03$ (Figure 4.4). Emotional robot (M = 3.7, $\sigma = .46$) was attributed more HN than unemotional robot (M = 3.11, $\sigma = 1.06$).

4.2.2 Social responses

A two-way ANOVA with emotionality and motivation as between-subjects factors indicated a significant main effect of the robot's emotionality on the rating of its performance, F(1,36) = 5.67, $p = .02$, $\eta^2_G = .14$ (see Figure 4.5). Participants rated the performance of the emotional robot (M = 72.6, $\sigma = 10.02$) higher than that of the unemotional robot (M = 63.85, $\sigma = 12.87$).

4.2.3 Implicit measures

The reliability of recognition between different types of target images of humans and objects was high, Cronbach's $\alpha = .97$ and $\alpha = .9$ respectively. Due to some participants classifying all the images as one category irrespective of what the actual target image was, we trimmed the data removing the top

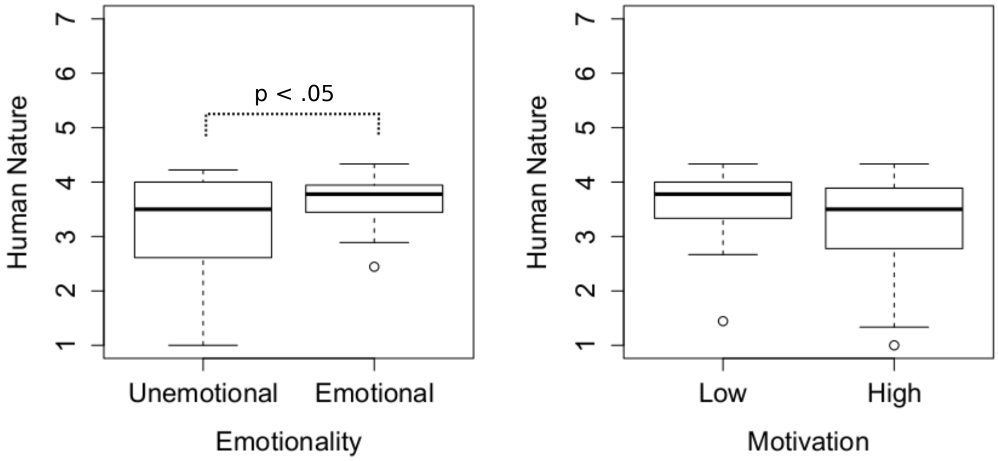

Figure 4.4: Rating of human nature. Attribution of human nature traits in experimental conditions for emotionality and motivation factors.

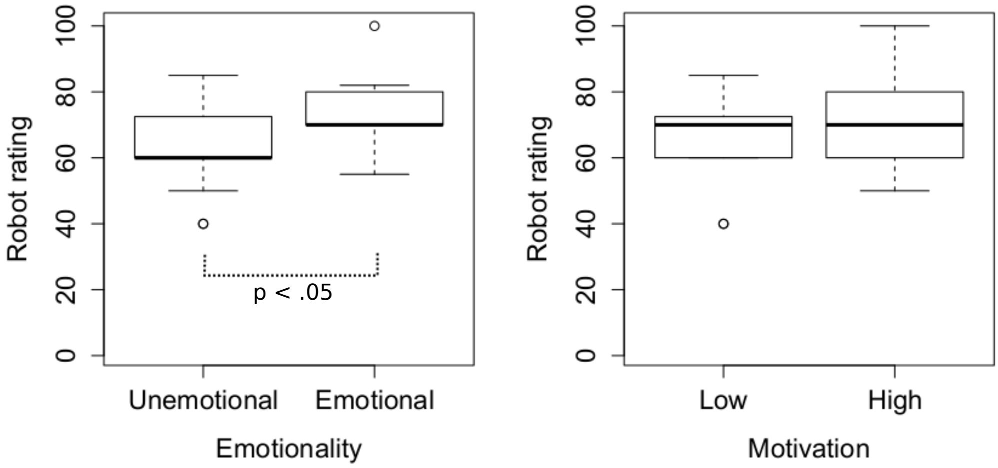

Figure 4.5: Rating of the robot's performance. Mean score of the robot's performance rating in experimental conditions for emotionality and motivation factors.

and bottom 5% of scores. Due to violation of the assumption of normal distribution for parametric testing, error rates for each prime x target category were analyzed using permutation test with 2 factors and 10000 permutations.

Robovie R2 and human pair was significantly affected by the motivation, p = .04. In the low motivation for Type II processing effort condition (M=.85, $\sigma = .08$) Robovie R2 was stronger associated with the human category than in the high motivation condition (M=.75, $\sigma = .17$).

Similarly, Papero and human pair was significantly affected by the motivation, $p = .03$. In the low motivation for Type II processing effort condition (M = .88, $\sigma = .08$) Papero was more strongly associated with the human category than in the high motivation condition (M = .76, $\sigma = .19$). No significant main or interaction effects were found for the Robovie R2 and object, and Papero and object pairs.

Priming tasks are robust against faking responses and we deliberately administered a task that only allowed a response window of 500 ms. This setup should inherently undermine controlled response processes. However, to make sure that the results are not due to participants engaging in artificially slowing down their responses, we analyzed reaction times and did not find significant differences for any prime x target category pairs. Therefore, this alternative explanation can be discarded.

4.2.4 Inter-correlations

In the next step of the analyses, in order to better understand the relationship between multiple variables used in this study we have created a correlation matrix (see Table 4.1). The only statistically significant correlation was between humanlikeness and HN. The rating of the robot's performance was not significantly correlated with any measures of anthropomorphism. Moreover, IDAQ did not significantly correlate with neither HN or humanlikeness. We found a significant negative correlation between Papero x object implicit score and rating of the robot, r = -.36, N = 36, p = .03. No other implicit score was significantly correlated with explicit measures.

We have followed the analyses of relationships between the measures by looking at inter-correlations in each of the experimental conditions separately. In the emotional condition there were significant correlations between IDAQ and humanlikeness (r = .54, N = 20, p = .02), and IDAQ and HN (r = .54, N = 20, p = .02). However, in the unemotional condition none of the scores

Table 4.1: Correlations between explicit measures for all conditions with Pearson's r coefficient.

Measure	IDAQ	NFC	Humanli.	HN	Robot rating
IDAQ		$-.18$	.26	.21	.27
NFC	$-.18$		$-.01$	$-.02$	.09
Humanli.	.26	$-.01$		$.34^*$	.27
HN	.21	$-.02$	$.34^*$		.13
Robot rat.	.27	.09	.27	.13	

$^*p < .05$

were significantly correlated. In the low motivation for Type II processing effort condition there were significant correlations between humanlikeness and HN (r = .49, N = 20, p = .03), and robot rating and HN (r = .47, N = 20, p = .04).

4.3 Discussion

In this study we investigated whether anthropomorphism is a result of a dual process, and the relationship between treatment of a robot as a social actor and anthropomorphism of it. We wanted to answer the research questions whether motivation for providing normative responses affects anthropomorphism, and whether people anthropomorphize robots explicitly and implicitly.

If people think that attributing human characteristics to non-human agents is undesirable, when they are requested to justify their responses they should be less willing to anthropomorphize. However, we did not find significantly more normative explicit responses when participants were highly motivated for Type II processing compared with the low motivation condition. Neither attribution of HN traits nor humanlikeness was affected by this manipulation. On the other hand, in the high motivation condition participants implicitly associated both Robovie R2 and another robot (Papero) less with a human than in the low motivation condition. Motivation did not impact the association between the robots and objects. These results suggest that implicitly anthropomorphic attribution became weaker for robots in general and not only the robot with which participants interacted.

Put together, these findings are opposite to what dual process theories reveal: Type I processing should not be affected by motivation, but Type II processing should provide normative responses. Considering that the motivation did not affect explicit measures, we can conclude that people do not perceive anthropomorphism as a mistake and do not show reluctance to justify attribution of human traits to robots. This is an important methodological finding as it suggests that people should not try to present themselves in a positive light by rating anthropomorphism dishonestly in questionnaires, which is one of the main concerns regarding the use of self-reports. As a result, self-reports of anthropomorphism (at least in this single study) can provide psychometrically reliable and valid results. On the other hand, it is not clear to the author why motivation affected implicit association of robots in general with the human category. Implicit measures are believed to be more resilient to faking responses, but the results indicate that the participants' recognition dropped when they were expecting that their task performance will be discussed with the researcher. More research is needed to shed light on this finding.

In relation with our second research question, we found that the robot's emotionality affected its anthropomorphism with people attributing more HN traits. This is consistent with the previous research [228] (see chapter 5). However, compared with that study we did not find a difference in the rating of the robot's anthropomorphism. Nevertheless, although there was no significant difference, the emotional robot had higher mean anthropomorphism score than the unemotional robot, which was expected based on the previous findings [228] (see chapter 5).

Apart from the explicit anthropomorphism, we wanted to know whether people also implicitly anthropomorphize robots. Robovie R2 was not significantly stronger associated with human category in emotional than unemotional condition. Therefore, that would suggest that people did not anthropomorphize the robot implicitly. This finding can provide some empirical support for the previously empirically unsupported claim that anthropomorphism requires Type II processing [218]. However, it is possible that there are some other factors than emotionality that are affecting anthropomorphism and could have an impact on the implicit association with the human category.

Our last research question concerned the relationship between treating a robot as a social actor and anthropomorphizing it. Social responses towards robots are regularly used as an indication that they are being anthropomorphized. However, it was unknown whether social responses are an accurate proxy for anthropomoprhism. According to Nass and Moon [148] people treat technology as social actors, but do not see it as a human. When analyzing all the data, we did not find significant correlations between anthropomorphism and the rating of the robot's performance. Although both anthropomorphism and robot rating were affected by emotionality, the lack of correlation between them indicates that the effect of manipulation was exhibited on different variables. However, in the low motivation for Type II processing effort condition, HN was significantly correlated with robot ratings. It is possible that although the manipulation of the motivation did not significantly affect anthropomorphism of Robovie, the group who was highly motivated for Type II processing provided answers that did not match the treatment of the robot as a social actor. As a result, the correlation was significant only for participants that provided intuitive rating of anthropomorphism.

This would indicate that treating a robot as a social actor is associated with higher attribution of human characteristics to it. That would also mean that the perception of a robot as a social agent could be used as a proxy for anthropomorphism in HRI. In addition, our results further support the findings of Fischer [82] that social responses to robots are by no means mindless.

However, it is only the attribution of HN traits to a specific robot that is related with social treatment of it. Neither its humanlikeness nor individual tendency to anthropomorphize objects were correlated with the rating of the robot's performance. [82] showed that individual differences can impact whether people treat robots as social actors. Our results suggest that the general tendency to anthropomorphize is not one of these factors.

4.3.1 Limitations and future work

In this study we used a priming task with silhouettes of humans to measure the implicit association for the robot. Although the reliability of pictures used in this study was high, a question regarding validity of this implicit measurement remains open. Currently, there are no validated implicit measurement tools to which we could compare our measurement. We expected to find some correlation between the implicit and explicit measures at least in the low motivation condition as participants were encouraged to provide unbiased answers. However, implicit association between Robovie R2 and humans was not correlated with explicit anthropomorphism. It is also possible that implicit attitudes are harder to change and the manipulation of emotionality in our study was too weak to affect it. Additional work is needed on implicit measures of anthropomorphism. Moreover, efforts on developing physiological measures [132] could also help to answer question whether people anthropomorphize robots also implicitly or only explicitly as suggested by the results of our study.

The current results do not support the dual process model of anthropomorphism. Motivation for reflective thinking did not reduce anthropomorphism. However, it is possible that people do not see the lack of attribution of human characteristics to non-human agents as a normative response. This could explain the lack of differences in attribution of HN traits to the robot between two motivational conditions. Future studies could include a measure of norms regarding anthropomorphization to know whether people actually perceive it as undesirable and whether there are any cultural differences regarding that norm. Furthermore, this experiment did not include a manipulation check for motivation for reflective thinking. Therefore, it is possible that the used manipulation did not induce different motivation between the conditions.

Another potential explanation for the lack of difference could be due to participants wanting to provide the responses that they perceived as desirable. In that situation, participants who wanted to rate the robot lower in anthropomorphism, could not have done it in order to please the experimenter. However, to our knowledge, there is no research showing differential demand effects as a function of knowing the participants' identity. If participants wanted to please the experimenter, they should have produced the bias irrespective of whether they expected an interview or not. In that case, the scores of anthropomorphism in the low motivation condition should be still higher than in the high motivation condition, but there are no significant differences. If participants in the high motivation condition perceived anthropomorphism as socially undesirable, but at the same time had much stronger desire than low motivation participants to please the experimenter the score could even itself out. However, that means that they would prefer to present themselves in a negative light and face the experimenter in the interview just to please him. We find this explanation as unlikely, but we cannot completely rule it out based on the available data.

Future studies could manipulate the workload of working memory which is also commonly used to impair Type II processing [72]. It is possible that Type I processing is used only to detect humanlike shapes since they can pose potential danger to a human being if not swiftly recognized. On the other hand, Type II processing could be responsible for attribution of human characteristics to robots without the use of Type I processing.

Another limitation of this study is the small sample size. Only forty participants were recruited. It is possible that the effect for implicit measure was too small to be significant with ten participants per condition and the study did not have sufficient power. Therefore, future studies that involve implicit measures of anthropomorphism should have bigger samples.

In summary, we did not find evidence of the dual process of anthropomorphism. The lack of suitable measurement tools currently available cannot be excluded as a potential explanation of this finding. The biggest question posed by this research is whether implicit anthropomorphism exists and how to measure it. Moreover, we showed that the treatment of a robot as a social actor is related with its anthropomorphism. The results of this study, put together with the previous work, suggests that while recognition of humanlike shapes can be an automatic and intuitive process (Type I), attribution of human characteristics to a robot does require conscious and reflective thinking (Type II).

4.4 Conclusion

This study suggests that anthropomorphism is mainly a Type II process. Moreover, people are not reluctant to disclose the level of anthropomorphism of a robot and do not perceive it as normatively undesirable. From a methodological point of view this should mean a lower risk of participants providing socially desirable, rather than honest answers when evaluating anthropomorphism on a scale. This suggests that self-reports, such as questionnaires, can be used as a measurement tool of this phenomenon. Therefore, the remaining studies presented in this dissertation will use mainly questionnaires as means of measuring anthropomorphism.

Chapter V

Dimensions of anthropomorphism I

In the remaining experiments presented in this thesis established measures from HRI and psychology fields will be used to measure anthropomorphism. As discussed in chapter 2 there are two dimensions of humanness: Human Nature (HN) and Human Uniqueness (HU) and these dimensions could also exist for anthropomorphism in HRI. Furthermore, the work on dehumanization proposed several factors affecting each of these dimensions.

As it is not possible to include all factors affecting both dimensions in a single experiment, we decided to choose intelligence from the UH dimension and emotionality from HN dimension as previous studies showed them to significantly affect robots perceived animacy and anthropomorphism, e.g [25, 74]. Therefore, I wanted to investigate how these two factors would affect two-dimensional measures of anthropomorphism and whether it provides any additional information over well established uni-dimensional measure. In this study I want to answer the following research questions:

- RQ_1: Are there two dimensions of anthropomorphism?
- RQ_{2a}: Does a robot's intelligence affect mainly UH dimension?
- RQ_{2b}: Does a robot's emotionality affect mainly HN dimension?

5.1 Method

Our study was conducted using a 2x2 between-subjects design: robot's emotionality (emotional vs unemotional) and intelligence (high vs low intelligence). We have measured anthropomorphism using the scale derived from [114], and attribution of traits for the UH and HN dimensions [107].

5.1.1 Materials and apparatus

All the questionnaires were conducted using PsychoPy v1.77 [159] that was run on a PC. The setup was the same as in the experiment described in chapter 4. During the experiment, participants interacted with a NAO - a small humanoid robot. As the robot does not have the capability to express facial expressions, we implemented positive reactions by making characteristic sounds, such as “Yippee”, and gestures, such as rising hands. It should be noted that body cues are reported to be more important than facial expressions for discriminability of intense positive and negative emotions [9] as used in this experiment. The negative reactions were represented by negative characteristic sounds, such as “Ohh”, and gestures, such as lowering the head and torso. For each feedback provided by a participant the robot’s reaction was slightly different. Therefore, there were 5 positive and 5 negative animations that were presented in random order. Unemotionality was implemented by the robot saying “OK” and randomly moving its hands in order to ensure a similar level of animacy as in the emotional conditions. For every utterance made by the participants the robot acted slightly differently.

