The Roles of Robots in our Future Society

Master Thesis

Q. L. Zuurbier

Main supervisor: Dr. M.M.A. de Graaf
Secondary supervisor: Dr. P.W. Woźniak
Student number: 4151909

Department of Information and Computing Sciences
Faculty of Science

November 26, 2019

Abstract

With robots becoming popular in a variety of areas of life and an expected growth of robots in our future society, the impact of robotics on society is likely to increase (Enz, Diruf, Spielhagen, Zoll, & Vargas, 2011). To make robots successful artificial companions, the concerns and attitudes among potential human users need to be addressed (Enz et al., 2011). This is why the present study investigated the evaluation of potential future robot roles from the perspective of the general public. The research reported here is explicitly explorative and the findings should help pave the way to formulating guidelines that will inform technology design with the goal of developing acceptable robots. An online questionnaire was conducted on Amazon Turk (*n*= 510) where participants evaluated potential future robot application scenarios in order to map how humanlike people desire the future robots in our society to be in terms of appearance and cognition, and to examine people's attitudes towards potential future roles of robots. The results of this study show that people's nationality, age, gender, and level of education influence their attitude towards potential future roles of robots and how humanlike people desire future robots to be. Lastly, preliminary guidelines meant to propel the development of future robots in a desirable direction are presented.

Keywords - Human-Robot Interaction, Social robots, Design guidelines, Future robot roles, Robot appearance, Robot cognition, Robot design

TABLE OF CONTENTS

1 Introduction

1.1 Motivation

Robots are becoming popular in a variety of areas of life and are receiving more and more attention because they promise to be a superior alternative for humans in contexts with dull, dangerous, or dirty tasks (Enz et al., 2011). For example, robots will be able to replace human soldiers on the battlefield (Arkin, 2009). Moreover, robot technology is beginning to be utilized in our everyday social and private space. Examples are robots that aim to monitor and assist people who are physically or mentally impaired, such as nursebots (Morris et al., 2003), the robot seal "Paro" (Wada & Shibata, 2007; Chang, Šabanovic, & Huber, 2013), and RIBA (Mukai et al., 2010). The United Nations & The International Federation of Robotics (2006) predict that this field of service and personal robotics will continue to grow in the future, signifying the importance of the role robots will play in our future society (Royakkers & van Est, 2015).

Robots have been researched as long-term companions for human users in a variety of settings as part of an EU sponsored project named LIREC (LIving with Robots and InteractivE Companions). The project ran from 2007 until 2012 and produced several academic publications that explored how we could live with digital and interactive companions. LIREC argues that in order to make such future systems acceptable, i.e. to make robots successful artificial companions, the concerns and attitudes among potential human users need to be addressed. LIREC's argument makes sense because people's attitudes reflect their evaluations of the social world around them and influence their thoughts and behavior (Eagly & Chaiken, 2005). In 1993, Eagly and Chaiken studied the psychology behind people's attitudes and theorized that attitudes consist of an affective, a cognitive, and a behavioral component. Moreover, Ajzen (2001) and Bizer, Tormala, Rucker, & Petty (2006) describe how attitudes link affective and cognitive components with behavioral components, which allows for the prediction of future behavioral reactions towards a person, object, or event. In the context of Human-Robot Interaction (HRI), this predictive power of attitudes means that people's attitudes towards robots allow for the prediction of people's behavioral reactions towards robots. In a similar vein, Enz et al. (2011) point out that people's attitudes regarding the social roles of robots are of interest because the success of social robots in a future society depends on the complex relationship between the cognitive and affective components of people's attitudes towards these robots.

Studies have shown that a robot's appearance and its behavior shape people's attitudes towards it and determine how people perceive the robot. So did DiSalvo, Gemperle, Forlizzi, and Kiesler (2002) show that the number of facial features greatly influence how humanlike robot heads are perceived, and did Haring, Watanabe and Mougenot (2013) show that robots with ears are expected to react to sound while robots with eyes are expected to respond to visual input. It has also been shown that a feature-based approach can produce a systematic understanding of the relationship between humanlike features in anthropomorphic robots and their overall physical human-likeness (Phillips, Zhao, Ullman, & Malle, 2018). Moreover, the appearance of a robot shapes people's expectations of the robot and the more features a robot has, the more intelligent it is perceived. This plays a crucial part in HRI because a mismatch between the robot's appearance and its behavior (cognition) could result in a rejection of this new technology. This means that the appearance features and cognitive capacities of a robot should take the intended use of the robot into account to prevent such a mismatch and enable an intuitive interaction between human and robot (Haring, Silvera-Tawil, Takahashi, Watanabe & Velonaki, 2016).

However, most studies concerning the public opinion on robot applications in social spaces have focused on one application area, and thus one robot role (intended use), at the time. The few studies that have focused on the general evaluation of robot roles in society did not focus on how humanlike people desire the future robots to be, or on what features people deem necessary for a particular robot role. Hence, the primary goals of this study are to investigate how humanlike people desire the future robots in our society to be in terms of appearance and cognition, and to examine people's attitudes towards the potential future roles of robots.

1.2 The need for guidelines

On the one hand, the feasibility of international guidelines for the development and use of robots is doubted, mostly because national laws and regulations differ as well as cultural notions about the most appropriate use of robots (Guo & Zhang, 2009). An example is the difference in value that cultures place on the development of independence in infants and toddlers, which could lead to contradictory views on the use of robots as caregivers for children. On the other hand, the need for guidelines is underlined by many researchers in the field of HRI for several application areas such as service robots in the home, health care, or edutainment (Veruggio & Operto, 2006; Vargas, Ho, Lim, Enz, & Aylett, 2009). In the field of social robotics, most applications are currently being developed without accompanying ethical research (RAE, 2009). This means an increased possibility of bad design, due to incomplete or false assumptions about users and their behavior being embedded in the programming of robots. Briefly put, ethical issues should not be left for programmers to decide. Instead, researchers should study society's concerns and objections and systematically transform these into guidelines for designers and developers (RAE, 2009). Similarly, Enz et al. (2011) reach the same conclusion as they point out that designers and programmers are influenced by their implicit and explicit assumptions about the views of stakeholders, instead of explicit knowledge. This study aims to gain explicit knowledge about the views of the stakeholders, which in this case is the general public. A secondary goal of this study is to examine how to transform findings from previous studies and people's opinions about the appearance features and cognitive capacities of robots into design guidelines for future robot roles.

2 Literature review

2.1 Literature research protocol

A literature review is conducted to show what has already been researched in the field, to place it in the right context, and to expose the current gap in knowledge. The search engines Google Scholar, Springerlink, ACM, ScienceDirect, and IEEE are used to gather relevant books, journal papers, and conference papers. Table 1 shows the concepts that are used as search items in the search engines, along with the synonyms that are used interchangeably for the concepts. In addition to the literature gathered through the search engines, several papers are recommended by the thesis supervisor of this study. This collection of literature serves as the starting set of papers and is subsequently used for the snowballing procedure as described by Wohlin (2014). The snowballing procedure essentially finds additional relevant work by taking a closer look at the references of the papers that make up the starting set.

Concept 1	**Concept 2**	**Concept 3**
Human-robot interaction	Human-robot collaboration	People's attitudes
Human-computer interaction	Human-robot cooperation	People's evaluations
HRI	Collaborative human-robot interaction	Public opinion
Concept 4	**Concept 5**	**Concept 6**
Humanoid	Anthropomorphism	Robot applications
Humanlike	Anthropomorphic robot	Robot roles
Concept 7	**Concept 8**	
Design guidelines	Robot technology	
Guidelines	Robotics	

Table 1 - Concepts used during the literature search

2.2 People's attitudes towards robots

In 2012, The European Commission gauged the public opinion towards robots among EU citizens and found that people (especially those with high interest in scientific discoveries and technological developments) think of robots more as instrument-like machines than humanlike ones. Moreover, 70% of the participants reported a positive attitude towards robots. On the other hand, a majority of the participants also reported being concerned that robots will take our jobs, and agree that the development of robots requires careful management. Similarly, Enz et al. (2011) researched people's hopes and fears related to the social roles that robots could fulfill in our future society. They found that people are skeptical when it comes to handing over control to robots in areas where they will act as social role models such as caretaking, education, and public security. They also found that males are generally more positive towards social robots than females are. It should be noted that all participants were students of the University of Bamberg in Germany and this study aims to examine whether this difference between males and females is found in other cultures as well. This is why this study will examine whether or not a similar difference between males and females is found for the participants of this study.

In 2016, de Graaf and Allouch identified the potential benefits and disadvantages of future robot applications in order to map the societal impact of robots. They found that robot applications that involve decreasing human exposure to dangerous situations or that involve leisure and convenience were rated positively. They also found that robot applications that challenge social norms, such as robots as romantic partners, were rated negatively. They conclude with the findings that people positively associate a future robot society with efficiency, a decrease of casualties, and convenience, and people negatively associate a future robot society with job loss and the lack of robots' social abilities. In 2008, Takayama, Ju, & Nass took a closer look at people's attitudes towards robot workers by identifying what characteristics of occupations people believe robots are qualified for. They conclude that people prefer robots for tasks that require memorization, keen perceptual skills, and service-orientation. Additionally, they found that people feel more positively toward robots collaborating with humans compared to robots replacing humans. This is in line with the findings of Eurobarometer (2012) and de Graaf and Allouch (2016) on people's concern about job loss due to robots replacing humans and will be taken into account while composing guidelines for the future robot roles in our society.

2.3 Interacting with robots

HRI is the multidisciplinary study of human-robot interaction and pursues to close the gap between human-human interaction and human-robot interaction. Within the field of HRI, a separate line of research has investigated how people evaluate several characteristics of robots.