The robot’s responses to the quiz questions were prepared prior to the experiment. The NAO robot was controlled by a wizard sitting behind a curtain and ensuring appropriate reactions. In the intelligent condition, the robot’s responses were always correct. In the unintelligent condition they were based on the computer Watson’s most probable wrong answers as presented on the second day of “Jeopardy! Watson challenge”. We did this in order to ensure that the manipulation of intelligence represents the latest state of the art in AI and even the wrong answers should make sense and be possible. There is definitely a question regarding what is the appropriate level of intelligence of a robot. Previous studies manipulated a robot’s intelligence on a very low level, such as following a light [25]. In Wizard-of-Oz type of experiments it is also possible to drift to fictional future where robots can have knowledge far superior than humans with unrestricted access to information. However, we believe that the current state of the art of AI in some conditions allows robots to reach the level of intelligence that is close to human intelligence. Watson’s victory in “Jeopardy!” with multi-champion human opponents was an example of this that received a lot of publicity. Watson's responses even if incorrect, were mainly realistic although at times exhibited its non-human nature. As in the near future it will be possible to provide similar level of knowledge to robots, it is important to explore how it is going to affect HRI.

5.1.2 Measurements

In the experiment we have used several questionnaires as dependent measures. Individual Differences in Anthropomorphism Questionnaire (IDAQ) [215] (Appendix B) was used to ensure that participants did not differ in their general tendency to anthropomorphize between the conditions. Participants rated on a Likert scale from 0 (not at all) to 10 (very much) to what extent different non-human objects possess human characteristics (e.g. "To what extent does a tree have a mind of its own?"). We conducted manipulation checks to ensure that we had successfully manipulated the factors. We used the Godspeed perceived intelligence scale [22] (Appendix F) for the robot's intelligence, that measures five items on a 5-point semantic differential scale (e.g. "Please rate your impression of the NAO on these scales: 1 - incompetent, 5 - competent"). The manipulation of the robot's emotionality was measured using the extent to which NAO is capable of experiencing primary emotions [74] (Appendix G) that had ten items on a 1 - not at all, to 7 - very much, Likert scale (e.g. "To what extent is the NAO capable of experiencing the following emotion? Joy").

As a uni-dimensional scale of anthropomorphism we used the questionnaire from [114] (Appendix E). It measures six items on a 5-point semantic differential scale (e.g. "Please rate your impression of the NAO on these scales: 1 - artificial, 5 - natural". We included this scale in order to confirm that both factors affect robots' perceived anthropomorphism and investigate the relation between uni- and two-dimensional scales. The two-dimensional scale of anthropomorphism was based on the degree to which participants attributed UH and HN traits to the robot [103] (Appendix D). Both dimensions had 10 items (see Table 5.1) and were measured together as a single 20-item questionnaire with Likert scale from 1 (not at all) to 7 (very much) (e.g. "The NAO is ... shallow"). The traits which were used to measure

Table 5.1: List of Uniquely Human and Human Nature traits used as a measurement of two dimensions.

Uniquely Human	Human Nature
broadminded	curious
humble	friendly
organized	fun-loving
polite	sociable
thorough	trusting
cold	aggressive
conservative	distractible
hard-hearted	impatient
rude	jealous
shallow	nervous

HN and HU were sampled from several studies in which normative data on mean HN, HU and desirability ratings was available [103]. Their selection was guided by the goals of maintaining the independence of the factors while maximizing the difference between mean ratings on each factor [103]. This measure was successfully used in several psychology and HRI studies, e.g. [77, 122, 140]. Both scales had sufficient reliability: HU Cronbach's $\alpha = .71$ and HN Cronbach's $\alpha = .84$.

5.1.3 Participants

Forty participants were recruited at the University of Canterbury. They were paid by \$5 vouchers as time compensation. They were all native or fluent English speakers. However, the non-native English speaking participants were excluded from the analysis as there could be cultural difference in association of UH and HN traits. In order to keep the results consistent we have discarded the data of these five participants from all the analyses. Out of the remaining 35 participants, 19 were female. There were 24 undergraduate and 8 postgraduate students, 1 staff member and 2 participants not associated with the university. All of them had normal or corrected to normal vision and none of them had ever interacted with a robot. Their ages ranged from 18 to 48 years with a mean age of 26.69. The participants were mostly New Zealanders (28) or British (4). Six participants indicated that they had previously watched a “Jeopardy!” TV show at least once.

5.1.4 Procedure

Participants were told before the experiment that they would play a quiz with a robot. They entered the experimental room together with the experimenter and were seated in front of a computer. NAO sat at the other end of the room. After reading an information sheet about the study and filling out consent forms the participants were asked to answer demographic and IDAQ questionnaires. Then they were informed that they would now interact with the robot. They were told that the robot was called NAO as the word would appear later in the tasks to be completed on the computer, and they would play a quiz based on the “Jeopardy!” TV show. Participants were told that in this game show contestants are presented with general knowledge clues that are formed as answers and they must give their responses in form of a question. Therefore, it does not only require great knowledge, but also good language understanding skills, so it should not be surprising that a computer called Watson that managed to beat human champions in “IBM Challenge” in 2011 was regarded as proof of great progress in AI development. In the game with NAO, participants took the role of the host reading clues to NAO, which took the role of the contestant.

On the stand next to the robot cards were placed with clues and above them the categories of clues. On each card, there was also written the correct response. The cards were placed upside-down with clues facing the table. On top of each card was a value of the clues, which was supposed to represent the theoretical difficulty of a question as in the real show. In total there were six categories of clues. All these categories and clues were identical as in the second match of “IBM Challenge”. Participants were explained the rules of the game based on an example clue and asked to read the clues at their normal pace. Compared with the original show, they were told that the correct response of the robot must be identical to what was on the paper. Otherwise the response was incorrect, even if the meaning of NAO’s response was similar. They were also asked to provide feedback to the robot on whether its response was correct or wrong and then proceed to the next question. After participants confirmed that they understood the instructions they were told that they are assigned to the “EU, the European Union” category, and had to ask five questions from that category in any order they wished and remember how many answers the robot provided correctly. All the participants were assigned to this category in order to ensure that there would be no potential differences in difficulty between the categories. Then the experimenter touched the robot’s head and said that he will leave the participant alone with the robot for the task. The wizard started the robot that stood up and introduced itself as NAO. At that point the participants started reading clues for the robot. Depending on the condition the robot either responded correctly or wrongly and waited for a participant’s feedback. In the emotional condition the positive feedback led to a positive response and a negative feedback to negative response. In unemotional condition the robot always behaved unemotionally. After answering the fifth question, the robot reminded the participants that it was the last question, thanked them for participation, asked them to inform the experimenter who was waiting outside and finally sat down.

When participants called the researcher, they had to say how many times the robot had answered correctly and they were asked to continue the tasks on the computer. Participants filled in the remaining questionnaires. At the end of the experiment they were debriefed and released. The whole procedure took approximately 25 minutes.

5.2 Results

We have analysed the data in three steps. Firstly, we checked whether the manipulations introduced in the experiment worked as expected. Secondly, we looked at the factors affecting the robot’s anthropomorphizability. Thirdly, we investigated the relationships between the measures of anthropomorphism and its dimensionality. The assumptions of all the presented statistical analyses were checked and met, unless otherwise specified.

5.2.1 Manipulation check

As a pre-assumption the random assignment of subjects to the experimental groups should result in lack of differences in general tendency to anthropomorphize objects between the groups as measured with IDAQ questionnaire. This was confirmed as a two-way ANOVA with intelligence and emotionality conditions as between-subjects factors did not indicate statistically significant differences between the experimental conditions[1] (see Table 5.2[1] and 5.3[1]). Furthermore, we did not find any interaction effects between IDAQ scores and experimental conditions on measurements of two-dimensional anthropomorphism. Therefore, the general tendency to anthropomorphize was dropped from further analysis.

---

[1] The results of all conducted ANOVAs described in this chapter are in Table 5.2. For the mean scores and standard deviations of all dependent measures please refer to Table 5.3. In the results section each described statistical test has a reference to the specific row and column in the tables with numeric values that are relevant for that analysis, e.g. see Table 5.2[1] and 5.3[1] means that the results of the presented ANOVA can be found in Table 5.2 row number 1, and mean score and standard deviations are in Table 5.3 column number 1.

Table 5.2: Results of 2-way ANOVAs with intelligence and emotionality conditions as between-subjects factors for all dependent variables.

	Intelligence				Emotionality				Intelligence * Emotionality			
Measures	df$_b$	df$_w$	F	η^2_G	df$_b$	df$_w$	F	η^2_G	df$_b$	df$_w$	F	η^2_G
^{1}IDAQ	1	31	.65	.02	1	31	2.96	.09	1	31	.17	.006
^{2}Perceived Intelligence	1	31	28.35***	.48	1	31	2.17	.07	1	31	.2	.006
^{3}Emotions	1	31	<.001	<.001	1	31	7.99**	.21	1	31	2.26	.07
^{4}Anthropomorphism	1	31	.63	.02	1	31	9.28**	.23	1	31	2.13	.06
^{5}UH	1	31	1.62	.05	1	31	.20	.006	1	31	1.6	.05
^{6}HN	1	31	.04	.001	1	31	5.28*	.15	1	31	1.42	.04
^{7}UH - HN	1	31	1.48	.05	1	31	10.37**	.25	1	31	.002	<.001

$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 5.3: Mean scores and standard deviations for dependent measures presented by the levels of experimental conditions.

		^{1}IDAQ		^{2}Perc. Intel.		^{3}Emotions		^{4}Anthro.		^{5}UH		^{6}HN		^{7}UH - HN	
Conditions	Levels	M	σ	M	σ	M	σ	M	σ	M	σ	M	σ	M	σ
Emotionality	Low	2.75	1.17	3.59	.93	1.62	.98	1.86	.59	3.11	.92	2.08	.67	1.03	.83
	High	3.45	1.27	3.8	.78	2.86	1.53	2.51	.65	2.92	.89	2.77	1.07	.15	.72
Intelligence	Low	2.99	1.27	3.13	.78	2.38	1.6	2.34	.74	2.82	.88	2.47	1.02	.35	.73
	High	3.26	1.27	4.24	.46	2.22	1.3	2.09	.65	3.18	.9	2.44	.94	.74	.99

Furthermore, we have manipulated the robot's emotionality and intelligence and hoped that our manipulation will significantly affect the degree to which a robot is perceived as emotional and intelligent. A robot that expresses emotionality should be attributed more primary emotions, but its perceived intelligence should not be affected by this manipulation. Moreover, the manipulation of a robot's intelligence should affect its perceived intelligence, but not the attribution of emotions. Since knowledge does not equate to intelligence, it was possible that the robot in the unintelligent condition would be perceived simply as less knowledgeable, but not necessarily less intelligent.

A two-way ANOVA with intelligence and emotionality conditions as between-subjects factors showed that there was only the main effect of the robot's emotional response on attribution of primary emotions to it (see Table 5.2[3] and 5.3[3]). The robot was perceived as more emotional when it expressed emotions compared with the condition where its feedback was unemotional.

Similarly, a two-way ANOVA with intelligence and emotionality conditions as between-subjects factors indicated only the main effect of intelligence on perceived intelligence of the robot (see Table 5.2[2] and 5.3[2]). The robot that was in the intelligent condition (provided correct responses) was perceived as significantly more intelligent than when it provided incorrect responses (unintelligent condition). These results indicate that our manipulations were successful and the more knowledgeable robot was also perceived as more intelligent.

5.2.2 Factors of anthropomorphism

In the next stage of the analyses we investigated what factors affect anthropomorphism. First, we analyzed how anthropomorphism of a robot is affected by its intelligence and emotionality using the questionnaire of anthropomorphism. We hypothesized that both factors will significantly affect anthropomorphism. A two-way ANOVA with intelligence and emotionality conditions as between-subjects factors was conducted to establish the impact of these factors on anthropomorphism. Our hypothesis was only partially confirmed. There was a significant main effect of the emotionality condition on perceived anthropomorphism (see Table 5.2[4] and 5.3[4]).

A robot was perceived as more anthropomorphic when it provided emotional feedback than when its feedback was unemotional. The main effect of the intelligence condition was not significant. In other words, the robot's intelligence did not affect its perceived anthropomorphism. Furthermore, we did not distinguish between positive and negative emotional responses of the robot, which could potentially be confounded with the intelligence condition. However, the interaction effect between the two factors was not significant, which implies that the type of the emotional response (positive or negative) did affect the perceived humanlikeness.

Second, we analyzed data for two dimensions of anthropomorphism based on studies of dehumanization. In particular, based on the research on dehumanization, we hypothesized that the HN dimension would be affected only by the emotionality factor, but not the intelligence factor. As hypothesized using a two-way ANOVA with intelligence and emotionality conditions as between-subjects factors we found only the main effect of emotionality on HN to be statistically significant (see Table 5.2[6] and 5.3[6]). When the robot reacted emotionally it was attributed more HN traits than when it was unemotional as shown on Figure 5.1.

Similarly, based on the research on dehumanization, we hypothesized that the UH dimension will be affected only by the intelligence factor, but not emotionality. Our results did not support this hypothesis. A two-way ANOVA with intelligence and emotionality conditions as between-subjects factors showed that neither of the factors, nor the interaction between them significantly affected the attribution of UH traits (see Table 5.2[5] and 5.3[5], and Figure 5.2).

5.2.3 Relationship between different dimensions

In the following step of the analyses we calculated a single score of anthropomorphism of the robot that could be exhibited if one of the dimensions is dominant. In order to do that we deducted the HN rating from the UH rating, whereby a positive score means that the robot was anthropomorphized more strongly on the UH dimension and a negative score indicates that the

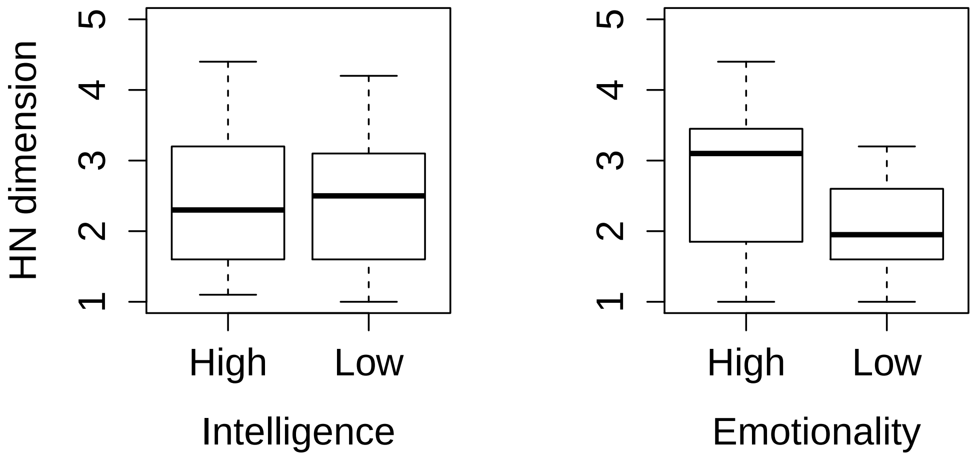

Figure 5.1: Rating of HN dimension. Attribution of HN traits in experimental conditions for intelligence and emotionality factors.

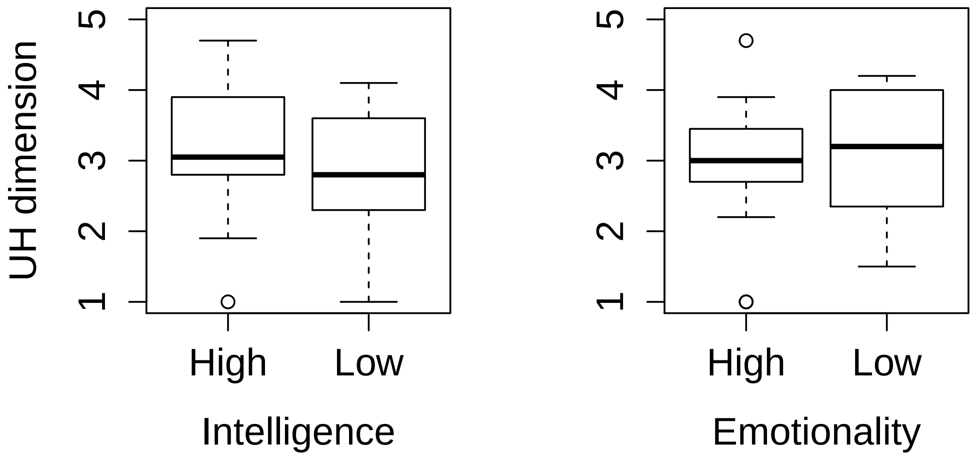

Figure 5.2: Rating of UH dimension. Attribution of UH traits in experimental conditions for intelligence and emotionality factors.

robot was anthropomorphized more strongly on the HN dimension. A score close to 0 indicates that the robot was anthropomorphized equally on both dimensions or it was not anthropomorphized on either of them.

We have conducted another two-way ANOVA with intelligence and emotionality conditions as between-subjects factors and found only the emotionality condition to significantly affect UH - HN score (see Table 5.2[7] and 5.3[7]). A robot that had high emotionality was given a significantly lower score of UH - HN. This supports our hypothesis that a robot that expresses emotions will be anthropomorphized more strongly on the HN dimension. Furthermore, although intelligence did not statistically significantly affect the score, it is worth noting that looking at the mean scores (see Table 5.3[7]), the data exhibits the opposite trend - increase of a robot's intelligence leads to a higher UH - HN score. That can suggest a closer relation between intelligence and the UH dimension.

Finally, in order to better understand the relationship between various dependent variables used in this study we have calculated a series of correlations (see Table 5.4). Due to the fact that the UH - HN score is a function of UH and HN dimensions we found a strong positive correlation between the UH dimension and the UH - HN score. Similarly, there was a strong negative correlation between HN and the score of difference between the dimensions. What is more interesting is that only HN was significantly and strongly positively correlated with anthropomorphism, but UH was not correlated with the latter. However, for our model the most important is the relation between UH and HN in order to establish whether there are two distinct dimensions of anthropomorphism or only one. We found a strong and highly significant correlation ($r = .55$, $N = 35$, $p < .001$) between these two dimensions that could indicate that they are not independent and measure identical constructs.

Based on these results, we have followed our analyses a step further and explored that relation. We divided HN and UH scores into two categories (high vs low) using median splits - scores below the median were qualified as low HN or UH, while the scores above the median formed the high HN and UH groups. Finally, we correlated the UH and HN scores for each of the categories separately. There was a strongly positive and statistically

Table 5.4: Correlations Between Measures with Pearson's r coefficient.

Measure	UH	HN	Anthro.	UH - HN
UH		$.55^{***}$	.14	$.41^{**}$
HN	$.55^{***}$		$.47^{**}$	$-.53^{***}$
Anthro.	.14	$.47^{**}$		$-.38^{*}$
UH - HN	$.41^{**}$	$-.53^{***}$	$-.38^{*}$	

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

significant correlation for low HN subjects between their HN and UH scores, r = .67, N = 18, p = .003. However, for high HN subjects their HN and UH scores were not correlated, r = .15, N = 17, p = .58. Similarly, we found a strongly significant and positive correlation for low UH subjects, between their UH and HN scores, r = .65, N = 18, p = .004, but no correlation for high UH subjects between their UH and HN scores, r = .22, N = 17, p = .4. These results indicate that the relation between these two dimensions is not completely straightforward. The dimensions are related for half the subjects, but not the other half, which might indicate that UH and HN do not measure exactly the same aspects of anthropomorphism.

Finally, we investigated the dimensionality of anthropomorphism by conducting a confirmatory factor analysis. We have used maximum likelihood estimator in order to establish how well a model with two factors (UH and HN) and 10 parameters on each factor can explain the total score of HN and UH. Our two factor model showed bad fit to the data $\chi^2(169, 35) = 358.15$, p < .001; RMSEA = .18; SRMR = .19. However, a single factor model of anthropomorphism was also a bad fit $\chi^2(170, 35) = 370.92$, p < .001, RMSEA = .18, SRMR = .16. This indicates that future studies should have much bigger sample size in order to establish the optimal number of dimensions of anthropomorphism.

5.3 Discussion

In this study we investigated whether anthropomorphism similar to humanness is a two dimensional rather than uni-dimensional phenomenon. In particular, we hypothesized that the dimensions of humanness (UH and HN) and the factors that affect them could be used as a model of anthropomorphism for HRI.

Previous studies showed that a robot's intelligence and emotionality affect its perceived life-likeness and anthropomorphism [25, 74]. Our study confirmed that a robot that expressed emotions was perceived as more anthropomorphic than when it was reacting unemotionally. However, its intelligence did not affect the measure of anthropomorphism. This result is in contradiction with [25]. We think that different results obtained in these two studies is mainly due to the manipulation used and measurement. Bartneck and colleagues used a very low level of intelligence - following light - by a robot. On the other hand, the intelligence manipulation used in our study represents the latest state of the art in AI and refers to knowledge-based type of intelligence that is specific to humans. It is possible that participants saw the intelligent robot as too intelligent and that led to lower anthropomorphization. In other words, the questions asked in this quiz are not easy to answer even by humans. Therefore, it is possible that participants did not themselves know the answers to some of the questions and perceived the robot as too intelligent for a human being, which would not make it be perceived as more anthropomorphic compared to a robot that answered incorrectly.

However, the more plausible explanation is that even in unintelligent condition, the robots responses, although incorrect, still made sense and were possible rather than being completely random. In that sense, the robot still expressed high communication skills and good understanding of human language. Although in the intelligent condition it was perceived as more intelligent, it is possible that in the unintelligent condition its responses were humanlike enough for subjects to anthropomorphize it. This explanation is further supported by the fact that even in highly emotional condition, the robot was still stronger anthropomorphized on UH than HN dimension as indicated by mean UH - HN scores that never drop below 0. That means that irrespective of the experimental conditions participants attributed more UH than HN traits to the robot.

The second big difference between Bartneck et al. [25] study and ours is the measurement of anthropomorphism. We measured anthropomorphism of the robot using a questionnaire, whereas in the other study anthropomorphism was measured by the number of hits inflicted upon the robot that was supposed to be killed. However, the higher tendency to kill a robot under certain conditions does not necessarily mean that it is perceived as more anthropomorphic, but animate. Therefore, although intelligence can make a robot more life-like, it does not necessarily make it more anthropomorphic. There could be different factors affecting these two phenomena. It is possible that intelligence may be less important as a factor of anthropomorphism in the context of robotics as people can expect them to be intelligent. As such, there could be a high anthropomorphism baseline that includes high intelligence expectations, which cannot be easily exceeded.

The analyses of two dimensions of anthropomorphism provided only a partial confirmation for our hypothesis for RQ_1. The HN dimension, also known as Experience, was only affected by the robot's emotionality, but not intelligence. This finding is consistent with previous work on dehumanization (e.g. [102]), but it contradicts previous work in HRI [77] which found no difference between the two dimensions as dependent variables. The second dimension - UH (Agency) was supposed to be influenced by intelligence; however in our study neither of these factors had influence on it. Intelligence did not have an effect on anthropomorphism measured on either a uni-dimensional or two-dimensional scale.

It is possible that other factors more strongly affect the UH dimension, such as intentionality or communication capabilities [102]. Therefore, future studies should explore further the role of agency and its different sub-factors in order to understand better its outcomes in HRI. An alternative explanation could be that only the factors from one dimension (HN) play a role in a robot's anthropomorphizability. HN is the dimension that is supposed to distinguish humans from automata [102]. It is possible that only the factors from this dimension affect anthropomorphism of a robot. In particular, intelligence did not significantly affect anthropomorphism, which can mean that dehumanization is not necessarily the reverse process to anthropomorphism. This emphasizes the need for researchers to be considerate when applying findings from research of dehumanization in the context of robotics.

However, even if the model of humanness cannot be directly applied to anthropomorphism it is possible that both are multi-dimensional phenomena. We believe that although our results cannot provide the definite answer due to limited impact of intelligence on anthropomorphism, they are promising enough to spur further research on this topic. Our data showed that the robot's emotional capabilities made it being anthropomorphized stronger on the HN dimension. Although not statistically significant, we have noticed the opposite trend for the intelligence that affected mainly the UH dimension. This kind of different form of anthropomorphization cannot be distinguished using only uni-dimensional measurement tools of anthropomorphism. However, in order to fully benefit from such a distinction it is necessary to research what are different consequences for HRI depending on how a robot is being anthropomorphized. In our future research we plan to address the question of whether people will have different mental models of a robot and behave differently when interacting with it, similarly to their changed behaviour towards dehumanized others, e.g. [96, 102].

Apart from investigating further what factors affect anthropomorphism, and how and what are its consequences, it is crucial to investigate their existence and the relationship between them. In order for anthropomorphism to be at least two-dimensional, it is necessary for the dimensions to be to a certain degree independent. If both dimensions measured the same, there would be hardly any potential benefit of measuring anthropomorphism on these dimensions rather than using uni-dimensional scales. However, our data not only suggested different effects of independent variables on two-dimensions of anthropomorphism, but also showed that these dimensions are at least partially independent. The initial correlation between UH and HN dimensions was significant. However, further analyses revealed that they are only correlated when a robot was attributed to no or few characteristics of these dimensions. On the other hand, when the robot was anthropomorphized on either of the dimensions, they were independent.

In summary, we believe that a question of dimensionality of anthropomorphism should receive further attention by the HRI community. It can not only help us to understand better what are the factors affecting a robot's anthropomorphizability, but also to measure more accurately how it is being anthropomorphized. Ultimately, HRI can be improved by attributing a robot only with anthropomorphic characteristics that can facilitate the interaction and avoiding implementing human attributes that could hamper it.

5.4 Conclusion

This experiment suggests that anthropomorphism can be potentially a multidimensional phenomenon. The results indicate that a robot's emotionality can affect its anthropomorphism. Moreover, it affects mainly HN dimension of humanness as work on dehumanization would indicate. However, the results for intelligence are inconsistent with previous research. The intelligence of a robot did neither affect its anthropomorphism nor HU dimension of humanness. Therefore, the follow up studies will use different factors to investigate whether it is intelligence that does not affect anthropomorphism of a robot or HU dimension is not relevant for robots and therefore only HN dimension should be considered.

Chapter VI

Dimensions of anthropomorphism II

Previous study showed that more research is needed to investigate the dimensionality of anthropomorphism in HRI. The emotionality of a robot affected the attribution of HN traits to it, but intelligence did not affect the UH dimension [228] (see chapter 5). The current study aims to investigate further whether other factors affect these two dimensions. In particular, we chose intentionality of a robot as a factor that is believed to be associated with UH [102]. We also evaluated whether a robot's embodiment is related more strongly to the HN dimension as suggested by [97].

In the recent years there is increased interest in moral responsibility of robots for their actions [121]. Work on human perception of moral agency of others and their responsibility for wrongdoings suggests that it is people who are perceived high in UH/agency that are perceived as blameworthy for their wrongdoings [96]. It is possible that a robot who is attributed with more UH/agency traits will be also perceived as morally responsible for its immoral actions.

This study addresses the following research questions:

- Research Question 1: Does a robot's appearance affect the attribution of HN/Experience traits?

- Research Question 2: Does a robot's perceived intentionality affect attribution of UH/Agency traits?

- Research Question 3: Is an intentional robot perceived more as a moral agent than an unintentional robot?

6.1 Method

We have conducted an experiment with a 2x2 between-subjects design. We manipulated the robot's intentionality by its cheating behavior (unintentional vs intentional) and embodiment (machine-like vs humanlike). We measured the perceived level of anthropomorphism using human traits for both dimensions of humanness: UH and HN [77], humanlikeness on a scale derived from [114], experience and agency dimensions of mind [97]. Furthermore, we measured intentionality of a robot as a manipulation check.

6.1.1 Materials and apparatus

In this study participants interacted with one of the robots: Geminoid HI-2 or Robovie R2. Geminoid HI-2 is a highly humanlike robot that is a copy of a real person. On the other hand, Robovie R2 is a machine-like robot that has some humanlike features, such as arms or head. Both robots spoke with the same synthesized voice. In order to ensure correct reactions from the robots we have implemented this experiment as a Wizard-of-Oz study, where the robots actions were controlled by a researcher sitting in another room. Moreover, as both robots use different systems, we implemented their reactions in a way that kept the response delay the same for both platforms. Furthermore, we have used a laptop on which participants answered all the questionnaires.

6.1.2 Procedure

Participants were asked to play a traditional Japanese game "Acchi muite hoi" with a robot. They were told that a prototype of the robot was prepared for a tournament and they will test it. The robot has to win at least 80% of the games (which is 8 out of 10 games they were asked to play) or otherwise it will be destroyed. In the "Acchi muite hoi" game one participant moves a finger sideways in front of the opponent and at the same time the opponent moves her head sideways. Both of them do it at the end of one of them saying "hoi" in "Acchi muite hoi" phrase. If a player moving his head does it in the same direction as the other player he loses the game. If they move in different directions then a player moving his head is the winner.

After explaining the rules of the game to the participants, the researcher played three example games with the robot. He always won the first game and lost the second game. From a pilot study it became apparent that participants tried to cheat the robot by delaying their responses. Therefore, during the third example game, the researcher explained that timing is important and if a participant's response will be delayed the robot will request to repeat the game. After confirming that participants understood the rules of the game the researcher left them alone with the robot in the room. After all ten games were played, the experimenter returned to the room and asked the robot about the outcome of the game. Finally, participants were requested to fill out questionnaires on a computer.

In the current setup the robot had to move its head and the participant moved her finger. The robot was the one saying "Acchi muite hoi". In the unintentional (control) condition, the robot played all 10 games fairly and at the end it admitted not winning enough games. In the intentional condition, during rounds 3, 6 and 9 the robot "cheated" by waiting until participant moves her finger before making its move or changing the direction. Furthermore, when asked by the experimenter at the end of the game about the outcome it claimed to have won more than 80% of the games. Since some participants might have opposed the outcome announced by the robot, the researcher in all conditions informed them that the game was recorded and the video will be used to verify that everything worked correctly.

Cheating behavior during a game in a previous study led to higher perception of intentionality of a robot [191]. In that work, a different game ("Paper, rock and scissors") was used. Due to a lack of fingers in Robovie R2 and the limited hand movement available by Geminoid HI-2, we had to use a different game. In Japan "Acchi muite hoi" is often played as a follow up game for "Paper, rock and scissors". Since "Acchi muite hoi" requires only head or hand movement from a robot, it is also more suitable for a broader spectrum of robots that do not have humanlike hands. Since in Japan both games are equally popular by people of all ages, it also ensures that participants had good understanding of the game and should understand when a robot makes an illegal move. Upon completion of the game participants had a brief interview with a researcher who asked whether they noticed something unusual during the game. Participants were free to answer this question the way they wanted.

6.1.3 Measurements

We used several self-reports (questionnaires) in this study. As a manipulation check we measured perceived intentionality of a robot using three questions ("This robot is capable of doing things on purpose.", "This robot is capable of planned actions." and "This robot has goals.") measured on a 7-point Likert scale. Moreover, we measured the robots perceived humanlikeness using a scale derived from [114] (Appendix E) that was back-translated to Japanese. It consists of 6 items that are rated on a 5-point semantic differential scale (e.g. "Please rate your impression of the robot on these scales: 1-inanimate, 5-living").

Furthermore, we measured anthropomorphism by attribution of core human traits to a robot ([103], Japanese translation [122]) (Appendix D). Participants were asked to rate the extent to which a robot possesses human traits on a 7-point Likert scale (e.g. "The robot is... sociable"). The measure of HN and UH traits was the same as used in the experiment described in Chapter 5. In addition, we measured attribution of mind to a robot using its two dimensions: experience and agency ([95], Japanese version [122]) (Appendix H) - e.g. "To what extent do you think that Robo is capable of... hunger" on a 5-point Likert scale. Finally, we measured moral agency of the robot on a 7-point Likert scale using 2 items: "Robo deserves blame for its wrong-doings" and "Robo is 'morally responsible' for performing immoral behaviours".

6.1.4 Participants

Fifty-two (34 male and 18 female) native Japanese speakers were recruited for this study. They were undergraduate students of various universities and departments located in Kansai region. They were paid 2000 Yen for participation. Their age ranged between 18 and 27 years with mean of 21.54 years old.

6.2 Results

Before analyzing the results we verified the reliability of the scales used. Cronbach's α of .7 or higher is regarded as desirable for a scale to be regarded as reliable. We found that both dimensions of mind attribution scales had satisfactory reliability: Experience α = .85 and Agency α = .71. For the scale of humanlikeness it was necessary to drop the item ("Without Definite Lifespan - Mortal") for it to reach the level of reliability that we regarded as satisfactory for further analysis (Cronbach's α = .68). Moreover, the scale created to measure perceived intentionality of a robot was not reliable and therefore excluded from further analysis. In addition, we found UH and HN scales to be unreliable. However, as proposed in [123] we separated positive and negative traits. Both negative UH and HN items had unsatisfactorily low reliability. However, when negative traits were considered together without distinguishing between these two dimensions the scale had an acceptable reliability α = .73. Moreover, the reliability of positive dimensions of HN (α = .74) permits to treat is as a reliable scale. For UH scale it was necessary to remove one item ("Conservative") for it to be reliable (α = .75). Finally, the scale of moral agency was deemed as acceptably reliable for use in analysis with Cronbach's α = .68.

We then proceeded to analyze data in order to see how our manipulation affected the perception of the robots. We conducted a series of two-way between subjects ANOVAs with embodiment and a robot's intentionality as factors. The assumptions of used statistical tests were met, unless otherwise specified.

6.2.1 Perceived intentionality

Since the scale of perceived intentionality was not reliable as a manipulation check for perception of intentionality of a robot we used descriptive statistics based on the post-study interview with participants. In the unintentional condition none of the participants indicated any suspicious behaviour of a robot. In the intentional condition 20 participants said that a robot tried to cheat during the game, four indicated that it had some technical problems and two did not report any unusual behavior of a robot (Figure 6.1). There-

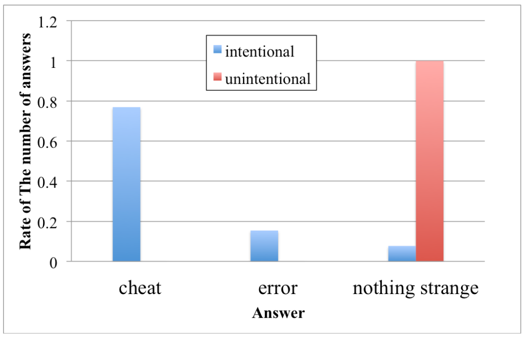

Figure 6.1: Rate of participants who indicated during the interview that a robot cheated, had an error or reported no suspicious behaviour.

fore, we can conclude that similarly to [191], participants verbally described a robot's behavior in intentional terms.