2.3.1 Aspects of a robot's appearance that influence human-robot interaction

A relatively large portion of studies within the field of HRI focus on the characteristics of appearance. Research has shown that facial features (DiSalvo et al., 2002), gaze, height, gender, voice (Eyssel, De Ruiter, Kuchenbrandt, Bobinger, & Hegel, 2012), and proximity to human partners (Dragan, Bauman, Forlizzi, & Srinivasa, 2015) all play a role in how humans respond to robots. Additionally, Haring et al. (2016) showed that the way people perceive robots differs significantly based on appearance alone because the appearance of a robot influences how it is initially percepted and shapes people's expectations towards the robot. Likewise, Haring et al. (2016) suggest that while designing the appearance and behavior of a robot, the intended use of the robot should be taken into account in order to enable intuitive human-robot interaction. Phillips et al. (2018) support this suggestion by postulating that, because many of the people interacting with robots will be novice users with few prior experiences and no training, their initial impressions are likely to be intuitively formed based on the robot's appearance. This is why the aspects of a robot's appearance will be taken into account while composing guidelines for the future robot roles in our society. What aspects of a robot's appearance play a role in human-robot interaction are discussed below.

To what extent a robot appears as a human being plays a role in how humans respond to the robot. In 2009, Riek, Rabinowitch, Chakrabarti, & Robinson showed that people empathize more strongly with humanoid robots and less with mechanical-looking robots. This finding is supported by Krach et al. (2008), who showed that, as the degree of anthropomorphization increases, brain areas related to how we process other minds become more active. Anthropomorphism is defined as "the tendency to imbue the real or imagined behavior of nonhuman agents with humanlike characteristics, motivations, intentions, or emotions" (Epley, Waytz, & Cacioppo, 2007, *p.864*). According to Epley et al. (2007), humanizing nonhumans represents a psychological strategy to satisfy an individual's need for controlling his or her environment. To clarify, robots (especially unpredictable ones) represent a source of uncertainty to an individual's environment, thus causing stress. In an attempt to reduce the experienced uncertainty and make sense of a robot's behavior, people attribute humanity to robots. Because people who draw inferences reduce their uncertainty, the process of anthropomorphization contributes to a more pleasant and efficient HRI. This makes it relevant to take this phenomenon into account while composing guidelines for the future robot roles in our society. The findings of Riek et al. (2009) and Krach et al. (2008) are in line with the Simulation Theory which states that people mentally 'simulate' the situation of other agents (i.e. putting ourselves in other's shoes) in order to understand their mental and emotional state, and that the more similar the other agent is, the stronger the empathy is (Goldman, 2006).

Robots have also started to appear in societal domains where moral decisions are taken, from care for the elderly to education and security. Malle, Scheutz, Forlizzi, & Voiklis (2016) examined what people expect and demand of robots that make moral decisions. They found that the appearance of a robot affects people's moral judgments about the robot. They demonstrated that people expect robots with a mechanical appearance to make different moral decisions compared to humanoid robots. Making it relevant to take these findings into account while composing guidelines for future robot roles where moral decisions need to be made.

Humans are efficient when interacting with another person, using as few words as we need to communicate our meaning (Clark & Brennan, 1991). Thus, Powers et al. (2005) examined how this principle can be utilized in human-robot interaction. They showed that simple changes in a robot's appearance (male or female voice / with or without lipstick), causes the amount of information elicited from users to change. These findings suggest that if we want a robot to have a minimal and efficient conversation with users, the robot should fit the stereotype associated with the social role it fulfills. On the other hand, if we want users to provide a lot of information and to explain themselves to the robot, then the robot should violate the stereotype associated with the social role it fulfills. These findings will be taken into account while composing guidelines for future robot roles where conversations with humans are part of the human-robot interaction.

The use of robots in the workplace is likely to grow substantially, meaning that robots will increasingly work together with people, each relying on the other for parts of the tasks where the other has the better skills (Hinds, Roberts, & Jones, 2004). Hinds et al. (2004) researched the effects of a robot's appearance and relative status (subordinate / peer / supervisor) on human-robot collaboration. They found that participants felt more responsible for the successful completion of a task when working with a mechanical-looking robot as compared with a humanoid robot, especially when the mechanical-looking robot was a subordinate. They suggest that humanoid robots are appropriate for settings in which people have to delegate responsibility to the robots or when the task is too demanding for people to do. In a similar vein, Bartneck and Forlizzi (2004) suggest that a robot's interaction with people is shaped by the degree and type of autonomy of the robot. The interactions range from: people doing all the work, people teaming with robots to accomplish tasks, robots providing a simple social response to human interaction, and a fully reciprocally social robot. According to Goetz, Kiesler, & Powers (2003), a comfortable experience with appropriate feedback is required in order for robots to gain the cooperation of its users. Thus, they examined how variations in the appearance and social behavior of a robot affect people's responses to the robot. They found that people prefer robots for jobs when the robot's human likeness matches the sociability required in those jobs. And that people comply more with a robot whose appearance matches the seriousness of the task. This means that humanoid robots are not always the better choice because people expect a robot to look and act appropriately given the context of the task. These findings will be taken into account while composing guidelines for future robot roles where robots and humans will collaborate.

More contexts in which humanoid robots are not always the better choice come to mind when considering the Uncanny Valley. The Uncanny Valley theory posits that as robots become more humanlike they become more familiar (and thus more likeable), until a point is reached at which its subtle imperfections make the robot seem eerie (Mori, 1970). The Uncanny Valley influences people's implicit decisions concerning a robot's social trustworthiness because humans infer trustworthiness from subtle facial expressions (Mathur & Reichling, 2016). This phenomenon is not limited to robots but is applicable to any type of humanlike object, such as dolls, masks, and avatars in virtual reality (see Figure 1). Researchers have hypothesized that this phenomenon occurs because humanoid features remind people of death (MacDorman & Ishiguro, 2006), or because abnormal features violate evolutionary aesthetics (Seyama & Nagayama, 2007). However, Ho, MacDorman and Pramono (2008) investigated this phenomenon further and suggest that it is not a robot's overall degree of human likeness that places it in an "uncanny valley" but rather a mismatch among elements. Meaning some aspects of a robot's form, motion quality, or interactivity may seem more human than others and that it is this mismatch we find disturbing. An example of this would be a robot with humanlike eyes and teeth combined with an absence of skin and mechanical jerkiness of movement. Ho et al. (2008) suggest that designers need to consider many details concerning the appearance of a humanoid robot and especially the performance of its facial aspects as this will have a big impact on the robot's overall impression. Because this phenomenon may affect the acceptance of humanoid robots, it will be taken into account while composing guidelines for the future robot roles where robots will have a humanlike appearance.

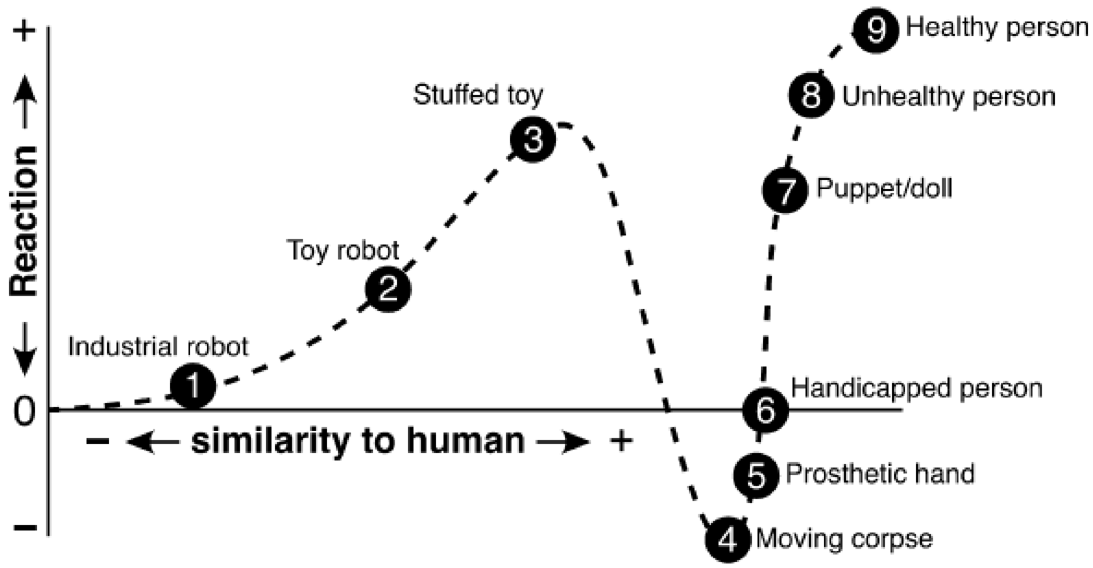


Figure 1 - Mathur and Reichling's (2016) adaptation of The Uncanny Valley.

2.3.2 Aspects of a robot's cognition that influence human-robot interaction

In order to interact with others (whether it is a device, robot, or another person) it is essential to have a good conceptual model of how the other operates. This means that, in order to optimize the interaction between humans and robots, robots should help people form a conceptual model of how it operates. Such a model makes it possible to explain and predict what the other is about to do, its reasons for doing it, and how to elicit the desired behavior from it (Norman, 2001). A robot can help a person form this model by communicating its internal state through visual cues or continual feedback (Breazeal & Brooks, 2005). In a similar vein, the "computers as social actors" approach posits that people apply the same social heuristics to computers and robots as they use for human interactions (Reeves & Nass, 1996). Moreover, Nass and Moon (2000) found that most people do so relatively automatically and without being aware of it. Research has also suggested that endowing robots with emotions results in a more natural and meaningful HRI (Breazeal, 2003). Similarly, Eyssel, Hegel, Horstmann and Wagner (2010) have demonstrated that when a robot conveys emotional reactions to a human interaction partner, the robot is perceived as more humanlike. In order to close the gap between human-human interaction and human-robot interaction, these findings will be taken into account while composing guidelines for the future roles of robots.

Whether people anticipate interacting with a robot or not and whether a robot's behavior is characterized as predictable or unpredictable impact people's view of a robot. Eyssel, Kuchenbrandt and Bobinger (2011) showed this by demonstrating that people who anticipate an encounter with an unpredictable robot anthropomorphize the robot more strongly compared to anticipating an encounter with a predictable robot. They also found that anticipating an encounter with an unpredictable robot increases people's willingness to spend their free time with the robot. In order to increase the acceptance of robots, these findings will be taken into account while composing guidelines for the future robot roles where people choose to spend their free time with a robot.

As HRI pursues to close the gap between human-human interaction and human-robot interaction, designers of robots that verbally interact with humans should also take note of the robot's vocal content and paralinguistic cues such as volume and speech rate. These aspects play important roles in human-human interaction because they express personality and emotion (Apple, Streeter, & Krauss, 1979; Pittam, 1994; Tusing & Dillard, 2000; Nass & Lee, 2001). These aspects will be taken into account while composing guidelines for the future robot roles where robots will verbally interact with humans. Lastly, how a robot should behave in order to enable pleasant and efficient HRI is dependent on the personality of the human. Tapus, Ţăpuş and Matarić (2008) suggest that extroverted individuals, who like social interactions, may prefer to have the robot physically closer than introverted individuals, who may perceive the robot as invading their space. Designers of future robots should take this matching of personalities into account if the personality of the humans that will interact with the robot is known.

2.3.3 Mind perception

Perceiving the mind of others is a crucial part of social interaction. However, people do not always ascribe minds to other people, and sometimes minds are ascribed to non-people such as God, animals, and robots. Ascribing a mind to an entity confers moral rights and makes its actions meaningful (Waytz, Gray, Epley, & Wegner, 2010), making it relevant to investigate what entities are ascribed minds and what entities are not. Moreover, DiSalvo et al. (2002) suggest that whether or not people ascribe minds to robots, and in what way, influences their decision whether to use the robot or not. They found that older (retired) people were more likely to use a robot if they held a positive attitude towards robots and perceived robot minds to have little agency. Similarly, Stafford, MacDonald, Jayawardena, Wegner and Broadbent (2014) found that participants who attributed robots high agency ratings were less likely to use the robots. According to Stafford et al. (2014), a possible explanation for this is that if a robot is perceived to have an 'unbalanced mind' i.e. the capacity for agency but not for empathy, it is perceived to be missing the checks and balances that promote both predictable and desirable behavior. This emotional response to an entity which misses the checks and balances that promote predictable and desirable behavior is similar to the ‘Uncanny Valley’ theory discussed earlier. One of the goals of this study is to examine how to transform findings from previous studies and people’s opinions about the appearance features and cognitive capacities of robots into design guidelines for future robot roles. Because whether robots are ascribed a mind or not may affect the acceptance of several future robot roles, it will be taken into account while composing these design guidelines.

By conducting online surveys, Gray, Gray and Wegner (2007) compared how people view the mental capacities of various human and nonhuman entities. They found that people intuitively think about other minds in terms of two distinct dimensions: experience (the capacity to sense and feel) and agency (the capacity to plan and act). Subsequently, the entities were compared with each other on each dimension. In the context of the present study, interesting findings include that the sociable robot Kismet scored higher on agency than a chimpanzee or a baby did, but Kismet scored lower on agency than a 5-year-old girl (see Figure 2). This means participants thought Kismet is more capable of planning and acting than a chimpanzee or baby, but less capable of planning and acting than a 5-year-old girl. On the other hand, Kismet scored lower than 11 of the 12 other entities on experience, only God scored equally low on experience as Kismet did. This means that participants thought Kismet is less capable of feeling sensations such as hunger, fear, and pain than a 7-week-old fetus or a man in a persistent vegetative state are. According to the participants of Gray et al. (2007), even a dead woman is more capable of feeling these sensations than Kismet. The finding that participants thought robots have a higher capacity for agency than capacity for experience is in line with the findings of Stafford et al. (2014). Because of the way people view the mental capacities of robots may affect the acceptance of several future robot roles, these findings will be taken into account while composing design guidelines.

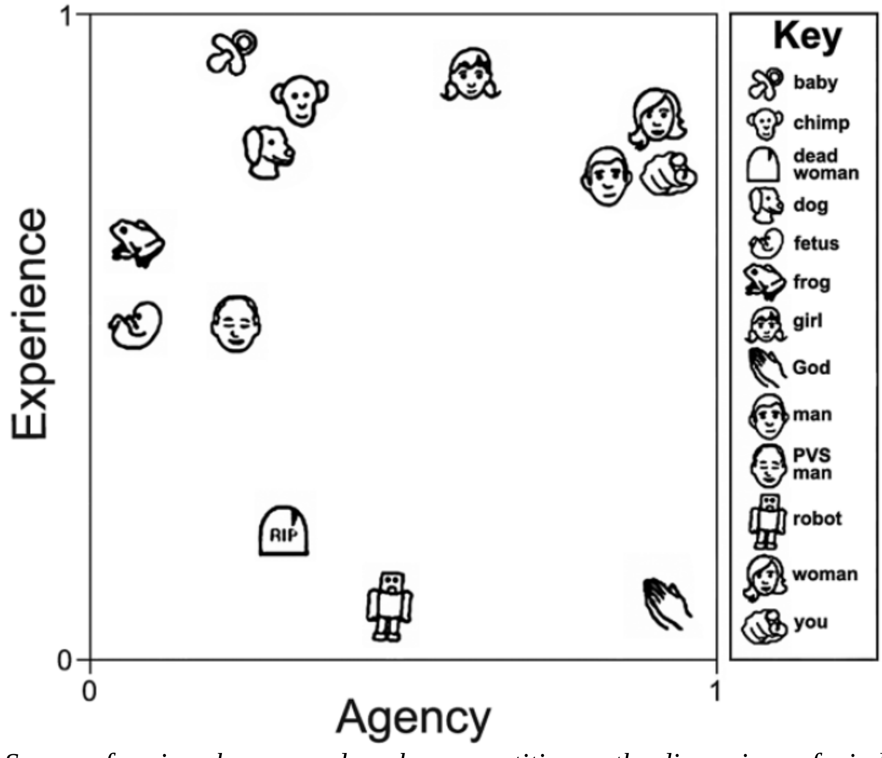


Figure 2 - Scores of various human and nonhuman entities on the dimensions of mind perception (Gray et al., 2007).

2.3.4 Utilizing existing HRI knowledge in the present study

With the goal of closing the gap between human-human interaction and human-robot interaction, the present study utilizes the knowledge gathered by the above mentioned studies by selecting which findings are included in this study's design guidelines and which findings are not. The above mentioned studies show how people evaluate the characteristics of robots, this study aims to transform this knowledge into design guidelines for future robots roles. Because knowledge of existing design guidelines is needed in order to be able to compose such design guidelines, the following section will discuss present-day as well as older design guidelines for robots.

2.4 Existing guidelines

Within the field of HRI, there is a lack of design guidelines for one or more (future) robot roles. Beer et al. (2012) identified preliminary design recommendations for home robots that support older adults to remain in their own homes as they age, and these recommendations could serve as a springboard for design guidelines, but other than that, design guidelines that focus on one or more (future) robot roles do not exist. This does not mean that it is a subject that is not being studied, but rather that past and current research has viewed the development of robots from a different perspective, such as philosophical or ethical.

In 1942, science fiction author Isaac Asimov was first to devise a set of rules concerning the development of robots (Asimov, 1968). His Three Laws of Robotics (often known as Asimov's Laws) are intended as a safety feature for how robots should operate. The laws read as follows:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey orders given to it by human beings, except where such orders would conflict with the first law.
3. A robot must protect its own existence as long as such protection does not conflict with the first or second law.

Asimov later extended his set with a 'Zeroth law', which is aimed at humanity as a whole and implies that a robot is ultimately under control of a human user even though it might follow orders and make decisions autonomously. Asimov's Laws have been used as a reference point for discussions within the philosophical and AI community regarding the ethical considerations of robots in society (Anderson, 2008; Sloman, 2006). Even medical doctors have discussed the future potential of robotic surgery in the context of the three laws (Moran, 2008). Although several of these discussions criticize the vagueness of the laws or argue that machines aren't capable of making those decisions for themselves, the laws may still prove useful as a first step towards the development of ethical guidelines for social robots.

Inspired by Asimov's Laws, Murphy & Woods proposed an alternative set of laws in 2009 based on what humans and robots can realistically accomplish in the foreseeable future (Murphy & Woods, 2009). The laws read as follows:

1. A human may not deploy a robot without the human-robot work system meeting the highest legal and professional standards of safety and ethics.
2. A robot must respond to humans as appropriate for their roles.
3. A robot must be endowed with sufficient situated autonomy to protect its own existence as long as such protection provides smooth transfer of control to other agents consistent the first and second laws.

Murphy & Woods argue that Asimov's laws assume functional morality, meaning robots will have sufficient agency and cognition to make moral decisions. Although this is not an incorrect assumption in itself, it does ignore 'operational morality', which represents the legal and professional responsibility of those who design and deploy robots. Similarly, Woods and Hollnagel (2006) point out that no matter how far the autonomy of robots will advance, the important challenges of accountability and liability will remain. By formulating these alternative laws, Murphy & Woods emphasize that the responsibility for the consequences of robots' successes and failures lies in the groups of human stakeholders.

The most recent attempt of composing guidelines that will govern the process of developing artificial intelligence originates from the European Union. On the 9th of April 2019, the European Union published a set of guidelines on how companies and governments should develop ethical applications of artificial intelligence. The guidelines were formulated by the High-Level Expert Group on Artificial Intelligence (AI HLEG) and have the objective to support the implementation of the European strategy on artificial intelligence. According to the guidelines, trustworthy AI should be:

1. Lawful - respecting all applicable laws and regulations.
2. Ethical - respecting ethical principles and values.
3. Robust - both from a technical perspective while taking into account its social environment.

Although these guidelines may prove useful as a step towards the development of ethical guidelines for social robots, they focus on a higher level of development and rather on the development of artificial intelligence than on the development of social robots. Because of this difference, the guidelines will not be taken into account while composing guidelines for the future robot roles in our society.

2.5 Research question and subquestions

Based on the above literature review and the limitations of previous studies, the main objective of this study was to investigate how humanlike people desire the future robots in our society to be by conducting a literature review and an online questionnaire. The main research question therefore was:
How humanlike do people desire the future roles of robots in our society to be?