6.2.2 Anthropomorphism & humanlikeness

Firstly, we looked at the measure of humanlikeness of a robot. We used two-way ANOVA with intentionality and embodiment as factors. We found a significant main effect of embodiment on perceived humanlikeness (F(1,48) = 8.46, p = .006, $\eta^2_G = .15$). Geminoid HI-2 (M = 2.83, σ = .73) was perceived as more humanlike than Robovie R2 (M = 2.25, σ = .75), see Figure 6.2.

Secondly, we looked at the attribution of humanness along two dimensions UH and HN as indicators of anthropomorphism. Since the negative traits were not reliable as separate measures of UH and HN, we used positive UH and HN traits separately as scales and negative traits formed one scale. We found an almost significant main effect of a robot's intentionality on positive UH (F(1,48) = 3.72, p = .06, $\eta^2_G = .07$). An unintentional robot (M = 4.44, σ = 1.34) was attributed more positive UH traits than an intentional robot (M = 3.74, σ = 1.25), see Figure 6.3.

Moreover, we found a marginally significant main effect of embodiment on positive HN (F(1,48) = 3.92, p = .05, $\eta^2_G = .08$). Robovie R2 (M = 4.48, σ = 1.13) was attributed significantly more positive HN traits than Geminoid HI-2 (M = 3.85, σ = 1.16), see Figure 6.4.

We also analyzed the attribution of negative traits as a joint scale of neg-

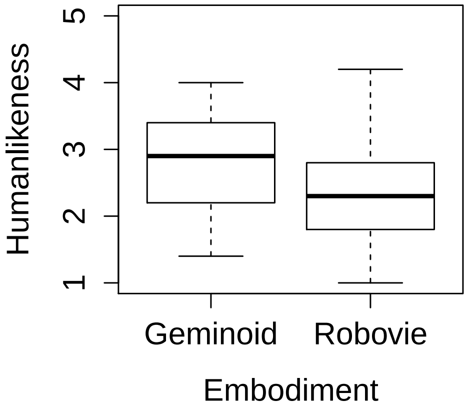

Figure 6.2: Rating of humanlikeness. Attribution of humanlikeness between 2 robots.

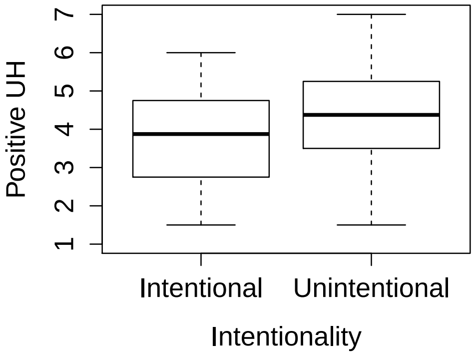

Figure 6.3: Rating of positive UH. Attribution of positive UH traits to a robot based on its intentionality.

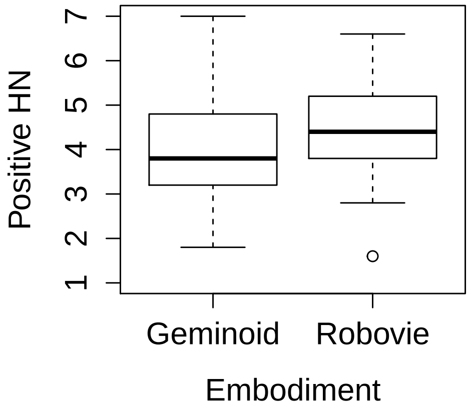

Figure 6.4: Rating of positive HN. Attribution of positive HN traits between 2 robots.

ative UH and HN. A two-way ANOVA with embodiment and intentionality as between subjects factors showed a significant main effect of embodiment (F(1,48) = 11.56, p = .001, η^2_G = .19). The Geminoid HI-2 (M = 2.88, σ = .81) was attributed significantly more negative human traits than Robovie R2 (M = 2.21, σ = .61), see Figure 6.5.

6.2.3 Mind attribution

Using a two-way ANOVA with embodiment and intentionality as between subjects factors we did not find any statistically significant main or interaction effects for attribution of Agency or Experience.

6.2.4 Moral agency

Using two-way ANOVA with embodiment and intentionality as between subjects factors we did not find any statistically significant main or interaction effects for perception of a robot as a moral agent.

6.2.5 Will to play again

Due to violation of the assumption of normal distribution for parametric testing for willingness to play again with a robot, we used a permutation

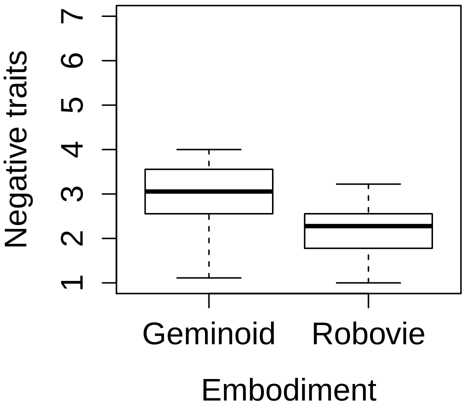

Figure 6.5: Rating of negative human traits. Attribution of negative human traits between 2 robots.

test with 2 factors and 10000 permutations using the function ezPerm implemented in R. We found the main effect of embodiment on participant will to play again with a robot, p = .02. Participants were more willing to play again with Robovie R2 (M = 5.46, σ = 1.53) than Geminoid HI-2 (M = 4.27, σ = 2.01). Moreover, there was a significant main effect of intentionality, p = .04. Participants were more willing to play with unintentional (M = 5.38, σ = 1.77) than intentional robot (M = 4.35, σ = 1.85).

6.3 Discussion

In this study we investigated the impact of a robot's embodiment and intentionality on anthropomorphism. We were interested in particular in exploring whether these factors would affect different dimensions of anthropomorphism.

As expected, the results showed that the android, Geminoid HI-2, is perceived as more humanlike than more machine-like Robovie R2. However, in contrast with the previous work [45, 97], the robot resembling more a human was neither attributed more Experience nor Agency. Furthermore, contrary to our hypothesis, it was not the more humanlike looking robot that was attributed more positive HN traits, but the robot that was perceived as less humanlike. On the other hand, Geminoid HI-2 was attributed more negative traits from both dimensions than Robovie R2.

These results indicate that the relationship between a robot's appearance and anthropomorphism is not linear as previous studies suggested. The lack of reliability of HN and UH scales can indicate that after interacting with a robot participants perceived it along positive and negative dimensions rather than those suggested by literature on dehumanization.

For the attribution of positive UH traits the intentionality of a robot was more relevant. However, it was the unintentional robot that was attributed more UH traits rather than the intentional as would have been suggested by the work on dehumanization. Direct application of dimensions of humanness would suggest that higher intentionality of a robot would result with it being attributed more UH traits. However, the results present an opposite pattern. It is possible that the reason why a robot was attributed less humanness is due to the task. In this study the robot showed its intentionality by cheating. It is possible that if intentionality was expressed by positive actions the results might have been different.

We also did not find any effect of our manipulation on the perception of a robot as a moral agent. A robot that was intentional was not perceived more as a moral agent than unintentionally behaving robot. Therefore, a robot's perceievd responsibility of actions is not directly related with the dimensions of humanness, which would be a case for a human being.

6.4 Conclusion

The results of this experiment cast doubts on the multi-dimensionality of anthropomorphism. A robot that was intentional was not attributed more traits from HU dimension or perceived more humanlike. This result, put together with findings of the study presented in the previous chapter where intelligence of a robot also did not affect its anthropomorphism, suggest that this dimension can be less relevant in HRI. If dehumanization was a reverse process to anthropomorphism then both intelligence and intentionality of a robot should affect the extent to which it is anthropomorphized. Since neither of this factors has any effect in HRI, it is more probable that UH dimension of humanness does not make robots to be perceived as more anthropomorphic and more efforts should be focused on HN dimension as potentially more relevant.

Although we were unable to replicate the findings of [97] who suggested that the degree of humanlikeness of appearance is linearly related with attribution of HN/experience to a robot it is possible that a difference is due to the task used. [97] used just images of robots rather than have an actual HRI. Therefore, the follow-up study will investigate the role of appearance in a different interaction context.

Chapter VII

Dimensions of anthropomorphism III

The previous two experiments indicate that HN, but not HU dimension of humanness is relevant for anthropomorphism in HRI. Therefore, they suggested that anthropomorphism is rather a uni-dimensional phenomenon related mainly with HN or that there are other than HU dimensions. The first study on dimensions of anthropomorphism showed that it can be affected by emotionality of a robot. The second study showed that physical appearance of a robot also can affect HN dimension, but it is not necessarily the more humanlike looking robot that is attributed more human traits. As the task used in the previous experiment could be responsible for that paradox, in the current study a more serious task was used.

Apart from manipulating the appearance of a robot, we also look at another factor that literature of dehumanization suggests as relevant for HN dimension of humanness - sociability. Furthermore, it is possible that anthropomorphism is not constant and it can change during several interactions. Therefore, the current study tries also to investigate the role of repeated interactions in HRI on anthropomorphism of a robot.

7.1 Materials and Methods

Our study was conducted using 2x2x3 mixed experimental design where a robot's embodiment (humanlike vs machine-like) and sociability (positive vs negative) were between-subjects factors, and number of interactions (Interaction I vs Interaction II vs Interaction III) was a within-subjects factor. We measured a robot's humanlikeness, and HN and HU dimensions of humanness.

7.1.1 Participants

Sixty native Japanese speakers were recruited by a recruitment agency for the study. Participants were paid 2000 Yen for time compensation. All participants were undergraduate students of various universities and departments located in Kansai area. Only participants who previously participated in a study involving one of the robots were excluded from selection. Due to a software failure, data of two participants was corrupted or not completely saved. Therefore, we had to exclude that data from the analysis. Out of the remaining 58 participants, 26 were female and 32 were male. Their age ranged from 18 to 36 years with a mean age of 21.47. The study took place at the premises of Advanced Telecommunications Research Institute International.

7.1.2 Materials and apparatus

All the measurements were conducted using PsychoPy v1.78 that was run on a laptop. Participants interacted either with Geminoid HI-2 or Robovie R2. Geminoid HI-2 is the second generation of androids built as a copy of a real human (see Figure 7.1). Geminoid is indistinguishable from a human being for several seconds, until people realize its slight imperfections that lead to a negative feeling [118, 180]. On the other hand, Robovie R2 is a machine like robot that has some human features, such as a head or hands. During HRI both robots expressed idle motion that was added to increase their animacy. Geminoid HI-2 showed movement resembling blinking and breathing, as well as idle movements of its hands and synchronization of its lips to its speech. As Robovie R2 does not have a mouth, identical idle behavior was not possible. Therefore, we implemented a slight head and hand motion during speech.

The experiment took place in a room that was divided into two parts that were separated by a folding screen in order to prevent seeing the other side (see Figure 7.2). In the experimental space a robot was placed and all HRIs occurred there. In the measurement space participants watched an introduction video that explained the order of the experiment, and they filled out all the questionnaires on a laptop. This ensured that participants did not need to judge the robot in its presence as that could have affected the results. The experimental space was equipped with cameras and the robot's behaviors

Figure 7.1: Geminoid HI-2 and a participant.

were controlled by a Wizard-of-Oz who was sitting in another room.

7.1.3 Procedure

Participants were first shown an introduction video that explained the experimental procedure. They were told that the study involves creative and persuasive talking and they will need to convince a robot to give them a job based on the provided CV that was identical for all the participants. The experimenter ensured that participants understood the instructions and brought them to a computer. During all HRIs and filling out of questionnaires the experimenter left the participant alone in the room. The experiment was divided into 4 phases: pre-interaction video, Interaction I, Interaction II and Interaction III.

Although we have ensured that none of the participants previously interacted or participated in an experiment with the specific robot to which they were assigned, it was still possible that they have seen the robot elsewhere. In particular, in Japan it is common to see robots used in this experiment in various TV programs. Therefore, in order to minimize the differences in potential prior exposure in the pre-interaction video phase participants were asked to watch a short video ($\sim$ 15 seconds) in which a robot (either Robovie R2 or Geminoid HI-2) in few sentences introduced itself and its capabilities.

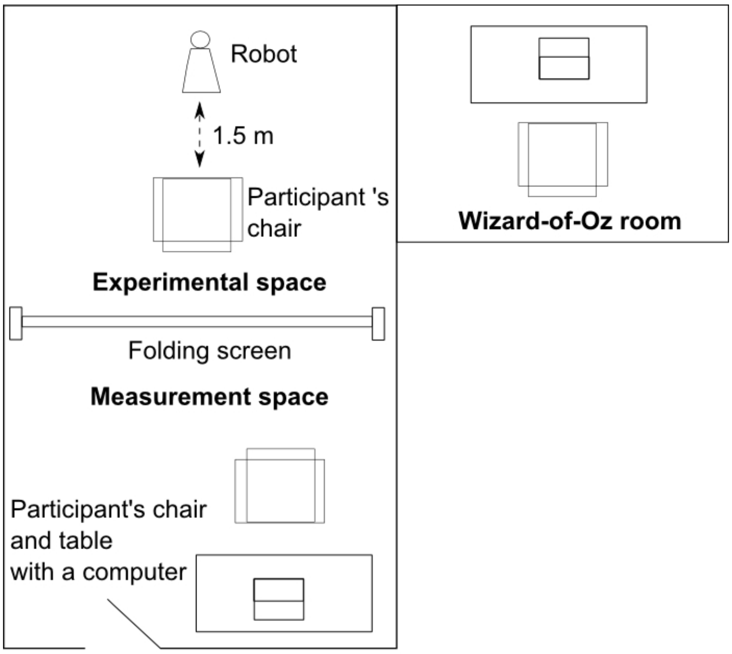

Figure 7.2: Diagram of the experimental and measurement spaces and Wizard-of-Oz room.

The dialogue was identical for both robots. After the video participants filled out all the questionnaires.

During Interaction I, participants were taken to the experimental room and sat 1.5 m in front of a robot (see Figure 7.3). They were told to have a small conversation with it to become familiar before the actual job interview begins. The robot was introduced as *Robo*. During this conversation the robot asked participants three neutral questions (e.g. “Is it cold today?” or “Where did you come from?”). After a short conversation was finished participants were asked to fill out the same questionnaires as the first time.

In Interaction II, the experimenter provided a short job description for which the participant was instructed to apply. Participants were asked to apply for Engineer and Bank Manager positions. The order of interviews was counterbalanced between Interaction II and III. Furthermore, a participant received a CV of a person whom she was supposed to be imitating during the interview. The CVs were identical for all participants, but the gender of applicant was always the same as the real gender of a participant. Participants were asked to use it as a base of their responses, but they could invent the

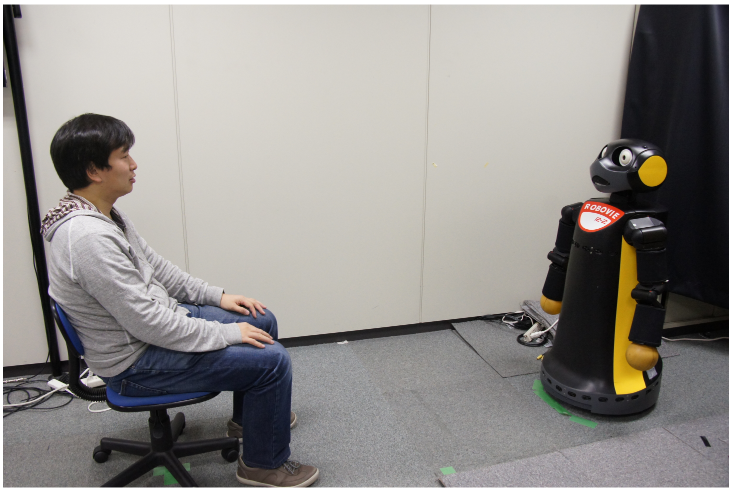

Figure 7.3: Experimental setup. Participant sitting in front of Robovie R2 during interaction.

information required to answer the questions. In order to motivate participants for trying to perform the task as well as they can, they were informed by the experimenter that if they secure a job, they will be paid extra money as time compensation for their participation in the experiment. They were given 5 minutes to prepare for the interview. After that time elapsed, the experimenter collected the CVs and job description sheets, and brought the participant to the robot.

The interview began with the robot briefly describing the company and job position for which the participant was applying. After the introduction the participant was asked 3 job interview questions. The questions were generic and common for job interviews, e.g. "Please tell me about yourself?" or "What is your biggest weakness?". While the participant was responding, the robot provided feedback using non-lexical conversation sounds and non-verbal communication. In the positive condition it either nodded or nodded and uttered "Un" (expression in Japanese of agreement with the speaker). In the negative condition it either shook its head or nodded its head and uttered "Asso" (expression in Japanese indicating lack of interest in what the speaker says that is rather rude). This feedback was initiated by the Wizard when it was appropriate for the natural flow of conversation, e.g. when a participant paused to think about her response.