To answer this research question, the following sub-questions were defined:
SQ1: How humanlike do people desire the future roles of robots in our society to be in terms of cognition?
This sub-question aimed to map how humanlike people desire a future robot role or cluster of future robot roles to be in terms of cognition using the capacities presented in '3.1.1 Cognitive capacities'.

SQ2: How humanlike do people desire the future roles of robots in our society to be in terms of appearance?
This sub-question aimed to map how humanlike people desire a future robot role or cluster of future robot roles to be in terms of appearance using the features presented in '3.1.2 Appearance features'.

SQ3: What are people's attitudes towards the potential future roles of robots in our society?
This sub-question examined how realistic and how positive or negative people view a future robot role or cluster of future robot roles using the questions presented in '3.1.3 Attitude'.

SQ4: What guidelines can be composed for the design of future robots?
This sub-question examined how to transform findings from previous studies and people's opinions about the appearance features and cognitive capacities of robots into guidelines that inform the design direction for the future robot roles in society.

3 Research method

De Graaf and Allouch (2016) evaluated potential future applications of robots in society by presenting participants with future robot application scenarios. Each scenario described a role that may be fulfilled by robots in our future society. The idea behind providing role descriptions was to limit the various ideas of future robots that people have developed based on science fiction literature and cinema (Ray et al., 2008). The present study built on the study conducted by de Graaf and Allouch (2016) and used similar role descriptions. The role descriptions were written in common language and without the use of terminology. This research method was chosen because earlier research on future robot applications has successfully applied similar research methods. Moreover, the results of studies that applied similar research methods are well suited for comparison, which in turn increases the validity of the results of this study. Although similar to the study conducted by de Graaf and Allouch (2016), this study viewed people's evaluation of future robot applications from a different perspective. Where de Graaf and Allouch (2016) investigated people's anticipated positive or negative consequences of several future robot applications, this study investigated how humanlike people desire the future robots in our society to be in terms of appearance and cognition, and examined people's attitudes towards the potential future roles of robots.

Data for this investigation were collected through an anonymous online questionnaire that was conducted in July and August 2019. It should be noted that this methodology relied on the imagination of people, which is common practice in marketing research as long as the technology is not yet in place (Vriens, Oppewal, & Wedel, 1998). The rationale behind providing written descriptions instead of movies or pictures was that participants should focus on the robot's role instead of on the robot's embodied appearance. This methodology has been successfully applied to future robot application scenarios before, e.g. Enz et al. (2011) and de Graaf and Allouch (2016), and the present study applied it as well. The descriptions of the future robot roles consisted of three parts. First, the role of the robot was given. For example "robots can serve as household help". Secondly, three capacities of that robot role were given. For example "robots are available day and night (1), follow our orders without conflict (2), and are easy to adjust to our personal preferences (3)". Lastly, three tasks of that robot role were given. For example "they will clean the house (1), do our groceries (2), and maintain our garden (3)". The descriptions of all the robot roles are presented in Appendix A.

3.1 Materials and measurement

Inspired by earlier research on future robot applications, such as Enz et al. (2011), Ray et al. (2008) and Arras and Cerqui (2005), 15 future robot roles were formulated by de Graaf and Allouch (2016). The present study extended this list of roles by adding a 'security guard' role (Wu, Gong, Chen, Zhi, & Xu, 2009). This means that the questionnaire aimed to map what features and capacities were deemed necessary for, and what people's attitudes were towards, the following 16 future robot roles: Housekeeper, Companion, Citizen, Entertainer, Factory worker, Caregiver, Personal assistant, Personal trainer, Nanny, Romantic partner, Security guard, Sexual partner, Soldier, Teacher, Tour guide, and Driver. Participants were asked to consider this set of 16 future robot roles. To decrease the burden on the participants, each participant was asked to consider five randomly assigned future robot roles. Similar to the rating scale used by Malle and Thapa Magar (2017), participants used a 7-point Likert scale to answer the questions.

3.1.1 Cognitive capacities

Informed by previous work on people's beliefs about robot capacities (Gray et al., 2007; Sytsma, 2004), Malle and Thapa Magar (2017) and Malle (2019) have taken the first step in developing a measure that can be used to assess people's desires for future robots' mental capacities. Malle (2019) took the item pool of 18 mental capacities of Gray, Gray and Wegner (2007) as a starting point and classified the items into five groups: Negative Affect, Positive Affect, Social Cognition, Morality, and Reality Interaction. Each group consisted of 7, 8, or 9 cognitive capacities. Drawing on Malle's (2019) classification, the questionnaire of this study distinguished between 5 cognitive domains: Physiological, Affective, Cognitive, Moral, and Agentic / Perceptual. To indicate whether participants deemed a particular cognitive domain necessary for a particular robot role, they answered the following question for each robot role:
"*Which of the following cognitive capacities would you like to see in a *future robot role*?*"
For each cognitive domain, participants used a 7-point Likert scale ranging from -3 (Not at all) to +3 (Very much) to indicate how much they would like the future robot role to possess the cognitive capacities belonging to the cognitive domain:

- *"Feeling pleasure, feeling happy, experiencing gratitude, and loving specific people or things."* (Affective)
- *"Feeling pain, feeling stressed, feeling tired, and experiencing fear."* (Physiological)
- *"Telling right from wrong, disapproving immoral actions, upholding moral values, and praising moral actions."* (Moral)
- *"Inferring a person's' thinking, understanding others' minds, planning for the future, setting goals."* (Cognitive)
- *"Communicating verbally, seeing and hearing the world, learning from instruction, and moving on their own."* (Agentic / Perceptual)

A screenshot of how this question was displayed to the participants is presented in Appendix B.

3.1.2 Appearance features

Inspired by Ezer (2008), a collection of possible robot appearance features was developed by Phillips et al. (2018). By eliminating 10 rare, redundant, or too equivocal features, the final collection of Philips et al. (2018) consisted of 19 possible robot appearance features. By conducting a Principal Component Analysis (PCA), Philips et al. (2018) uncovered four distinct appearance dimensions: Surface look, Body-manipulators, Facial features, and Mechanical locomotion. The questionnaire of the present study used these four appearance dimensions. To indicate whether participants deemed a particular appearance dimension necessary for a particular robot role, they answered the following question for each robot role:
*"Which of the following appearance features would you like to see in a *future robot role*?"*
For each appearance dimension participants used a 7-point Likert scale ranging from -3 (Not at all) to +3 (Very much) to indicate how much they would like the future robot role to possess the appearance features belonging to the appearance dimension:

- *"Head hair, nose, eyebrows, eyelashes, gender, and clothes."* (Surface look)
- *"Arms, legs, hands, and fingers."* (Body-manipulators)
- *"Head, face, eyes, and mouth."* (Facial features)
- *"Wheels or tracks."* (Mechanical locomotion)

A screenshot of how this question was displayed to the participants is presented in Appendix C.

3.1.3 Attitude

Taking into account that the attitudes among potential human users need to be addressed in order to make robots successful artificial companions, participants' attitudes towards a particular robot role were assessed by asking them the following questions for each robot role:

*"How realistic do you consider the *future robot role* to be fulfilled by robots in the near future?"*
Participants used a 7-point Likert scale ranging from 1 (Completely unrealistic) to 7 (Completely realistic) to indicate how realistic they viewed the future robot role.

*"How do you view the *future robot role* role for robots?"*
Participants used a 7-point Likert scale ranging from 1 (Completely negative) to 7 (Completely positive) to indicate how negative or positive they viewed the future robot role.

A screenshot of how these questions were displayed to the participants is presented in Appendix D.

3.2 Procedure

Studies such as Dautenhahn et al. (2005) and Ray, Mondada and Siegwart (2008) presented their participants with robot technology in an experimental or trade fair setting. These types of settings may have led to answers that were influenced by social desirability. In order to minimize the influence of social desirability on the participants, the present study conducted an anonymous online questionnaire in order to map what appearance features and cognitive capacities were deemed necessary for, and what people's attitudes were towards, the future robot roles. Robot roles for which similar features and capacities were deemed necessary formed a cluster of robot roles. Likewise, robot roles that were rated similarly in terms of how participants thought of them (unrealistic vs. realistic and negative vs. positive) formed a cluster of robot roles.

The questionnaire consisted of four parts. In the first part, the participants were presented with an informed consent form which explained how the participants could receive their compensation, what their rights as a participant were, and how they could contact the principal investigator of this study if they had any questions or remarks. After indicating that they wanted to participate in the study, the participants continued to the second part of the questionnaire. In the second part, the design and the goal of the questionnaire were explained, and it was made clear that there were no right or wrong answers. The third part asked the participants to provide demographic details, namely age, nationality, gender, and level of education. The fourth and last part of the questionnaire consisted of the questions that were aimed at mapping what appearance features and cognitive capacities the participants deemed necessary for, and what people's attitudes were towards, the future robot roles.

3.3 Data analysis

The data collection method was quantitative and the collected data was used for statistical analysis. As explained in 3.1.1, participants rated five cognitive capacities for each future robot role in terms of how much they would like that future robot role to possess the cognitive capacity. These five ratings made up the cognitive subscale and were aggregated into a 'Cognition score' per future robot role. This score represented how humanlike the participants wanted a particular future robot role to be in terms of cognition. In order to determine whether it was justifiable to interpret the aggregated 'Cognition score', a reliability analysis was conducted. The cognitive subscale was found to have acceptable reliability (5 items; $\alpha = .783$).

As explained in 3.1.2, participants rated four appearance features for each future robot role in terms of how much they would like that future robot role to possess the appearance feature. The inverse scores were used from the scores concerning Mechanical locomotion (wheels or tracks) because a high Mechanical locomotion score indicated low human likeness instead of high human likeness. These four ratings made up the appearance subscale and were aggregated into an 'Appearance score' per future robot role. This score represented how humanlike the participants wanted a particular future robot role to be in terms of appearance. In order to determine whether it was justifiable to interpret the aggregated 'Appearance score', a reliability analysis was conducted. The appearance subscale was found to have acceptable reliability (4 items; $\alpha = .731$).

3.4 Participants

3.4.1 Participant recruitment

Participants for the present study were recruited through a combination of the Amazon Mechanical Turk platform (MTurk) as well as a 'convenience approach'. MTurk provided a way to overcome the barrier of participant recruitment costs and difficulties by providing easy and inexpensive access to non-student adult participants. The convenience approach entailed a non-probability sampling technique where participants were selected because of their convenient accessibility and proximity to the researcher. For the present study, this involved contacting students from the University of Utrecht. Participants who participated through the MTurk platform received a compensation of $0.10 if their response was approved. In order to increase the proportion of reliable responses, the MTurk participants had to meet two requirements in order for them to be qualified and able to participate in the study. The first requirement was that they had to have a minimum of 1000 approved Human Intelligent Tasks (HITs). An approved HIT is a task that they completed on the MTurk platform and was subsequently approved by the person who requested the task to be completed. The second requirement was that they had to have a HIT Approval Rate of 98% or higher. The HIT Approval Rate represented the proportion of completed tasks that were approved by the persons who requested the tasks to be completed.

The reliability of responses from MTurk participants has been examined by several studies. Berinsky, Huber and Lenz (2012) pointed out that MTurk participants are notably younger and more ideologically liberal than the general public. Nevertheless, Buhrmester, Kwang and Gosling (2011) showed that data obtained through MTurk were at least as reliable as data obtained with traditional methods. Moreover, Paolacci, Chandler and Ipeirotis (2010) suggested MTurk is a viable alternative for data collection because participants exhibit the classic heuristics and biases and pay attention to directions at least as much as participants from traditional sources. Lastly, to check if the demographic information about participants on MTurk is trustworthy, Rand (2012) used IP address logging to verify the participant's self-reported country of residence. He found that 97% of the responses were accurate. He also compared the consistency of a range of demographic variables reported by the same participants across two different studies and found between 81% and 98% agreement, depending on the variable.

3.4.2 Participant demographics

On average, participants completed the survey in 7 minutes and 21 seconds. Responses of participants who completed the survey in less than 3 minutes were considered unreliable and were not approved. Similarly, responses of participants that contained untruthful, and thus unreliable answers, were not approved. Examples of this are participants who indicated that they were born before 1900. 743 participants completed the questionnaire, of which 724 through the MTurk platform. 232 participant responses were considered unreliable, meaning 510 (68.6%) responses were considered reliable. The unreliable responses were discarded and were not taken into account while analyzing the data. 153 (29.5%) of the participants did not share their demographic details. The demographic characteristics of the 357 participants that did share their demographic details are presented below.

Two-thirds of the participants had a North American nationality and almost one-third of the participants originated from Asia, of which 95 from India (See Table 2). This aligns with previous research into the demographics of MTurk participants (Hara et al., 2019; Difallah et al., 2018). It should be noted that from this point onwards, this study categorizes people from India as Asian people. Almost half of the European participants had the Dutch nationality. The nationalities of participants were grouped by continent. The included continents were Asia, Europe, and North America, because very few participants originated from Africa, Australia, and South America, these continents were grouped together as 'Other'.

Nationality	Number of participants
North America	235 (65.8%)
Asia	102 (28.6%)
Europe	15 (4.2%)
Other	5 (1.4%)

Table 2 - Nationality of participants per continent

Most participants were females (See Table 3), this is explained by the fact that North America has more female MTurk participants than male MTurk participants (Hara et al., 2019). On the other hand, India has almost twice as many male MTurk participants as female MTurk participants (Difallah, Filatova, & Ipeirotis, 2018). However, because the number of North American participants in this study was more than twice the number of Indian participants, the dominance of female participants was not surprising and is in line with previous research into MTurk participants.

Gender	Number of participants
Female	207 (58%)
Male	150 (42%)

Table 3 - Gender of participants

The majority of participants had a high level of education (See Table 4), the questionnaire distinguished between 6 levels of education. A 'Low' level of education referred to participants who did not complete high school. A 'Medium' level of education referred to participants who completed high school but did not hold a college degree. A 'High' level of education referred to participants who either held a bachelor's, master's, or doctoral degree.

Level of education	Number of participants
High	327 (91.6%)
Medium	29 (8.1%)
Low	1 (0.3%)

Table 4 - Education level of participants

70.3% of the participants were under the age of 45 (See Table 5), with an average of 39 years. This is in line with previous research into MTurk participants (Difallah et al., 2018).

Age	Number of participants
18-29	111 (31.1%)
30-44	140 (39.2%)
45-60	67 (18.8%)
60+	39 (10.9%)

Table 5 - Age of participants

4 Results

4.1 Differences between age groups

4.1.1 Differences in Cognition & Appearance

Cognition

A one-way between subjects ANOVA was conducted to compare the effect of the participant's age on how future robot roles were rated in terms of desired cognition. The age groups were 18-29, 30-44, 45-60, and 60+. There was a significant effect of the participant's age group on the 'Cognition score' at the $p<.05$ level for the four conditions [$F(3, 1781) = 53.65$, $p < 0.001$]. Post hoc comparisons using the Tukey HSD test indicated that the mean score for the age group 18-29 ($M = 1.42$, $SD = 1.15$) was significantly different ($p < 0.001$) than the three other age groups. The Tukey HSD test also indicated that the mean score for the age group 30-44 ($M = 0.74$, $SD = 1.37$) was significantly different ($p < 0.001$) than the age groups 18-29 and 45-60 ($M = 0.34$, $SD = 1.55$). Interestingly, the age group 60+ ($M = 0.53$, $SD = 1.51$) only differed significantly ($p < 0.001$) from the age group 18-29. This shows that the age of the participants influences how humanlike the participants want a particular future robot role to be in terms of cognition. Particularly, this shows that younger participants want future robots to be cognitively more humanlike compared to older participants (see Figure 3).

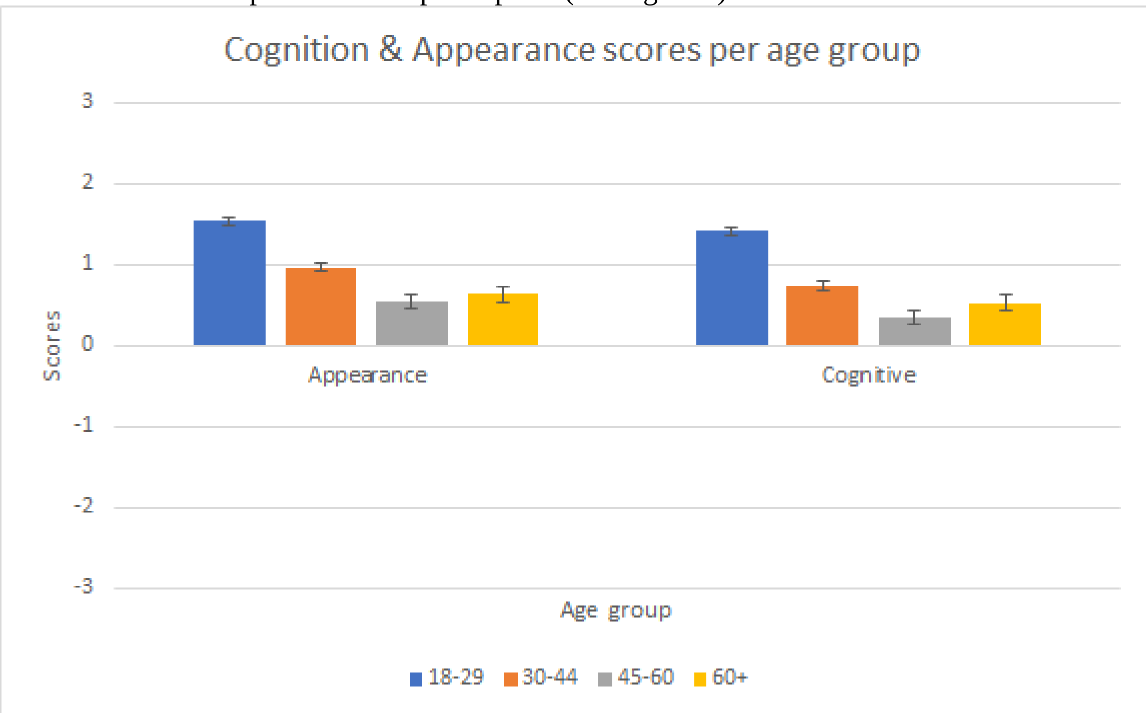


Figure 3 - Cognition and Appearance scores per age group

Appearance

Another one-way between subjects ANOVA was conducted to compare the effect of the participant's age on how future robot roles were rated in terms of desired appearance. There was a significant effect of the participant's age group on the 'Appearance score' at the $p<.05$ level for the four conditions [$F(3, 1781) = 42.16$, $p < 0.001$]. Post hoc comparisons using the Tukey HSD test indicated that the mean score for the age group 18-29 ($M = 1.54$, $SD = 1.20$) was significantly different ($p < 0.001$) than the three other age groups. The Tukey HSD test also indicated that the mean score for the age group 30-44 ($M = 0.97$, $SD = 1.46$) was significantly different ($p < 0.001$) than the age groups 18-29 and 45-60 ($M = 0.56$, $SD = 1.56$). Interestingly, the age group 60+ ($M = 0.63$, $SD = 1.51$) only differed significantly ($p < 0.001$) from the age group 18-29. This shows that the age of the participants influences how humanlike the participants want a particular future robot role to be in terms of appearance. Particularly, this shows that younger participants want future robots to appear more humanlike compared to older participants (see Figure 3).

4.1.2 Differences in Realistic & Opinion

Realistic

A one-way between subjects ANOVA was conducted to compare the effect of the participant's age on how realistic they considered the future robot roles. The age groups were 18-29, 30-44, 45-60, and 60+. There was a significant effect of the participant's age group on the 'Realistic score' at the $p<.05$ level for the four conditions [$F(3, 1781) = 54.58$, $p < 0.001$]. Post hoc comparisons using the Tukey HSD test indicated that the mean score for the age group 18-29 ($M = 5.67$, $SD = 1.55$) was significantly different ($p < 0.001$) than the three other age groups. The other age groups did not differ significantly from each other. This shows that the age of the participants influences how realistic the participants view a particular future robot role to be. Particularly, this shows that younger participants view future robots to be more realistic compared to older participants (see Figure 4).