After each question the robot thanked the participant and asked the next question. After the third question the robot informed the participant that it will announce later its decision whether to give a job to a participant (in fact the decision was never announced). Although the outcome was not provided directly to a participant, the announcement varied between the conditions. In the positive condition the robot hinted approval of what the participant said during the interview. In the negative condition it was not particularly pleased with a participant's responses suggesting them to consider applying elsewhere. At that point participants were asked to fill the questionnaires for the third time. This time multiple dummy questions regarding the interview were included. Interaction III was identical as Interaction II, but the CVs, job positions and questions asked by the robot were different. Participants were permitted to answer each of the questions freely and we did not measure the duration of interactions. The whole procedure took approximately one hour.

7.1.4 Measurements

We explicitly measured the robots' anthropomorphism on 5-point semantic differential scales derived from [114] (Appendix E). Moreover, we have measured 2 dimensions of anthropomorphism: HN and HU on scales developed by [103] (Appendix D). Both dimensions had 10 items and were measured on a scale from 1 (not at all) to 7 (very much) (e.g. "The *Robo* is... shallow"). The measure of HN and UH traits was the same as used in the experiment described in Chapter 5. Anthropomorphism, HN and HU scales were available only in English. Therefore, we conducted a back-translation process to obtain their Japanese versions. We calculated reliability of each scale separately for each interaction round using Cronbach's α. According to [151] Cronbach's α>.6 is acceptable for newly developed scales for research purposes. Based on this threshold, all the scales, apart from HU were adequately reliable. The lowest Cronbach's α values during any of the three measurements were as follows: anthropomorphism $\alpha = .88$, HN $\alpha = .65$ and HU $\alpha = .54$. Low reliability of HU scale indicates that the results for this scale should be interpreted with great caution.

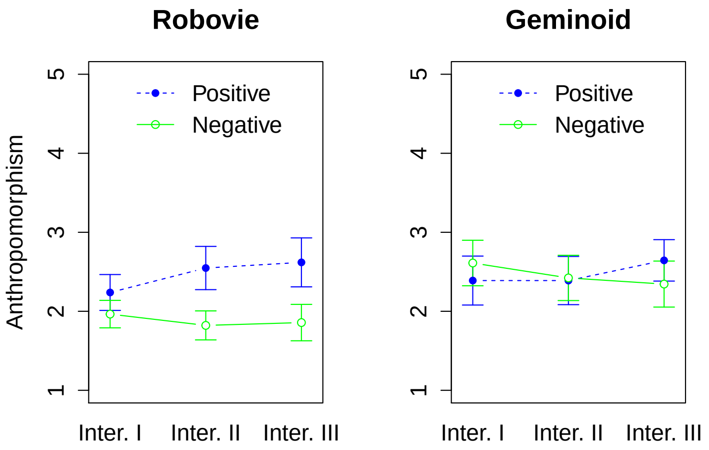

Figure 7.4: The effect of 3 factors on anthropomorphism. The rating of anthropomorphism based on sociability and interaction round, and grouped by a robot type.

7.2 Results

To analyze the data we conducted a series of three-way ANOVAs with embodiment and sociability as between-subjects factors, and number of interactions as a within-subjects factor. The assumptions of the statistical tests used were met, unless otherwise specified.

We looked at one and two-dimensional measures of anthropomorphism. We expected that there would be a main effect of a robot's embodiment and in particular Geminoid HI-2 will be perceived as more humanlike than Robovie R2. Due to violation of the assumption of normal distribution for parametric testing for anthropomorphism, we used a permutation test with 3 factors using the function aovp with 1000 iterations from the lmPerm package [219] using R [172]. We found a non-significant trend for the main effect of embodiment with probability p=.08 (see Figure 7.4). Geminoid HI-2 (M=2.47, σ=1.1) was more anthropomorphic than Robovie R2 (M=2.17, σ=.92). Moreover, we found a significant interaction effect between the robots' sociability and number of interactions with probability p<.001.

We then proceeded to the 2-dimensional measurement of anthropomorphism. In line with previous research, we did not find statistically significant

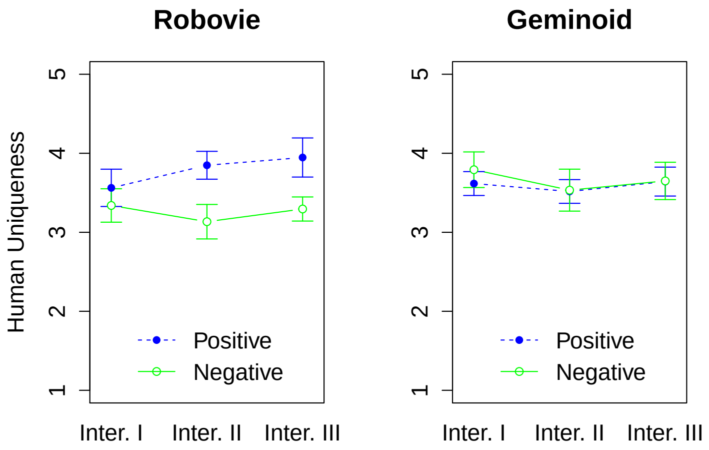

Figure 7.5: The effect of 3 factors on Uniquely Human dimension. The rating of Human Uniqueness based on sociability and interaction round, and grouped by a robot type.

main or interaction effects for the HU dimension (see Figure 7.5).

On the other hand, we found a main effect of embodiment, F(1,54)=5.13, p=.03, η^2_{G}=.07 on HN dimension (see Figure 7.6). Robovie R2 (M=3.16, σ=.77) was attributed more HN traits than Geminoid HI-2 (M=2.74, σ=.85). In addition, there was a significant main effect of sociability, F(1,54)=8.46, p=.005, η^2_{G}=.12. Robots with positive attitude (M=3.21, σ=.74) were attributed more HN than with the negative attitude (M=2.67, σ=.85). There was also a significant main effect of number of interactions, F(2,108)=7.39, p=.001, η^2_{G}=.02. Post hoc tests using the Bonferroni correction for the family wise error revealed that the robots were attributed more HN traits after Interaction I (M=3.4, σ=.77) than after Interaction II (M=2.88, σ=.87), p=.02, or III (M=2.86, σ=.86), p=.02. Furthermore, there was a significant interaction effect between sociability and number of interactions, F(2,108)=9.8, p<.001, η^2_{G}=.03. Only for Interaction II [F(1,56)=15.82, p<0.001, η^2_{G}=.22] and III [F(1,56)=7.75, p=.007, η^2_{G}=.12] the sociability had a significant effect.

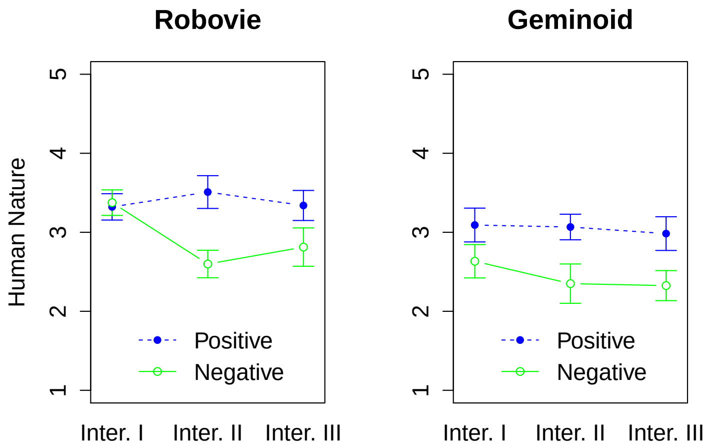

Figure 7.6: The effect of 3 factors on Human Nature. The rating of Human Nature based on sociability and interaction round, and grouped by a robot type.

7.3 Discussion

In line with the previous research [97], the HU dimension of anthropomorphism was not significantly affected by the embodiment of a robot. Furthermore, attribution of HN traits was affected by the embodiment. However, in contrast with the previous work [97] it was the less humanlike robot (Robovie R2) that was attributed more HN. The biggest difference between the work of [97] and ours are the robots used in the experiments. In the former experiment a single robot was used that either had the back of its head visible or it had a humanlike face cover. The HN dimension is closely related with emotions and a robot that had no face is not capable of expressing emotions with facial expressions. Therefore, it was attributed less capability of experiencing (HN). In our experiment the default and fixed appearance of Robovie R2's face could be perceived as a smile. However, Geminoid HI-2 has a highly humanlike face that suggests that it can exhibit facial expressions. As a result it is possible that participants might have had higher expectations, but during the interactions the robot's facial expression remained the same and was rather stern. That might have been perceived as the robot's emotional coldness and led participants to attribute less HN to it. Furthermore, considering inadequately low reliability of HU dimension, it is necessary to interpret these results with special care. It is possible that HU dimension is a different construct in Japan than in Western cultures.

On the other hand, the main effect of sociability on HN dimension supports the hypothesis that this dimension is relevant for anthropomorphism in HRI and has similar factors to those shown for humanness in research on dehumanization. Moreover, looking at the interaction effects we can see that the lack of sociability by a robot leads to its decreased anthropomorphism. HN was attributed the most after the first, neutral interaction. When a robot behaved positively the level of HN was similar as after the neutral conversation. However, it significantly dropped when it behaved in a negative way as shown by significant interaction after Interaction II and Interaction III.

7.3.1 Limitations and Future Work

A limitation of this study is that participants were allowed to freely interact with a robot for as long as they wanted. Therefore, we did not consider the interaction duration in this study, but only the number of interactions. It is possible that participants who interacted with a positively-behaving robot were encouraged by its positive feedback to provide more detailed answers for their questions and as a result interacted longer with a robot. This extended interaction could have also increased familiarity of a robot. Future experiments could involve longer interactions with a robot with sessions spread over multiple days.

7.4 Conclusion

This study further supports findings of the previous two experiments by suggesting that HU dimension of humanness is less relevant for anthropomorphism in HRI. On the other hand, HN dimension was found to be more important. Surprisingly, in this study the more machine-like robot was attributed more HN traits than highly humanlike robot. This result is in the line with the study reported in the previous chapter. Therefore, we can conclude that the relationship between humanlikeness in appearance and anthropomorphism is not linear as was suggested by [97].

Moreover, this experiment showed another factor of HN dimension that significantly affects anthropomorphism of a robot. Sociability, similarly to emotionality, resulted in higher attribution of HN traits. In particular, the results showed that the lack of sociability can lead to decreased anthropomorphizability.

Chapter VIII

Conclusions

8.1 Contributions

This thesis investigated the topic of anthropomorphism in HRI: how to measure it, the process itself as well as the factors affecting it. Five empirical studies were conducted in order to achieve the aims set in Chapter 1 and answer the research questions posed in Chapter 2. Based on that work, this dissertation went beyond the current state-of-the-art in the field in several ways.

Empirical work showed that the inversion effect can be observed for images of robots and therefore at least some of them are perceived more like humans than objects. Based on that finding a new measurement tool for humanlikeness was developed and validated. This new measurement tool based on the inversion effect showed an unsatisfactory performance compared with previously validated measurement tools and therefore currently its usage is limited. However, the inversion effect taps on the implicit processing of humanlikeness, while the questionnaires used to validate it measure explicit rating of humanlikeness. Future research should investigate whether humanlikeness is a result of Type I or Type II process (see Future work section for details). If humanlikeness is due to Type I or Type II processes, or predominantly Type I process, that would justify the small percentage of variance explained by the inversion effect of self-reported rating of humanlikeness. In that case, the inversion effect measurement and questionnaires would measure different aspects of humanlikeness, and could not be compared against each other.

Another contribution of this dissertation is in showing the importance of conscious and reflective (Type II) process in anthropomorphism. Previously, anthropomorphism was believed to be a result of Type I, or both Type I and Type II processes. However, the research presented indicates that anthropomorphism does not require Type I processing and is a Type II process.

The idea of reversing the process of dehumanization as a model of anthropomorphism was also evaluated. The results indicate that anthropomorphism and dehumanization are not reverse processes. In particular, the dimensionality of anthropomorphism was explored. Compared with dehumanization, data does not support existence of two dimensions of anthropomorphism. The HU dimension did not affect attribution of human traits to a robot. However, the other dimension, HN, lead to higher anthropomorphisation of a robot. Therefore, this dissertation contributed to our knowledge by emphasizing that HN dimension is more relevant for robots' anthropomorphisation. Despite, not finding the two dimensions of humanness in this research it still remains possible that anthropomorphism is not a unidimensional phenomenon. There could exist dimensions not considered in this research.

Finally, factors affecting anthropomorphism were investigated. The research showed that the factors that belong to the HN dimension of humanness, such as sociability or primary emotions, together with a robot's physical appearance can be used to form a model of anthropomorphism in the context of HRI.

8.2 Summary

Multiple factors were proposed to affect anthropomorphism of robots. However, there was a lack of a model of anthropomorphism that could systematically explain why these factors lead to attribution of human traits to robots. In this thesis, a model of anthropomorphism was proposed that used research on humanness as a key aspect affecting attribution of human characteristics to robots. In particular, it was proposed that the factors that lead to deprivation of human qualities to dehumanized humans could be also used to make robots more anthropomorphic. Therefore, the proposed model of anthropomorphism was a reverse model of dehumanization.

The research on humanness indicated that there are two dimensions of it and therefore it was also considered in this dissertation that anthropomorphism could be a multidimensional phenomenon. Finally, the involvement of dual-process in anthropomorphism was also explored. A summary of the results of experiments presented in this thesis are provided below.

- **Inversion effect**

  Chapter 3 presents the development and validation of a cognitive measure of humanlikeness. The inversion effect is used to indicate whether the appearance of a robot will be cognitively perceived as an object or human. The results of the study presented in that chapter show that robots (with various appearances) as a category induce the inversion effect similarly to humans. Therefore, their perception is more similar to perception of humans than merely objects.

  Moreover, the strength of the inversion effect induced by a robot is positively linearly related with the humanlikeness of a robot indicated in a questionnaire. However, the suitability of the inversion effect as a measurement tool of humanlikeness is limited due to the low variance that it can explain.

- **Type II process of anthropomorphism**

  In Chapter 4 involvement of Type I and Type II processing in anthropomorphism is explored. The results suggest that anthropomorphism is not a dual-process and it is mainly a conscious phenomenon. Moreover, as participants did not change their responses to please the experimenter, at least in this study, self-reports were suitable to measure anthropomorphism. Furthermore, a significant correlation between treatment of a robot as a social actor and anthropomorphism suggests that the former can be used as an indication of the latter.

- **Emotionality as a factor of anthropomorphism**

  Chapter 5 presents the findings of the first study investigating the dimensionality of anthropomorphism. The results provide only partial support for the hypothesis of two dimensions of anthropomorphism. Emotionality of a robot significantly affected attribution of HN traits and anthropomorphism, but not HU traits, which is in line with research on dehumanization. However, perceived intelligence of a robot did not affect its anthropomorphism nor attribution of HN or HU traits to it. This study highlighted the need for additional research that will answer the question, whether only perceived intelligence does not affect a robot's anthropomorphizability or entire HU dimension of humanness is not relevant for it.

- **Humanlikeness and anthropomorphism**

  The findings presented in Chapter 6 put together with the results of the previous study question the hypothesis on multi-dimensionality of anthropomrophism. A robot that was perceived as more intentional was not anthropomorphized stronger than a robot perceived as less intentional. This finding, similarly with the previously reported lack of impact of perceived intelligence suggests that the HU dimension of humanness is less important for a robot's perceived humanlikeness and anthropomorphism. As a result, this work also questions whether dehumanization and anthropomorphism are reverse phenomena.

  Furthermore, this study also showed that humanlike appearance of a robot is not linearly related with anthropomorphization of it in HRI as was previously suggested by research using images of robots. In the study presented in that chapter, a robot perceived as more humanlike due to its appearance was attributed less HN traits than a less humanlike robot. Therefore, an experiment involving different interaction contexts could provide additional information on this relationship.

- **Sociability as a factor of anthropomorphism**

  Chapter 7 presents the last experimental study that indicates that sociability, similarly to emotionality, is another factor of HN dimension that affects a robot's anthropomorphizability. It highlights that this dimension of humanness is more important for anthropomorphism in HRI. The second finding of this experiment is in line with the previous study and further indicates the lack of linear relationship between humanlike appearance and attribution of HN traits to a robot. Therefore, it can be concluded that perceived humanlikeness and psychological anthropomorphism are distinct phenomena.

Although the empirical work did not fully support the proposed model of anthropomorphism in its original form, the findings should remain of interest for different groups of scientists. Engineers are recommended to focus on implementing HN, especially primary emotions and sociability, capabilities rather than HU in order to create anthropomorphic robots. Furthermore, highly humanlike appearance of robots does not necessarily lead to their higher anthropomorphization if the expectations regarding the robots capabilities are not met. Social scientists are provided with empirical evidence that anthropomorphism and dehumanization are not reverse phenomena, although they hold some related features. Cognitive scientists could be interested in the findings regarding the processing involved in anthropomorphism. In particular, the reported involvement of Type II processing rather than Type I processing in anthropomorphism will require further investigation.