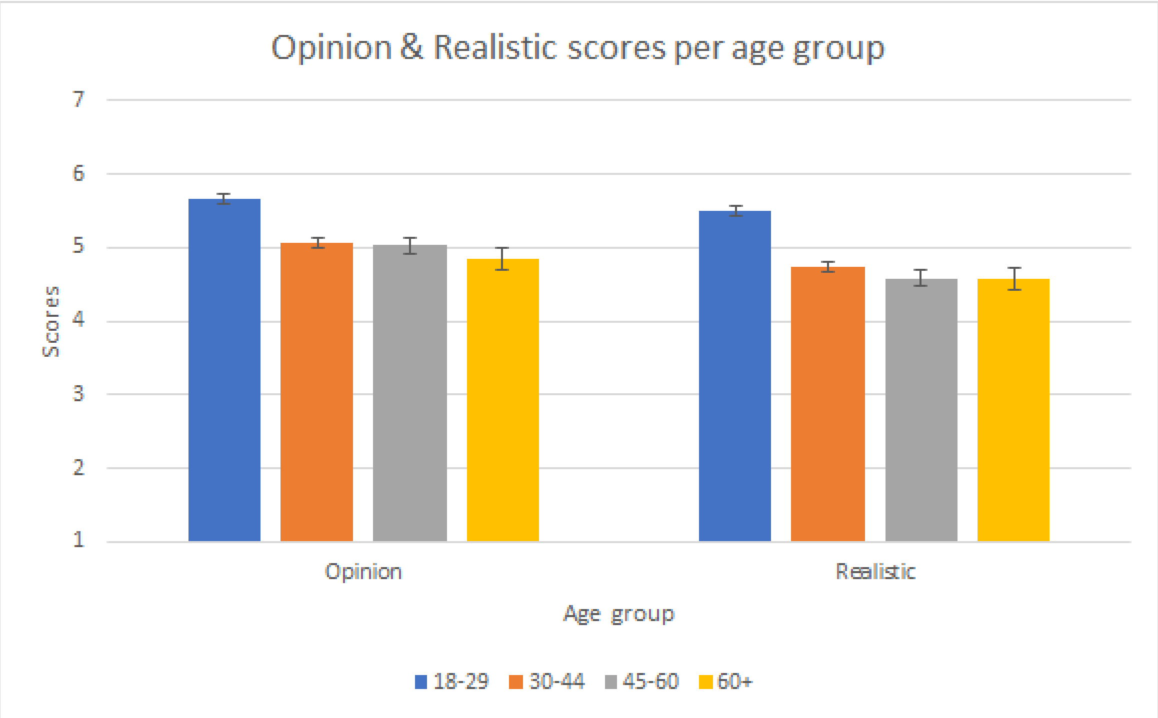


Figure 4 - Opinion and Realistic scores per age group

Opinion
Another one-way between subjects ANOVA was conducted to compare the effect of the participant's age on how positive or negative they viewed the future robot roles. There was a significant effect of the participant's age group on the 'Opinion score' at the $p<.05$ level for the four conditions [$F(3, 1781) = 24.57$, $p < 0.001$]. Post hoc comparisons using the Tukey HSD test indicated that the mean score for the age group 18-29 ($M = 5.49$, $SD = 1.62$) was significantly different ($p < 0.001$) than the three other age groups. The other age groups did not differ significantly from each other. This shows that the age of the participants influences how positive or negative the participants view a particular future robot role to be. Particularly, this shows that younger participants view future robots more positively compared to older participants (see Figure 4).

4.2 Differences between genders

4.2.1 Differences in Cognition & Appearance

Cognition
A one-way between subjects ANOVA was conducted to compare the effect of the participant's gender on how future robot roles were rated in terms of desired cognition. There was not a significant effect of the participant's gender on the 'Cognition score' at the $p<.05$ level for the two conditions [$F(1, 1783) = 0.36$, $p = 0.547$]. This shows that the gender of the participants does not significantly influence how humanlike the participants want a particular future robot role to be in terms of cognition (see Figure 5).

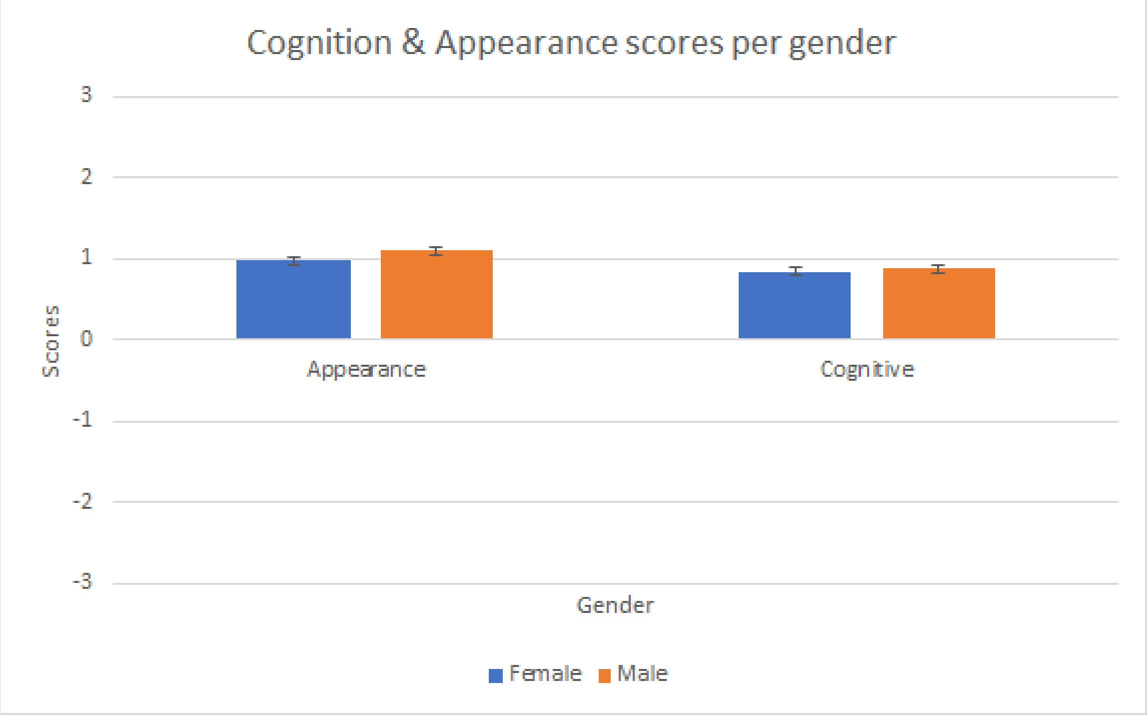


Figure 5 – Cognition and Appearance scores per gender

Appearance
Another one-way between subjects ANOVA was conducted to compare the effect of the participant's gender on how future robot roles were rated in terms of desired appearance. There was not a significant effect of the participant's gender on the 'Appearance score' at the $p<.05$ level for the two conditions [$F(1, 1783) = 3.13$, $p = 0.077$]. This shows that the gender of the participants does not significantly influence how humanlike the participants want a particular future robot role to be in terms of appearance (see Figure 5).

4.2.2 Differences in Realistic & Opinion

Realistic

A one-way between subjects ANOVA was conducted to compare the effect of the participant's gender on how realistic they considered the future robot roles. There was a significant effect of the participant's gender on the 'Realistic score' at the $p<.05$ level for the two conditions [$F(1, 1783) = 18.18$, $p < 0.001$]. This shows that the gender of the participants influences how realistic the participants view a particular future robot role to be. Particularly, this shows that male participants view future robots to be more realistic compared to female participants (see Figure 6).

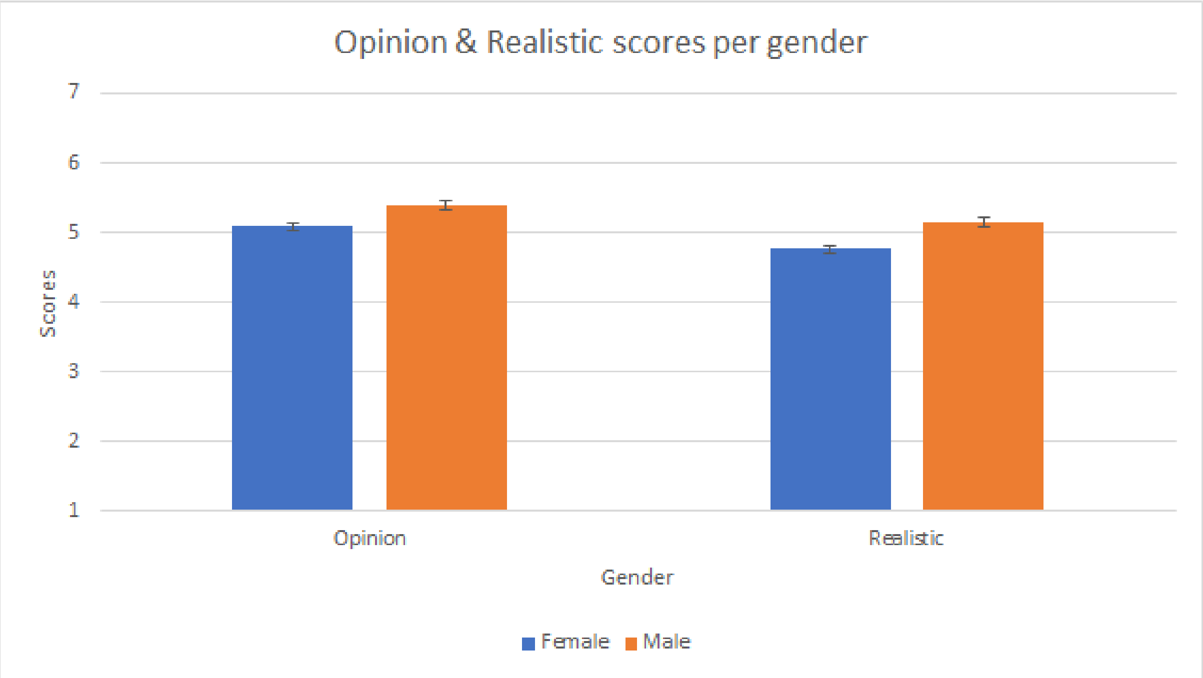


Figure 6 - Opinion and Realistic scores per gender

Opinion

Another one-way between subjects ANOVA was conducted to compare the effect of the participant's gender on how positive or negative they viewed the future robot roles. There was a significant effect of the participant's gender on the 'Opinion score' at the $p<.05$ level for the two conditions [$F(1, 1783) = 11.80$, $p = 0.001$]. This shows that the gender of the participants influences how positive or negative the participants view a particular future robot role to be. Particularly, this shows that male participants view future robots more positively compared to female participants (see Figure 6).

4.3 Differences between education level groups

4.3.1 Differences in Cognition & Appearance

Cognition

A one-way between subjects ANOVA was conducted to compare the effect of the participant's education level on how future robot roles were rated in terms of desired cognition. The education levels that used were 'Medium' and 'High' because only one participant had a 'Low' education level. There was a significant effect of the participant's education level on the 'Cognition score' at the $p<.05$ level for the two conditions [$F(1, 1778) = 8.36$, $p = 0.004$]. This shows that the level of education of the participants influences how humanlike the participants want a particular future robot role to be in terms of cognition. Particularly, this shows that participants with a 'High' education level want future robots to be cognitively more humanlike compared to participants with a 'Medium' education level (see Figure 7).

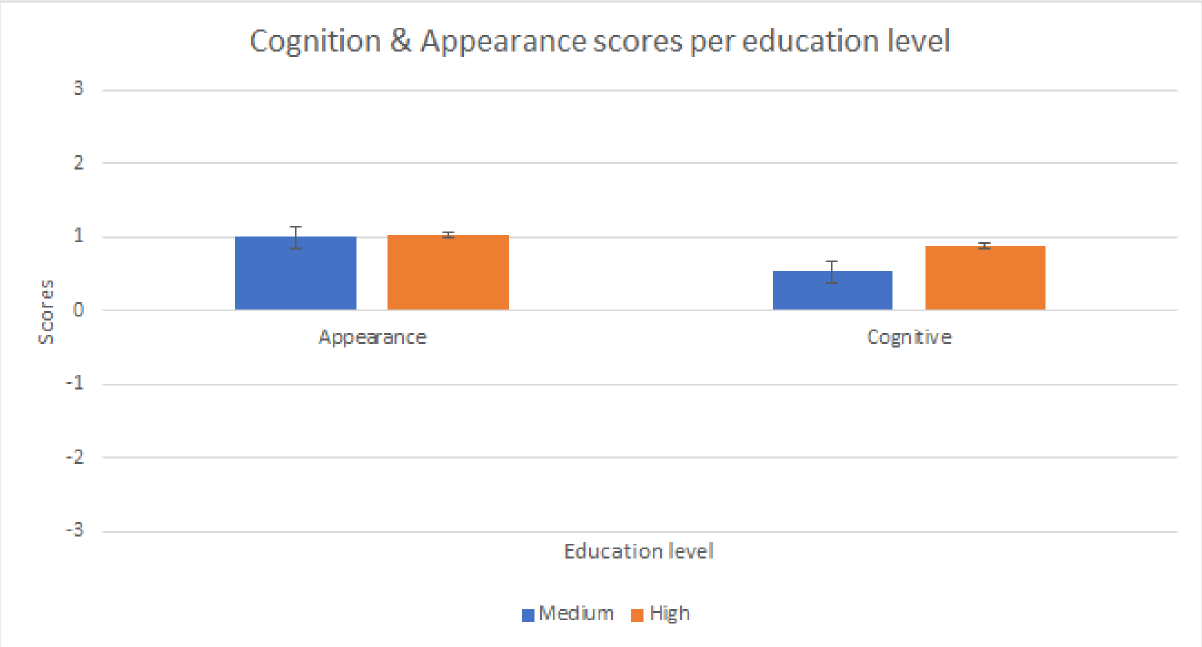


Figure 7 – Cognition and Appearance scores per education level

Appearance

Another one-way between subjects ANOVA was conducted to compare the effect of the participant's education level on how future robot roles were rated in terms of desired appearance. There was not a significant effect of the participant's education level on the 'Appearance score' at the $p<.05$ level for the two conditions [$F(1, 1778) = 0.06$, $p = 0.812$]. This shows that the level of education of the participants does not significantly influence how humanlike the participants want a particular future robot role to be in terms of appearance (see Figure 7).

4.3.2 Differences in Realistic & Opinion

Realistic

A one-way between subjects ANOVA was conducted to compare the effect of the participant's education level on how realistic they considered the future robot roles. There was not a significant effect of the participant's education level on the 'Realistic score' at the $p<.05$ level for the two conditions [$F(1, 1778) = 2.22$, $p = 0.136$]. This shows that the level of education of the participants does not significantly influence how realistic the participants view a particular future robot role to be (see Figure 8).

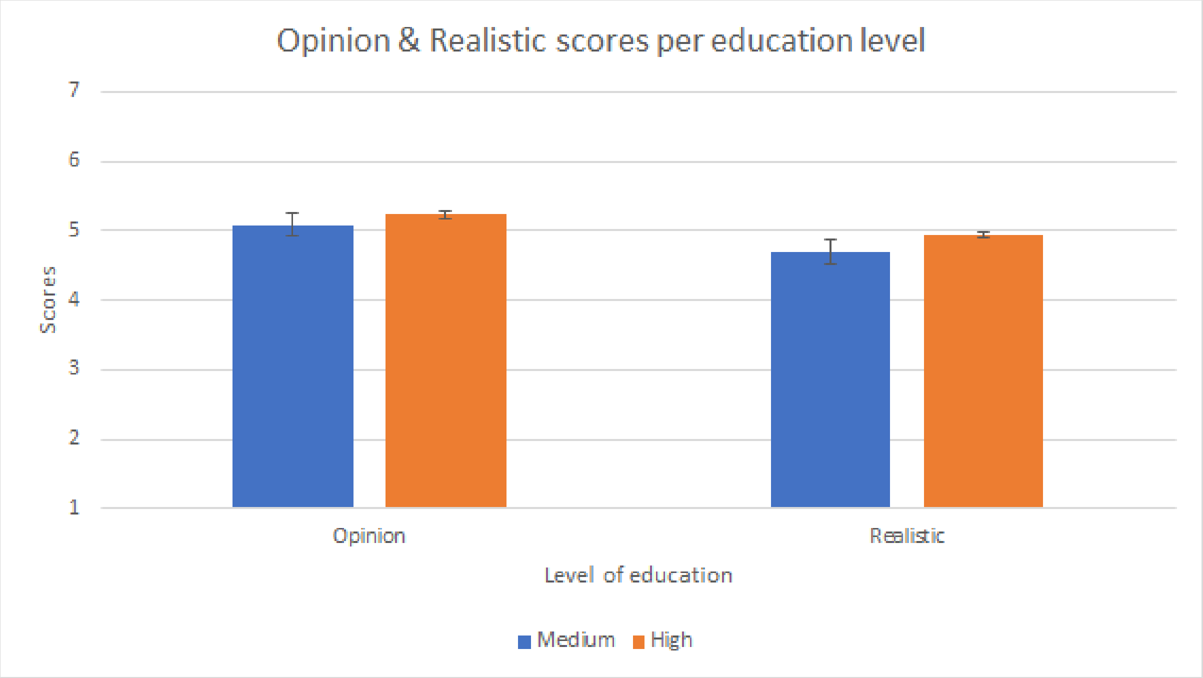


Figure 8 - Opinion and Realistic scores per education level

Opinion

Another one-way between subjects ANOVA was conducted to compare the effect of the participant's education level on how positive or negative they viewed the future robot roles. There was not a significant effect of the participant's education level on the 'Opinion score' at the $p<.05$ level for the two conditions [$F(1, 1783) = 0.78$, $p = 0.377$]. This shows that the level of education of the participants does not significantly influence how positive or negative the participants view a particular future robot role to be (see Figure 8).

4.4 Differences between continents

4.4.1 Differences in Cognition & Appearance

Cognition
A one-way between subjects ANOVA was conducted to compare the effect of the participant's nationality on how future robot roles were rated in terms of desired cognition. There was a significant effect of the participant's nationality group on the 'Cognition score' at the $p<.05$ level for the four conditions [$F(3, 1776) = 71.70$, $p < 0.001$]. Post hoc comparisons using the Tukey HSD test indicated that the mean score for North America ($M = 0.54$, $SD = 1.47$) was significantly different ($p < 0.001$) than the mean score of Asia ($M = 1.58$, $SD = 1.00$). The Tukey HSD test also indicated that the mean score for Europe ($M = 0.66$, $SD = 1.31$) was significantly different ($p < 0.001$) than the mean score of Asia. This shows that the nationality of the participants influences how humanlike the participants want a particular future robot role to be in terms of cognition. Particularly, this shows that participants from Asia want future robots to be cognitively more humanlike compared to participants from other continents (see Figure 9).