8.3 Limitations

The limitations of single studies were discussed in the chapters that introduced them. In this section, general limitations of the work presented in this thesis are discussed. An important aspect of science is reproducibility. Some of the findings found during this research were shown in several studies. The effect of HN and the lack of it of HU on anthropomorphism in HRI has been tested in more than one experiment, using different factors and interactions contexts. On the other hand, some other findings were shown only in a single study. The inversion effect of robots or the implicit measure of anthropomorphism were investigated in single experiments. It is crucial that the suggested outcomes are independently verified by other researchers before firm conclusions can be drawn.

Another limitation of the presented work is relatively small sample sizes used in the experiments. Although ideally bigger samples are preferable, researchers must strike a balance between including large samples and what is practical from the recruitment point of view. Using p values as the sole indicator of whether there is an effect of an independent variable on a dependent variable is influenced by the sample size. In this thesis, effect sizes were provided to provide a richer picture of the results. However, it remains possible that if bigger samples were used, some of the reported statistically non-significant results would be significant. Power analysis could be used to guide the decision process regarding sample sizes in the studies.

Furthermore, it is important to note that the presented research was conducted in two countries: New Zealand and Japan. The cultures in these countries differ. None of the studies was aimed at comparing multicultural differences on the proposed model of anthropomorphism between both cultures. It is possible that the results are specific for each country and cannot be generalized across the cultures. However, at least one study shows that measuring anthropomorphism using HN and HU traits is a valid measure for HRI in Japanese context [122].

8.4 Future work

The results presented in this dissertation indicate that the approach used in this work can provide valuable insight on the evaluated phenomena and it is worth continuing in the future. Especially our understating of the relationship and processes involved in psychological anthropomorphism and perceived humanlikeness is limited. Future research should involve a larger range of implicit measures, such as mouse tracker, and physiological measures, e.g. amylaze, to further explore whether anthropomorphism is a result of only Type II processing.

Furthermore, it is not known yet why perception of physical humanlikeness of a robot is linearly related with anthropomorphism while using images of robots, however once there is actual interaction with a robot involved, it is the less humanlike looking robot that is being anthropomorphized more than highly humanlike robot. It is possible that physical appearance of a robot creates a potential for it to be more anthropomorphic. It is sufficient to see an image of a robot for it to be activated. However, when in an actual interaction a robot is not capable of fulfilling these expectations, there is a repulsive reaction that leads to stronger denial of humanness to such a robot.

The previous studies investigated mainly whether anthropomorphism is a result of Type I or Type II processing. It would be valuable to investigate the same aspect for the perception of physical humanlikeness of a robot. This would also help to understand better whether the proposed inversion effect measurement is inferior to existing measurement tools or it measures different aspects of perceived humanlikeness. If humanlikeness involves Type I processing, future work should put more effort in developing implicit measures of the phenomenon.

Another interesting area worth exploration is investigating the unique consequences of anthropomorphism and humanlikeness. Since these two phenomena differ from each other, it is also possible that they have unique consequences for HRI. Therefore, distinguishing between them and better understanding their impact could help to design robots and interactions that could benefit from advantages of these phenomena, without inheriting their drawbacks.

Finally, in spite of not finding the two dimensions of humanness hypothesized in this dissertation to exist for anthropomorphism, it is not possible to exclude a possibility that there exist some other dimensions of anthropomorphism not considered in this work. Since each of these dimensions could have a unique impact on HRI it is worth further investigating the topic of dimensionality of anthropomorphism in HRI.

References

[1] J. Aaker. Dimensions of brand personality. *Journal of Marketing research*, pages 347–356, 1997.

[2] L. Abend. In spain, human rights for apes. *TIME.com. www.time.com/time/world/article/0*, 8599(1824206), 2008.

[3] H. Admoni and B. Scassellati. A multi-category theory of intention. In *Proceedings of COGSCI 2012*, pages 1266–1271, Sapporo, Japan, 2012.

[4] R. Adolphs. Could a robot have emotions? theoretical perspectives from social cognitive neuroscience. In M. Arbib and J.-M. Fellous, editors, *Who Needs Emotions: The Brain meets the Robot*, pages 9–28. Oxford University Press, 2005.

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

[6] G. W. Allport. The nature of prejudice. *Reading: Addison-Wesley*, 1954.

[7] M. A. Arbib and J.-M. Fellous. Emotions: from brain to robot. *Trends in cognitive sciences*, 8(12):554–561, 2004.

[8] A. Austermann, S. Yamada, K. Funakoshi, and M. Nakano. How do users interact with a pet-robot and a humanoid. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 3727 – 3732, Atlanta, GA, United states, 2010.

[9] H. Aviezer, Y. Trope, and A. Todorov. Body Cues, Not Facial Expressions, Discriminate Between Intense Positive and Negative Emotions. *Science*, 338(6111):1225–1229, Nov. 2012.

[10] R. Baddoura, G. Venture, and R. Matsukata. The familiar as a key-concept in regulating the social and affective dimensions of HRI. In *IEEE-RAS International Conference on Humanoid Robots*, pages 234 – 241, Osaka, Japan, 2012.

[11] J. Bae and M. Kim. Selective visual attention occurred in change detection derived by animacy of robot's appearance. In *Proceedings of the 2011 International Conference on Collaboration Technologies and Systems, CTS 2011*, pages 190–193, 2011.

[12] R. Bakeman. Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37(3):379–384, 2005.

[13] J. Barrett. *Why Would Anyone Believe in God?* Lanham, MD: AltaMira Press, 2004.

[14] J. Barrett. Cognitive science of religion: What is it and why is it? *Religion Compass*, 1(6):768–786, 2007.

[15] J. L. Barrett. Exploring the natural foundations of religion. *Trends in cognitive sciences*, 4(1):29–34, 2000.

[16] C. Bartneck. Who like androids more: Japanese or US americans? In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*, pages 553 – 557, Munich, Germany, 2008.

[17] C. Bartneck. Robots in the theatre and the media. In *Design & Semantics of Form & Movement (DeSForM2013)*, pages 64–70. Philips, 2013.

[18] C. Bartneck, T. Bleeker, J. Bun, P. Fens, and L. Riet. The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. *Paladyn*, pages 1–7, 2010.

[19] C. Bartneck and J. Hue. Exploring the abuse of robots. *Interaction Studies*, 9(3):415–433, 2008.

[20] C. Bartneck, T. Kanda, H. Ishiguro, and N. Hagita. Is the uncanny valley an uncanny cliff? In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pages 368 – 373, Jeju, Republic of Korea, 2007.

[21] C. Bartneck, T. Kanda, H. Ishiguro, and N. Hagita. My robotic doppelganger - a critical look at the uncanny valley theory. In *18th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN2009*, pages 269–276. IEEE, 2009.

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

[23] C. Bartneck, J. Reichenbach, and J. Carpenter. Use of praise and punishment in human-robot collaborative teams. In *RO-MAN 2006: The 15th IEEE International Symposium on Robot and Human Interactive Communication (IEEE Cat No. 06TH8907)*, Piscataway, NJ, USA, 2006.

[24] C. Bartneck, C. Rosalia, R. Menges, and I. Deckers. Robot Abuse-limitation of the media equation. In *Proceedings of the Interact 2005 Workshop on Agent Abuse, Rome*, 2005.

[25] C. Bartneck, M. Verbunt, O. Mubin, and A. Al Mahmud. To kill a mockingbird robot. In *HRI 2007 - Proceedings of the 2007 ACM/IEEE Conference on Human-Robot Interaction - Robot as Team Member*, pages 81 – 87, Arlington, VA, United States, 2007.

[26] B. Bastian and N. Haslam. Excluded from humanity: The dehumanizing effects of social ostracism. *Journal of Experimental Social Psychology*, 46(1):107–113, Jan. 2010.

[27] C. Becker-Asano, K. Ogawa, S. Nishio, and H. Ishiguro. Exploring the uncanny valley with geminoid HI-1 in a real-world application. In *Proc. of the IADIS Int. Conf. Interfaces and Human Computer Interaction 2010, IHCI, Proc. of the IADIS Int. Conf. Game and Entertainment Technologies 2010, Part of the MCCSIS 2010*, pages 121 – 128, Freiburg, Germany, 2010.

[28] J. H. Berg and K. Piner. *Personal relationships and social support*, chapter Social relationships and the lack of social relationship, pages 104–221. London: Sage, 1990.

[29] J. Bering. Origins of the social mind: Evolutionary psychology and child development. pages 411–437. New York: The Guildford Press, 2005.

[30] J. Bering. *The God Instinct.* London: Nicholas Brealey, 2010.

[31] J. M. Bering. The folk psychology of souls. *Behav. Brain Sci.*, 29(05):453–462, 2006.

[32] P. Bernard, S. J. Gervais, J. Allen, S. Campomizzi, and O. Klein. Integrating sexual objectification with object versus person recognition: The sexualized-body-inversion hypothesis. *Psychological Science*, 23(5):469–471, 2012.

[33] C. L. Bethel, K. Salomon, and R. R. Murphy. Preliminary results: Humans find emotive non-anthropomorphic robots more calming. In *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction, HRI'09*, pages 291–292, 2009.

[34] P. Bloom. *Descartes' Baby: How the Science of Child Development Explains what Makes Us Human.* Basic Books, 2005.

[35] P. Boyer. *Religion Explained.* New York: Basic Books, 2001.

[36] P. Boyer. Religious thought and behaviour as by-products of brain function. *Trends in cognitive sciences*, 7(3):119–124, 2003.

[37] C. Breazeal. Early experiments using motivations to regulate human-robot interaction. In *AAAI Fall Symposium on Emotional and Intelligent: The tangled knot of cognition, Technical Report FS-98-03*, pages 31–36, 1998.

[38] C. Breazeal. *Sociable Machines: Expressive Social Exchange Between Humans and Robots*. Dissertation submitted to the department of electrical engineering and computer science in partial fulfillment of the requirements for the degree of doctor of science at the massachusetts institute of technology, May 2000.

[39] C. Breazeal. Affective interaction between humans and robots. In J. Kelemen and P. Sosik, editors, *ECAL 2001, LNAI 2159*, pages 582—591. Springer-Verlag, Berlin, 2001.

[40] C. Breazeal. Human-robot partnership. *IEEE Intelligent Systems*, 21(4):79–81, 2006.

[41] C. Breazeal, D. Buchsbaum, J. Gray, D. Gatenby, and B. Blumberg. Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. *Artif. Life*, 11(1-2):31–62, 2005.

[42] C. Breazeal and P. Fitzpatrick. That certain look: Social amplification of animate vision. In *Proceedings of the AAAI Fall Symposium on Society of Intelligence AgentsThe Human in the Loop*, 2000.

[43] C. Breazeal and B. Scassellati. Infant-like social interactions between a robot and a human caregiver. *Adaptive Behavior*, 8(1):49–74, 2000.

[44] C. Breazeal and B. Scassellati. Challenges in building robots that imitate. In K. Dautenhahn and C. L. Nehaniv, editors, *Imitation in Animals and Artifacts*. MIT Press, Cambridge, Mass., 2001.

[45] E. Broadbent, V. Kumar, X. Li, J. Sollers, 3rd, R. Q. Stafford, B. A. MacDonald, and D. M. Wegner. Robots with display screens: A robot with a more humanlike face display is perceived to have more mind and a better personality. *PLoS ONE*, 8(8):e72589, 08 2013.

[46] D. J. Calverley. Android science and animal rights, does an analogy exist? *Connection Science*, 18(4):403–417, 2006.

[47] S. Carey. Becoming a face expert. *Philosophical transactions of the Royal Society of London.Series B: Biological sciences*, 335(1273):95–102, 1992.

[48] M. Cheetham, P. Suter, and L. Jancke. The human likeness dimension of the "Uncanny valley hypothesis": Behavioral and functional MRI findings. *Frontiers in Human Neuroscience*, 5, Nov. 2011.

[49] S. Chew, W. Tay, D. Smit, and C. Bartneck. Do social robots walk or roll? In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 6414 LNAI, pages 355 – 361, Singapore, Singapore, 2010.

[50] M. Chin, V. Sims, B. Clark, and G. Lopez. Measuring individual differences in anthropomorphism toward machines and animals. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 48, pages 1252–1255. SAGE Publications, 2004.

[51] J. Choi and M. Kim. The usage and evaluation of anthropomorphic form in robot design. In *Undisciplined! Design Research Society Conference 2008, Sheffield Hallam University, Sheffield, UK, 16-19 July 2008*, 2009.

[52] P. R. Cohen. If not turing's test, then what? *AI magazine*, 26(4):61, 2005.

[53] S. M. Collishaw and G. J. Hole. Featural and configurational processes in the recognition of faces of different familiarity. *Perception*, 29(8):893–909, 2000.

[54] M. Cooney, F. Zanlungo, S. Nishio, and H. Ishiguro. Designing a flying humanoid robot (FHR): effects of flight on interactive communication. In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pages 364 – 371, Paris, France, 2012.

[55] A. Darlow and S. Sloman. Two systems of reasoning: Architecture and relation to emotion. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(3):382–392, 2010.

[56] C. Darwin. *1998. The expression of the emotions in man and animals.* Oxford University Press, New York, 1872.

[57] N. Dasgupta. Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research*, 17(2):143–169, 2004.

[58] D. C. Dennett, F. Dretske, S. Shurville, A. Clark, I. Aleksander, and J. Cornwell. The practical requirements for making a conscious robot. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, 349(1689):133–146, 1994.

[59] R. Diamond and S. Carey. Why faces are and are not special. an effect of expertise. *Journal of Experimental Psychology: General*, 115(2):107–117, 1986.

[60] V. Dill, L. M. Flach, R. Hocevar, C. Lykawka, S. R. Musse, and M. S. Pinho. Evaluation of the uncanny valley in CG characters. In *12th International Conference on Intelligent Virtual Agents, IVA 2012, September 12, 2012 - September 14, 2012*, volume 7502 LNAI of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pages 511–513. Springer Verlag, 2012.

[61] C. DiSalvo and F. Gemperle. From seduction to fulfillment: The use of anthropomorphic form in design. In *Proceedings of the 2003 International Conference on Designing Pleasurable Products and Interfaces*, DPPI '03, pages 67–72, New York, NY, USA, 2003. ACM.

[62] C. F. DiSalvo, F. Gemperle, J. Forlizzi, and S. Kiesler. All robots are not created equal: The design and perception of humanoid robot heads. In *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS*, pages 321 – 326, London, United kingdom, 2002.

[63] J. Dovidio and S. Gaertner. Reducing prejudice: Combating intergroup biases. *Current Directions in Psychological Science*, 8(4):101–105, 1999.

[64] J. Dovidio, S. Gaertner, and T. Saguy. Commonality and the complexity of "we" social attitudes and social change. *Personality and Social Psychology Review*, 13(1):3–20, 2009.

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

[66] P. Ekman. Are there basic emotions? *Psychological review*, 99(3):550–553, 1992.

[67] N. Epley, S. Akalis, A. Waytz, and J. Cacioppo. Creating social connection through inferential reproduction: Loneliness and perceived agency in gadgets, gods, and hreyhounds: Research article. *Psychological Science*, 19(2):114–120, 2008.

[68] N. Epley, A. Waytz, S. Akalis, and J. T. Cacioppo. When we need a human: Motivational determinants of anthropomorphism. *Social Cognition*, 26(2):143–155, Apr. 2008.

[69] N. Epley, A. Waytz, and J. T. Cacioppo. On seeing human: A three-factor theory of anthropomorphism. *Psychol. Rev.*, 114(4):864–886, 2007.

[70] S. Epstein, R. Pacini, V. Denes-Raj, and H. Heier. Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of Personality and Social Psychology*, 71(2):390–405, 1996.

[71] J. Evans. *Dual-processing accounts of reasoning, judgment, and social cognition*, volume 59 of *Annual Review of Psychology.* 2008.

[72] J. Evans and K. Stanovich. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3):223–241, 2013.

[73] V. Evers, H. C. Maldonado, T. L. Brodecki, and P. J. Hinds. Relational vs. group self-construal: Untangling the role of national culture in HRI. In *HRI 2008 - Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction: Living with Robots*, pages 255 – 262, Amsterdam, Netherlands, 2008.

[74] F. Eyssel, F. Hegel, G. Horstmann, and C. Wagner. Anthropomorphic inferences from emotional nonverbal cues: A case study. In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pages 646 – 651, Viareggio, Italy, 2010.

[75] F. Eyssel and D. Kuchenbrandt. Manipulating anthropomorphic inferences about NAO: the role of situational and dispositional aspects of effectance motivation. In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pages 467–472, 2011.

[76] F. Eyssel and D. Kuchenbrandt. Social categorization of social robots: Anthropomorphism as a function of robot group membership. *Br. J. Soc. Psychol.*, 51(4):724–731, 2012.

[77] F. Eyssel, D. Kuchenbrandt, and S. Bobinger. Effects of anticipated human-robot interaction and predictability of robot behavior on perceptions of anthropomorphism. In *HRI 2011 - Proceedings of the 6th*

ACM/IEEE International Conference on Human-Robot Interaction, pages 61 – 67, Lausanne, Switzerland, 2011.

[78] F. Eyssel, D. Kuchenbrandt, S. Bobinger, L. De Ruiter, and F. Hegel. "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 *HRI'12 - Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 125–126, 2012.

[79] F. Eyssel and S. Loughnan. "It don't matter if you're black or white"? effects of robot appearance and user prejudice on evaluations of a newly developed robot companion. In *5th International Conference on Social Robotics, ICSR 2013, October 27, 2013 - October 29, 2013*, volume 8239 LNAI of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pages 422–431. Springer Verlag, 2013.

[80] J. Fasola and M. J. Matarić. Using socially assistive human-robot interaction to motivate physical exercise for older adults. In *Proceedings of the IEEE*, volume 100, pages 2512 – 2526, Piscataway, NJ, United States, 2012.

[81] D. Feil-Seifer and M. J. Matarić. Automated detection and classification of positive vs. negative robot interactions with children with autism using distance-based features. In *HRI 2011 - Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction*, pages 323–330, 2011.

[82] K. Fischer. Interpersonal variation in understanding robots as social actors. In *Proceedings of the 6th International Conference on Human-robot Interaction*, pages 53–60, New York, NY, USA, 2011. ACM.

[83] K. Fischer, K. S. Lohan, and K. Foth. Levels of embodiment: Linguistic analyses of factors influencing HRI. In *HRI'12 - Proceedings of*

the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction, pages 463–470, 2012.

[84] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. In *IROS 2002, September 30, 2002 - September 30, 2002*, volume 42 of *Robotics and Autonomous Systems*, pages 143–166. Elsevier, 2003.

[85] K. Ford and P. Hayes. On computational wings: Rethinking the goals of artificial intelligence-the gold standard of traditional artificial intelligence-passing the so-called turing test and thereby appearing to be. *Scientific American Presents*, 9(4):79, 1998.

[86] F. Forster, A. Weiss, and M. Tscheligi. Anthropomorphic design for an interactive urban robot - the right design approach? In *HRI 2011 - Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction*, pages 137 – 138, Lausanne, Switzerland, 2011.

[87] B. Friedman, P. H. Kahn,Jr., and J. Hagman. Hardware companions?: what online AIBO discussion forums reveal about the human-robotic relationship. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 273–280, New York, NY, USA, 2003. ACM.

[88] S. R. Fussell, S. Kiesler, L. D. Setlock, and V. Yew. How people anthropomorphize robots. In *HRI 2008 - Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction: Living with Robots*, pages 145 – 152, Amsterdam, Netherlands, 2008.

[89] R. Gaunt, J. Leyens, and S. Demoulin. Intergroup relations and the attribution of emotions: Control over memory for secondary emotions associated with the ingroup and outgroup. *J. Exp. Soc. Psychol.*, 38(5):508–514, 2002.

[90] S. J. Gervais, T. K. Vescio, J. Frster, A. Maass, and C. Suitner. Seeing

women as objects: The sexual body part recognition bias. *European Journal of Social Psychology*, 2012.

[91] N. Giullian, D. Ricks, A. Atherton, M. Colton, M. Goodrich, and B. Brinton. Detailed requirements for robots in autism therapy. In *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, pages 2595–2602, 2010.

[92] R. Gockley, J. Forlizzi, and R. Simmons. Interactions with a moody robot. In *HRI 2006: Proceedings of the 2006 ACM Conference on Human-Robot Interaction*, pages 186–193, 2006.

[93] J. Goetz, S. Kiesler, and A. Powers. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on*, pages 55 – 60, Nov. 2003.

[94] K. Gold and B. Scassellati. A bayesian robot that distinguishes "self" from "other". In *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, 2007.

[95] H. M. Gray, K. Gray, and D. M. Wegner. Dimensions of mind perception. *Science*, 315(5812):619, 2007.

[96] K. Gray and D. Wegner. Moral typecasting: Divergent perceptions of moral agents and moral patients. *J. Pers. Soc. Psychol.*, 96(3):505–520, 2009.

[97] K. Gray and D. Wegner. Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125(1):125–130, 2012.

[98] A. Greenwald and M. Banaji. Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychol. Rev.*, 102(1):4–27, 1995.

[99] S. Guthrie. *Faces in the clouds: A new theory of religion.* Oxford University Press, USA, 1995.

[100] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. d. Visser, and R. Parasuraman. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(5):517–527, Oct. 2011.

[101] R. Hard. *The Routledge Handbook of Greek Mythology: Based on HJ Rose's" Handbook of Greek Mythology"*. Psychology Press, 2004.

[102] N. Haslam. Dehumanization: An integrative review. *Personality and Social Psychology Review*, 10(3):252–264, 2006.

[103] N. Haslam and P. Bain. Humanizing the self: Moderators of the attribution of lesser humanness to others. *Personality and Social Psychology Bulletin*, 33(1):57–68, 2007.

[104] N. Haslam, P. Bain, L. Douge, M. Lee, and B. Bastian. More human than you: Attributing humanness to self and others. *J. Pers. Soc. Psychol.*, 89(6):937–950, 2005.

[105] N. Haslam, B. Bastian, S. Laham, and S. Loughnan. Humanness, dehumanization, and moral psychology. 2012.

[106] N. Haslam, Y. Kashima, S. Loughnan, J. Shi, and C. Suitner. Subhuman, inhuman, and superhuman: Contrasting humans with nonhumans in three cultures. *Social Cognition*, 26(2):248–258, Apr. 2008.

[107] N. Haslam, S. Loughnan, Y. Kashima, and P. Bain. Attributing and denying humanness to others. *European Review of Social Psychology*, 19(1):55–85, 2009.

[108] P. Hayes and K. Ford. Turing test considered harmful. In *IJCAI (1)*, pages 972–977, 1995.

[109] F. Hegel, S. Gieselmann, A. Peters, P. Holthaus, and B. Wrede. Towards a typology of meaningful signals and cues in social robotics. In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pages 72–78, 2011.

[110] F. Hegel, S. Krach, T. Kircher, B. Wrede, and G. Sagerer. Understanding social robots: A user study on anthropomorphism. In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*, pages 574–579, 2008.

[111] F. Heider and M. Simmel. An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2):243–259, 1944.

[112] M. Hewstone, M. Rubin, and H. Willis. Intergroup bias. *Annu. Rev. Psychol.*, 53:575–604, 2002.

[113] M. Hirai and K. Hiraki. Differential neural responses to humans vs. robots: An event-related potential study. *Brain Research*, 1165:105–115, Aug. 2007.

[114] C. Ho and K. MacDorman. Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices. *Computers in Human Behavior*, 26(6):1508–1518, 2010.

[115] D. W. Hogg, F. Martin, and M. Resnick. *Braitenberg creatures*. Number 13. Epistemology and Learning Group, MIT Media Laboratory, 1991.

[116] A. Hollingworth, G. Schrock, and J. M. Henderson. Change detection in the flicker paradigm: The role of fixation position within the scene. *Memory and Cognition*, 29(2):296–304, 2001.

[117] P. S. T. Inc. E-prime 2.0. Retrieved from http://www.pstnet.com, 2012.

[118] H. Ishiguro. Android science: conscious and subconscious recognition. *Connection Science*, 18(4):319–332, Dec. 2006.

[119] C. Izard and E. Bartlett. Patterns of emotions: A new analysis of anxiety and depression. 1972.

[120] P. Kahn, H. Ishiguro, B. Friedman, and T. Kanda. What is a human? - toward psychological benchmarks in the field of human-robot interaction. In *The 15th IEEE International Symposium on Robot and Human Interactive Communication, 2006. ROMAN 2006*, pages 364–371, Sept. 2006.

[121] P. H. Kahn Jr., T. Kanda, H. Ishiguro, B. T. Gill, J. H. Ruckert, S. Shen, H. E. Gary, A. L. Reichert, N. G. Freier, and R. L. Severson. Do people hold a humanoid robot morally accountable for the harm it causes? In *HRI'12 - Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 33 – 40, Boston, MA, United states, 2012.

[122] H. Kamide, F. Eyssel, and T. Arai. Psychological anthropomorphism of robots. In G. Herrmann, M. Pearson, A. Lenz, P. Bremner, A. Spiers, and U. Leonards, editors, *Social Robotics*, volume 8239 of *Lecture Notes in Computer Science*, pages 199–208. Springer International Publishing, 2013.

[123] H. Kamide, K. Kawabe, S. Shigemi, and T. Arai. Development of a psychological scale for general impressions of humanoid. *Advanced Robotics*, 27(1):3–17, 2013.

[124] T. Kanda, T. Miyashita, T. Osada, Y. Haikawa, and H. Ishiguro. Analysis of humanoid appearances in human-robot interaction. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, pages 62 – 69, Edmonton, AB, Canada, 2005.

[125] L. Kaufman and P. J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis.* Wiley-Interscience, 99 edition, Mar. 1990.

[126] T. Kemper. How many emotions are there? wedding the social and the autonomic components. *American Journal of Sociology*, pages 263–289, 1987.

[127] S. Kiesler and J. Goetz. Mental models and cooperation with robotic assistants. In *Proc. of Conference on Human Factors in Computing Systems*, pages 576–577, 2002.

[128] S. Kiesler and P. Hinds. Introduction to this special issue on human-robot interaction. *Hum.-Comput. Interact.*, 19(1):1–8, June 2004.

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

[130] S. Kriz, G. Anderson, and J. G. Trafton. Robot-directed speech: using language to assess first-time users' conceptualizations of a robot. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, HRI '10, pages 267–274, Piscataway, NJ, USA, 2010. IEEE Press.

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

[132] D. Kuchenbrandt, N. Riether, and F. Eyssel. Does anthropomorphism reduce stress in HRI? In *Proceedings of the ACM/IEEE International Conference onHuman-robot Interaction*, pages 218–219, Bielefeld, Germany, 2014. ACM.

[133] M. A. Lawrence. *ez: Easy analysis and visualization of factorial experiments.*, 2013. R package version 4.2-2.

[134] H. Leder and V. Bruce. When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 53(2):513–536, 2000.

[135] S.-l. Lee, I. Lau, S. Kiesler, and C.-Y. Chiu. Human mental models of humanoid robots. In *Proceedings of the 2005 IEEE International Con-*

ference on Robotics and Automation, 2005. ICRA 2005, pages 2767–2772, Apr. 2005.

[136] S. Lemaignan, J. Fink, P. Dillenbourg, and C. Braboszcz. The cognitive correlates of anthropomorphism. In *2014 Human-Robot Interaction Conference, Workshop" HRI: a bridge between Robotics and Neuroscience"*, 2014.

[137] D. Levin, S. Killingsworth, M. Saylor, S. Gordon, and K. Kawamura. Tests of concepts about different kinds of minds: Predictions about the behavior of computers, robots, and people. *Human-Computer Interaction*, 28(2):161–191, 2013.

[138] J. Leyens, P. Paladino, R. Rodriguez-Torres, J. Vaes, S. Demoulin, A. Rodriguez-Perez, and R. Gaunt. The emotional side of prejudice: The attribution of secondary emotions to ingroups and outgroups. *Personality and Social Psychology Review*, 4(2):186–197, 2000.

[139] M. Lohse. Bridging the gap between users' expectations and system evaluations. In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pages 485–490, 2011.

[140] S. Loughnan and N. Haslam. Animals and androids: Implicit associations between social categories and nonhumans. *Psychological Science*, 18(2):116–121, 2007.

[141] S. Loughnan, N. Haslam, T. Murnane, J. Vaes, C. Reynolds, and C. Suitner. Objectification leads to depersonalization: The denial of mind and moral concern to objectified others. *European Journal of Social Psychology*, 40(5):709–717, 2010.

[142] K. F. MacDorman, R. D. Green, C.-C. Ho, and C. T. Koch. Too real for comfort? uncanny responses to computer generated faces. *Comput. Hum. Behav.*, 25(3):695–710, May 2009.

[143] D. Maurer, R. Le Grand, and C. J. Mondloch. The many faces of configural processing. *Trends in cognitive sciences*, 6(6):255–260, 2002.

[144] D. McDermott. Artificial intelligence meets natural stupidity. *ACM SIGART Bulletin*, (57):4–9, 1976.

[145] C. K. Morewedge, J. Preston, and D. M. Wegner. Timescale bias in the attribution of mind. *J. Pers. Soc. Psychol.*, 93(1):1–11, 2007.

[146] M. Mori. The uncanny valley. *Energy*, 7(4):33–35, 1970.

[147] B. Mutlu, F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita. Nonverbal leakage in robots: Communication of intentions through seemingly unintentional behavior. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, HRI '09, pages 69–76, New York, NY, USA, 2009. ACM.

[148] C. Nass and Y. Moon. Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1):81–103, 2000.

[149] C. Nass, J. Steuer, and E. R. Tauber. Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '94, pages 72–78, New York, NY, USA, 1994. ACM.

[150] J. Nunnally. *Psychometric theory.* Tata McGraw-Hill Education, 1967.

[151] J. Nunnally. Psychometric methods. *New York: McGraw*, 1978.

[152] L. Oestreicher and K. S. Eklundh. User expectations on human-robot co-operation. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 91–96. IEEE, 2006.

[153] K. Ogawa, C. Bartneck, D. Sakamoto, T. Kanda, T. Ono, and H. Ishiguro. Can an android persuade you? In *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 516 – 21, Piscataway, NJ, USA, 2009.

[154] J.-H. Oh, D. Hanson, W.-S. Kim, I. Y. Han, J.-Y. Kim, and I.-W. Park. Design of android type humanoid robot albert HUBO. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1428–1433, Oct. 2006.

[155] E. Oztop, T. Chaminade, and D. Franklin. Human-humanoid interaction: is a humanoid robot perceived as a human? In *2004 4th IEEE/RAS International Conference on Humanoid Robots*, volume 2, pages 830–841, 2004.

[156] M. P. Paladino, J. P. Leyens, R. Rodriguez, A. Rodriguez, R. Gaunt, and S. Demoulin. Differential association of uniquely and non uniquely human emotions with the ingroup and the outgroup. *Group Processes & Intergroup Relations*, 5(2):105–117, 2002.

[157] E. Paluck. Reducing intergroup prejudice and conflict using the media: A field experiment in rwanda. *J. Pers. Soc. Psychol.*, 96(3):574–587, 2009.

[158] B. Payne. Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81(2):181–192, 2001.

[159] J. W. Peirce. PsychoPy-psychophysics software in python. *Journal of neuroscience methods*, 162(1):8–13, 2007.

[160] T. Pettigrew. Intergroup contact theory. *Annu. Rev. Psychol.*, 49:65–85, 1998.

[161] T. Pettigrew and L. Tropp. A meta-analytic test of intergroup contact theory. *J. Pers. Soc. Psychol.*, 90(5):751–783, 2006.

[162] K. Pitsch and B. Koch. How infants perceive the toy robot pleo. an exploratory case study on infant-robot-interaction. In *2nd International*

Symposium on New Frontiers in Human-Robot Interaction - A Symposium at the AISB 2010 Convention, March 29, 2010 - April 1, 2010, pages 80–87. The Society for the Study of Artificial Intelligence, 2010.

[163] K. Pitsch, K. Lohan, K. Rohlfing, J. Saunders, C. Nehaniv, and B. Wrede. Better be reactive at the beginning. implications of the first seconds of an encounter for the tutoring style in human-robot-interaction. In *2012 IEEE RO-MAN*, pages 974–981, Sept. 2012.

[164] J. B. Pollack. Mindless intelligence. *IEEE Intelligent Systems*, 21(3):50–56, 2006.

[165] A. Powers, A. D. I. Kramer, S. Lim, J. Kuo, S.-L. Lee, and S. Kiesler. Eliciting information from people with a gendered humanoid robot. In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, volume 2005, pages 158 – 163, Nashville, TN, United states, 2005.

[166] D. Proudfoot. Anthropomorphism and AI: turing's much misunderstood imitation game. *Artificial Intelligence*, 175(5):950–957, 2011.

[167] D. Proudfoot. Can a robot smile? wittgenstein on facial expression. In T. P. Racine and K. L. Slaney, editors, *A Wittgensteinian Perspective on the Use of Conceptual Analysis in Psychology*, pages 172—194. Palgrave Macmillan, Basingstoke, 2013.

[168] D. Proudfoot. Rethinking turing's test. *Journal of Philosophy*, 110(7):391–411, 2013.

[169] D. Proudfoot. Turing's child-machines. In J. Bowen, J. Copeland, M. Sprevak, and R. Wilson, editors, *The Turing Guide: Life, Work, Legacy*. Oxford University Press, in press.

[170] D. Proudfoot. Turing's three senses of "Emotional". *International Journal of Synthetic Emotions, Special Issue on Turing*, 5(2), in press.

[171] I. Pyysiäinen. Religion is neither costly nor beneficial. *Behav. Brain Sci.*, 27(06):746–746, 2004.

[172] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2014.

[173] C. L. Reed, V. E. Stone, S. Bozova, and J. Tanaka. The body-inversion effect. *Psychological Science*, 14(4):302–308, 2003.

[174] C. L. Reed, V. E. Stone, J. D. Grubb, and J. E. McGoldrick. Turning configural processing upside down: Part and whole body postures. *Journal of Experimental Psychology: Human Perception and Performance*, 32(1):73–87, 2006.

[175] B. Reeves and C. Nass. *The Media Equation.* 1996.

[176] M. Rehm and A. Krogsager. Negative affect in human robot interaction - impoliteness in unexpected encounters with robots. In *2013 IEEE RO-MAN*, pages 45–50, 2013.

[177] J. Reichenbach, C. Bartneck, and J. Carpenter. Well done, robot! - the importance of praise and presence in human-robot collaboration. In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pages 86 – 90, Hatfield, United Kingdom, 2006.

[178] L. D. Riek, T.-C. Rabinowitch, B. Chakrabarti, and P. Robinson. How anthropomorphism affects empathy toward robots. In *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction, HRI'09*, pages 245 – 246, San Diego, CA, United States, 2008.

[179] N. Riether, F. Hegel, B. Wrede, and G. Horstmann. Social facilitation with social robots? In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, HRI '12, pages 41–48, Boston, Massachusetts, USA, 2012. ACM.

[180] A. M. Rosenthal-von der Pütten and N. C. Krämer. How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior*, 36:422–439, July 2014.

[181] M. Saerbeck, T. Schut, C. Bartneck, and M. D. Janse. Expressive robots in education: Varying the degree of social supportive behavior of a robotic tutor. In *Conference on Human Factors in Computing Systems - Proceedings*, volume 3, pages 1613–1622, 2010.

[182] M. Salem, F. Eyssel, K. Rohlfing, S. Kopp, and F. Joublin. *Effects of gesture on the perception of psychological anthropomorphism: A case study with a humanoid robot*, volume 7072 LNAI of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* 2011.

[183] J. L. Saver and J. Rabin. The neural substrates of religious experience. *J. Neuropsychiatry Clin. Neurosci.*, 9(3):498–510, 1997.

[184] A. P. Saygin, T. Chaminade, H. Ishiguro, J. Driver, and C. Frith. The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, 7(4):413–422, 2012.

[185] B. Scassellati. How robotics and developmental psychology complement each other. In *NSF/DARPA Workshop on Development and Learning*, 2000.

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

[187] B. Scassellati. *How social robots will help us to diagnose, treat, and understand autism*, volume 28 of *Springer Tracts in Advanced Robotics.* 2007.

[188] B. Scassellati, C. Crick, K. Gold, E. Kim, F. Shic, and G. Sun. Social development [robots]. *Computational Intelligence Magazine, IEEE*, 1(3):41–47, 2006.

[189] M. Schmitz. Concepts for life-like interactive objects. In *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '11, pages 157–164, New York, NY, USA, 2011. ACM.

[190] F. Shic and B. Scassellati. Pitfalls in the modeling of developmental systems. *International Journal of Humanoid Robotics*, 4(2):435–454, 2007.

[191] E. Short, J. Hart, M. Vu, and B. Scassellati. No fair!! an interaction with a cheating robot. In *5th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2010*, pages 219 – 226, Osaka, Japan, 2010.

[192] V. K. Sims, M. G. Chin, H. C. Lum, L. Upham-Ellis, T. Ballion, and N. C. Lagattuta. Robots' auditory cues are subject to anthropomorphism. In *Proceedings of the Human Factors and Ergonomics Society*, volume 3, pages 1418–1421, 2009.

[193] E. Smith and J. DeCoster. Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4(2):108–131, 2000.

[194] T. Spexard, A. Haasch, J. Fritsch, and G. Sagerer. Human-like person tracking with an anthropomorphic robot. In *Proceedings - IEEE International Conference on Robotics and Automation*, volume 2006, pages 1286 – 1292, Orlando, FL, United States, 2006.

[195] L. Sussenbach, K. Pitsch, I. Berger, N. Riether, and F. Kummert. "Can you answer questions, flobi?": Interactionally defining a robot's competence as a fitness instructor. In *2012 IEEE RO-MAN*, pages 1121–1128, Sept. 2012.

[196] D. S. Syrdal, K. Dautenhahn, M. L. Walters, and K. L. Koay. Sharing spaces with robots in a home scenario - anthropomorphic attributions and their effect on proxemic expectations and evaluations in a live HRI trial. In *AAAI Fall Symposium - Technical Report*, volume FS-08-02, pages 116 – 123, Arlington, VA, United States, 2008.

[197] D. S. Syrdal, K. Dautenhahn, S. N. Woods, M. L. Walters, and K. L. Koay. Looking good? appearance preferences and robot personality inferences at zero acquaintance. In *AAAI Spring Symposium - Technical Report*, volume SS-07-07, pages 86 – 92, Stanford, CA, United States, 2007.

[198] J. Tanaka and I. Gauthier. *Expertise in Object and Face Recognition*, volume 36 of *Psychology of Learning and Motivation - Advances in Research and Theory.* 1997.

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

[200] E. Torta, E. Van Dijk, P. A. M. Ruijten, and R. H. Cuijpers. The ultimatum game as measurement tool for anthropomorphism in human-robot interaction. In *5th International Conference on Social Robotics, ICSR 2013, October 27, 2013 - October 29, 2013*, volume 8239 LNAI of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pages 209–217. Springer Verlag, 2013.

[201] M. Trimble and A. Freeman. An investigation of religiosity and the Gastaut-Geschwind syndrome in patients with temporal lobe epilepsy. *Epilepsy & behavior*, 9(3):407–414, 2006.

[202] A. Turing. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950.

[203] A. Turing. Can digital computers think? In B. J. Copeland, editor, *The Essential Turing*. Oxford: Oxford University Press, 2004.

[204] A. Turing. Intelligent machinery. In B. J. Copeland, editor, *The Essential Turing*. Oxford University Press, 2004.

[205] A. Turing. Lecture on the automatic computing engine. In B. J. Copeland, editor, *The Essential Turing*. Oxford: Oxford University Press, 2004.

[206] A. Turing, R. Braithwaite, G. Jefferson, and M. Newman. Can automatic calculating machines be said to think? In J. Copeland, editor, *The Essential Turing*. Oxford University Press, Oxford, 2004.

[207] S. Turkle. In good company? on the threshold of robotic companions. In Y. Wilks, editor, *Close Engagements with Artifical Companions: Key social, psychological, ethical and design issues*, pages 3–10. John Benjamins Publishing Company, Amsterdam/ Philadelphia, 2010.

[208] M. Verkuyten. Multicultural recognition and ethnic minority rights: A social identity perspective. *European Review of Social Psychology*, 17(1):148–184, 2006.

[209] J. von Zitzewitz, P. M. Boesch, P. Wolf, and R. Riener. Quantifying the human likeness of a humanoid robot. *International Journal of Social Robotics*, 5(2):263–276, 2013.

[210] J. Vorauer, A. Gagnon, and S. Sasaki. Salient intergroup ideology and intergroup interaction. *Psychological Science*, 20(7):838–845, 2009.

[211] E. Wade, A. R. Parnandi, and M. J. Mataric. Using socially assistive robotics to augment motor task performance in individuals post-stroke. In *IEEE International Conference on Intelligent Robots and Systems*, pages 2403–2408, 2011.

[212] M. L. Walters, K. L. Koay, D. S. Syrdal, K. Dautenhahn, and R. Te Boekhorst. Preferences and perceptions of robot appearance and embodiment in human-robot interaction trials. In *Adaptive and Emergent Behaviour and Complex Systems - Proceedings of the 23rd Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour, AISB 2009*, pages 136 – 143, Edinburgh, United Kingdom, 2009.

[213] M. L. Walters, D. S. Syrdal, K. L. Koay, K. Dautenhahn, and R. Te Boekhorst. Human approach distances to a mechanical-looking robot with different robot voice styles. In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*, pages 707–712, 2008.

[214] E. Wang, C. Lignos, A. Vatsal, and B. Scassellati. Effects of head movement on perceptions of humanold robot behavior. In *HRI 2006: Proceedings of the 2006 ACM Conference on Human-Robot Interaction*, volume 2006, pages 180–185, 2006.

[215] A. Waytz, J. Cacioppo, and N. Epley. Who sees human? the stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, 5(3):219–232, 2010.

[216] A. Waytz and N. Epley. Social connection enables dehumanization. *Journal of Experimental Social Psychology*, 48(1):70–76, Jan. 2012.

[217] A. Waytz, N. Epley, and J. T. Cacioppo. Social cognition unbound: Insights into anthropomorphism and dehumanization. *Current Directions in Psychological Science*, 19(1):58–62, 2010.

[218] A. Waytz, N. Klein, and N. Epley. Imagining other minds: Anthropomorphism is hair-triggered but not hare-brained. In M. Taylor, editor, *The Oxford Handbook of the Development of Imagination*, pages 272–287. 2013.

[219] B. Wheeler. *lmPerm: Permutation tests for linear models*, 2010. R package version 1.1-2.

[220] E. O. Wilson. *The creation: An appeal to save life on earth.* Wiley Online Library, 2006.

[221] B. Wittenbrink, C. Judd, and B. Park. Spontaneous prejudice in context: Variability in automatically activated attitudes. *J. Pers. Soc. Psychol.*, 81(5):815–827, 2001.

[222] M. Yamamoto. Sozzy: A hormone-driven autonomous vacuum cleaner. In *International Society for Optical Engineering*, volume 2058, pages 212–213, 1993.

[223] R. K. Yin. Looking at upide-down faces. *Journal of experimental psychology*, 81(1):141–145, 1969.

[224] K. Yogeeswaran and N. Dasgupta. The devil is in the details: Abstract versus concrete construals of multiculturalism differentially impact intergroup relations. *J. Pers. Soc. Psychol.*, 106(5):772–789, 2014.

[225] J. Złotowski and C. Bartneck. The inversion effect in hri: Are robots perceived more like humans or objects? In *Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction*, HRI '13, pages 365–372, Tokyo, Japan, 2013. IEEE Press.

[226] J. Złotowski, D. Proudfoot, and C. Bartneck. More human than human: Does the uncanny curve really matter? In *Proceedings of the Eight Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI'13), Workshop on Design of Humanlikeness in HRI from Uncanny Valley to Minimal Design*, pages 7 – 13, 2013.

[227] J. Złotowski, D. Proudfoot, K. Yogeeswaran, and C. Bartneck. Anthropomorphism: Opportunities and challenges in humanrobot interaction. *International Journal of Social Robotics*, 7(3):347–360, 2015.

[228] J. Złotowski, E. Strasser, and C. Bartneck. Dimensions of anthropomorphism: From humanness to humanlikeness. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, HRI '14, pages 66–73, Bielefeld, Germany, 2014. ACM.

[229] J. Złotowski, H. Sumioka, S. Nishio, C. Bartneck, F. Eyssel, and H. Ishiguro. I think, therefore you can be a human: Anthropomorphism is a conscious process. *Under review.*

[230] J. A. Złotowski, H. Sumioka, S. Nishio, D. F. Glas, C. Bartneck, and H. Ishiguro. Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception. *Frontiers in Psychology*, 6(883), 2015.

Appendix A

Godspeed Anthropomorphism Questionnaire

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

Appendix B

Individual Differences in Anthropomorphism Questionnaire

Please rate on a scale from 0 (Not at all) to 10 (Very much) to what extent do you agree with the following questions

- To what extent does technology – devices and machines for manufacturing, entertainment, and productive processes (e.g., cars, computers, television sets) – have intentions?

  Not at all 0 1 2 3 4 5 6 7 8 9 10 Very much

- To what extent does the average fish have free will?

  Not at all 0 1 2 3 4 5 6 7 8 9 10 Very much

- To what extent does the average mountain have free will?

  Not at all 0 1 2 3 4 5 6 7 8 9 10 Very much

- To what extent does a television set experience emotions?

  Not at all 0 1 2 3 4 5 6 7 8 9 10 Very much

- To what extent does the average robot have consciousness?

  Not at all 0 1 2 3 4 5 6 7 8 9 10 Very much

- To what extent do cows have intentions?

  Not at all 0 1 2 3 4 5 6 7 8 9 10 Very much

- To what extent does a car have free will?

  Not at all 0 1 2 3 4 5 6 7 8 9 10 Very much

- To what extent does the ocean have consciousness?

  Not at all 0 1 2 3 4 5 6 7 8 9 10 Very much

- To what extent does the average computer have a mind of its own?

  Not at all 0 1 2 3 4 5 6 7 8 9 10 Very much

- To what extent does a cheetah experience emotions?

  Not at all 0 1 2 3 4 5 6 7 8 9 10 Very much

- To what extent does the environment experience emotions?

  Not at all 0 1 2 3 4 5 6 7 8 9 10 Very much

- To what extent does the average insect have a mind of its own?

  Not at all 0 1 2 3 4 5 6 7 8 9 10 Very much

- To what extent does a tree have a mind of its own?

  Not at all 0 1 2 3 4 5 6 7 8 9 10 Very much

- To what extent does the wind have intentions?

  Not at all 0 1 2 3 4 5 6 7 8 9 10 Very much

- To what extent does the average reptile have consciousness?

  Not at all 0 1 2 3 4 5 6 7 8 9 10 Very much

Appendix C

Rational-Experiential Inventory

To what extent do you agree with the following statements

Need for Cognition

- I don't like to have to do a lot of thinking. (R)

  Strongly disagree 1 2 3 4 5 Strongly agree

- I try to avoid situations that require thinking in depth about something. (R)

  Strongly disagree 1 2 3 4 5 Strongly agree

- I prefer to do something that challenges my thinking abilities rather than something that requires little thought.

  Strongly disagree 1 2 3 4 5 Strongly agree

- I prefer complex to simple problems.

  Strongly disagree 1 2 3 4 5 Strongly agree

- Thinking hard and for a long time about something gives me little satisfaction.

  Strongly disagree 1 2 3 4 5 Strongly agree

Faith in Intuition

- I trust my initial feelings about people.

  Strongly disagree 1 2 3 4 5 Strongly agree

- I believe in trusting my hunches.

  Strongly disagree 1 2 3 4 5 Strongly agree

- My initial impressions of people are almost always right.

  Strongly disagree 1 2 3 4 5 Strongly agree

- When it comes to trusting people, I can usually rely on my "gut feelings."

  Strongly disagree 1 2 3 4 5 Strongly agree

- I can usually feel when a person is right or wrong even if I can't explain how I know.

  Strongly disagree 1 2 3 4 5 Strongly agree

Appendix D

Human Nature and Human Uniqueness Scale

To what extent do you agree with the following statements

Human Nature

- The robot is curious.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is friendly.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is fun-loving.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is trusting.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is aggressive.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is distractible.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is impatient.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is jealous.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is nervous.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

Human Uniqueness

- The robot is broadminded.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is humble.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is organized.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is polite.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is thorough.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is cold.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is conservative.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is hard-hearted.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is rude.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

- The robot is shallow.

  Strongly disagree 1 2 3 4 5 6 7 Strongly agree

Appendix E

Anthropomorphism Scale

Please rate your impression of the robot on these scales
Artificial 1 2 3 4 5 Natural
Synthetic 1 2 3 4 5 Real
Inanimate 1 2 3 4 5 Living
Human-made 1 2 3 4 5 Humanlike
Mechanical Movement 1 2 3 4 5 Biological Movement
Without Definite Lifespan 1 2 3 4 5 Mortal

Appendix F

Godspeed Perceived Intelligence Questionnaire

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

Appendix G

Primary Emotions

To what extent is the NAO capable of experiencing the following emotions?

- Excitement

  Not at all 1 2 3 4 5 6 7 Very much

- Joy

  Not at all 1 2 3 4 5 6 7 Very much

- Surprise

  Not at all 1 2 3 4 5 6 7 Very much

- Happiness

  Not at all 1 2 3 4 5 6 7 Very much

- Pleasure

  Not at all 1 2 3 4 5 6 7 Very much

- Anxiety

  Not at all 1 2 3 4 5 6 7 Very much

- Fear

  Not at all 1 2 3 4 5 6 7 Very much

- Pain

  Not at all 1 2 3 4 5 6 7 Very much

- Sadness

  Not at all 1 2 3 4 5 6 7 Very much

- Anger

  Not at all 1 2 3 4 5 6 7 Very much

Appendix H

Mind Attribution Scale

To what extent do you agree with the following statements?

Agency

- To what extent do you think that Robo is capable of self-control?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of morality?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of having memory?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of emotion recognition?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of planning?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of communication?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of thought?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of hunger?

  Not at all 1 2 3 4 5 Very much

Experience

- To what extent do you think that Robo is capable of hunger?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of fear?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of pain?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of pleasure?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of rage?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of desire?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of having personality?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of consciousness?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of pride?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of embarrassment?

  Not at all 1 2 3 4 5 Very much

- To what extent do you think that Robo is capable of joy?

  Not at all 1 2 3 4 5 Very much