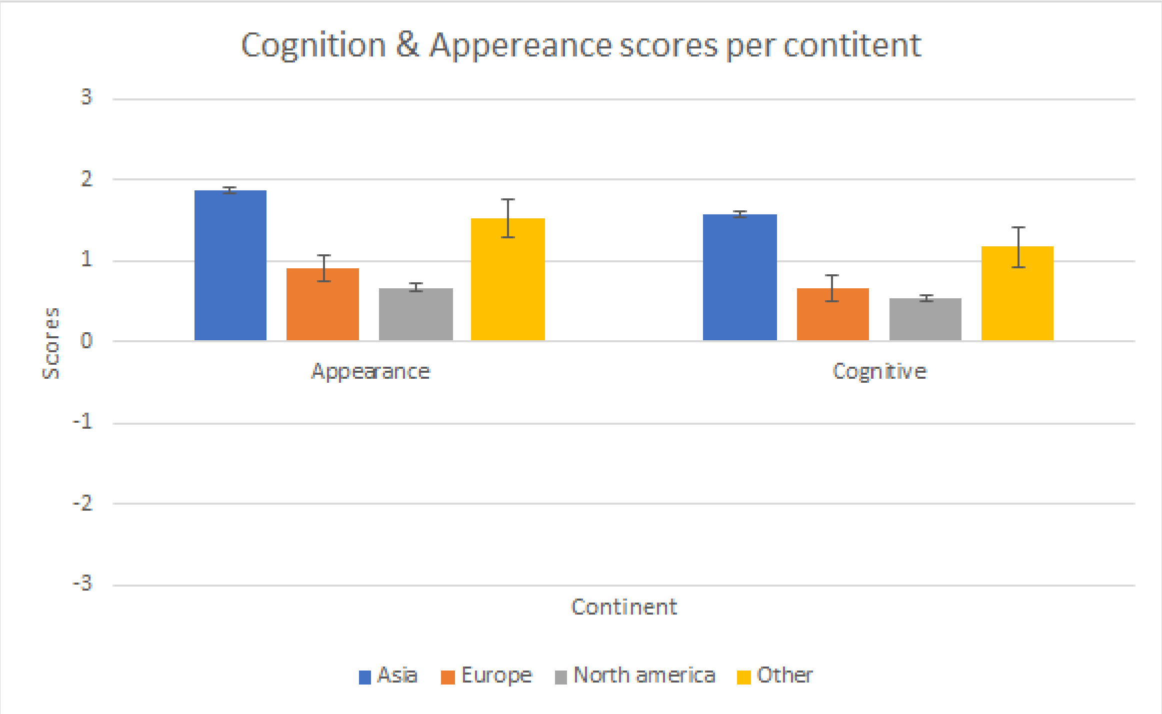


Figure 9 – Cognition and Appearance scores per continent

Appearance
Another one-way between subjects ANOVA was conducted to compare the effect of the participant's nationality on how future robot roles were rated in terms of desired appearance. There was a significant effect of the participant's nationality group on the 'Appearance score' at the $p<.05$ level for the four conditions [$F(3, 1776) = 93.64$, $p < 0.001$]. Post hoc comparisons using the Tukey HSD test indicated that the mean score for North America ($M = 0.66$, $SD = 1.50$) was significantly different ($p < 0.001$) than the mean score of Asia ($M = 1.87$, $SD = 0.96$). The Tukey HSD test also indicated that the mean score for Europe ($M = 0.91$, $SD = 1.40$) was significantly different ($p < 0.001$) than the mean score of Asia. This shows that the nationality of the participants influences how humanlike the participants want a particular future robot role to be in terms of appearance. Particularly, this shows that participants from Asia want future robots to appear more humanlike compared to participants from other continents (see Figure 9).

4.4.2 Differences in Realistic & Opinion

Realistic

A one-way between subjects ANOVA was conducted to compare the effect of the participant's nationality on how realistic they considered the future robot roles. There was a significant effect of the participant's nationality group on the 'Realistic score' at the $p<.05$ level for the four conditions [$F(3, 1776) = 55.42$, $p < 0.001$]. Post hoc comparisons using the Tukey HSD test indicated that the mean score for North America ($M = 4.54$, $SD = 2.00$) was significantly different ($p < 0.001$) than the mean scores of Asia ($M = 5.78$, $SD = 1.49$) and the Other group ($M = 6.54$, $SD = 1.17$). This shows that the nationality of the participants influences how realistic the participants view a particular future robot role to be. Particularly, this shows that participants from Asia view future robots to be more realistic compared to participants from other continents (see Figure 10).

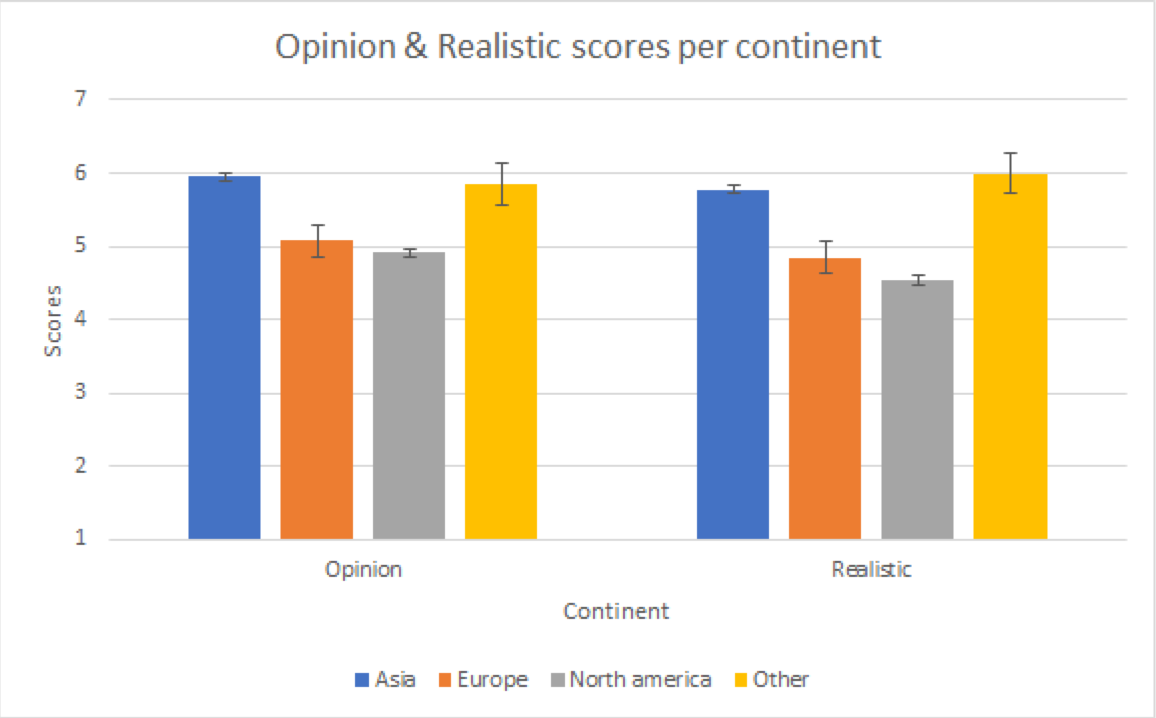


Figure 10 - Opinion and Realistic scores per continent

Opinion

Another one-way between subjects ANOVA was conducted to compare the effect of the participant's nationality on how positive or negative they viewed the future robot roles. There was a significant effect of the participant's nationality group on the 'Opinion score' at the $p<.05$ level for the four conditions [$F(3, 1776) = 39.85$, $p < 0.001$]. Post hoc comparisons using the Tukey HSD test indicated that the mean score for Asia ($M = 5.95$, $SD = 1.39$) was significantly different ($p < 0.001$) than the mean score of North America ($M = 4.91$, $SD = 1.97$) and significantly different ($p = 0.001$) than the mean score of Europe ($M = 5.08$, $SD = 1.92$). This shows that the nationality of the participants influences how positive or negative the participants view a particular future robot role to be. Particularly, this shows that participants from Asia view future robots more positively compared to participants from other continents (see Figure 10).

4.5 Differences between robot roles

4.5.1 Differences in Cognition & Appearance

Cognition
A one-way between subjects ANOVA was conducted to compare the effect of a particular role on how it was rated in terms of desired cognition. The 16 robot roles were Housekeeper, Companion, Citizen, Entertainer, Factory worker, Caregiver, Personal assistant, Personal trainer, Nanny, Romantic partner, Security guard, Sexual partner, Soldier, Teacher, Tour guide, and Driver. There was a significant effect of what particular role was rated on the 'Cognition score' at the $p<.05$ level for the 16 conditions [$F(15, 2534) = 6.49$, $p < 0.001$]. This shows that how humanlike the participants want a particular future robot role to be in terms of cognition depends on the specific future robot role.

Appearance
Another one-way between subjects ANOVA was conducted to compare the effect of a particular role on how it was rated in terms of desired appearance. There was a significant effect of what particular role was rated on the 'Appearance score' at the $p<.05$ level for the 16 conditions [$F(15, 2534) = 3.30$, $p < 0.001$]. This shows that how humanlike the participants want a particular future robot role to be in terms of appearance depends on the specific future robot role.

4.5.2 Differences in Realistic & Opinion

Realistic
A one-way between subjects ANOVA was conducted to compare the effect of what particular role was rated on how realistic they considered the future robot role. There was a significant effect of what particular role was rated on the 'Realistic score' at the $p<.05$ level for the 16 conditions [$F(15, 2534) = 23.31$, $p < 0.001$]. This shows that how realistic the participants view a particular future robot role depends on the specific future robot role.

Opinion
Another one-way between subjects ANOVA was conducted to compare the effect of what particular role was rated on how positive or negative they viewed the future robot role. There was a significant effect of what particular role was rated on the 'Opinion score' at the $p<.05$ level for the 16 conditions [$F(15, 2534) = 23.41$, $p < 0.001$]. This shows that how negative or positive the participants view a particular future robot role depends on the specific future robot role.

4.6 Clusters of robot roles

To further investigate how the significant differences between the different future robot roles emerged from the data, two separate cluster analyses were performed. For the first analysis, all future robot roles were plotted with their ‘Cognition score’ on the x-axis, and their ‘Appearance score’ on the y-axis. This produced an overview of possible clusters of future robot roles. Subsequently, a Hierarchical Cluster Analysis was conducted to confirm that certain future robot roles were statistically clustered together based on their ‘Cognition score’ and ‘Appearance score’. For the second analysis, all future robot roles were plotted with their ‘Realistic score’ on the x-axis, and their ‘Opinion score’ on the y-axis. This produced an overview of possible clusters of future robot roles. Subsequently, a Hierarchical Cluster Analysis was conducted to confirm that certain future robot roles were statistically clustered together based on their ‘Realistic score’ and ‘Opinion score’.

4.6.1 Cognition & Appearance cluster

Less clear clusters were found than initially expected (see Figure 11).

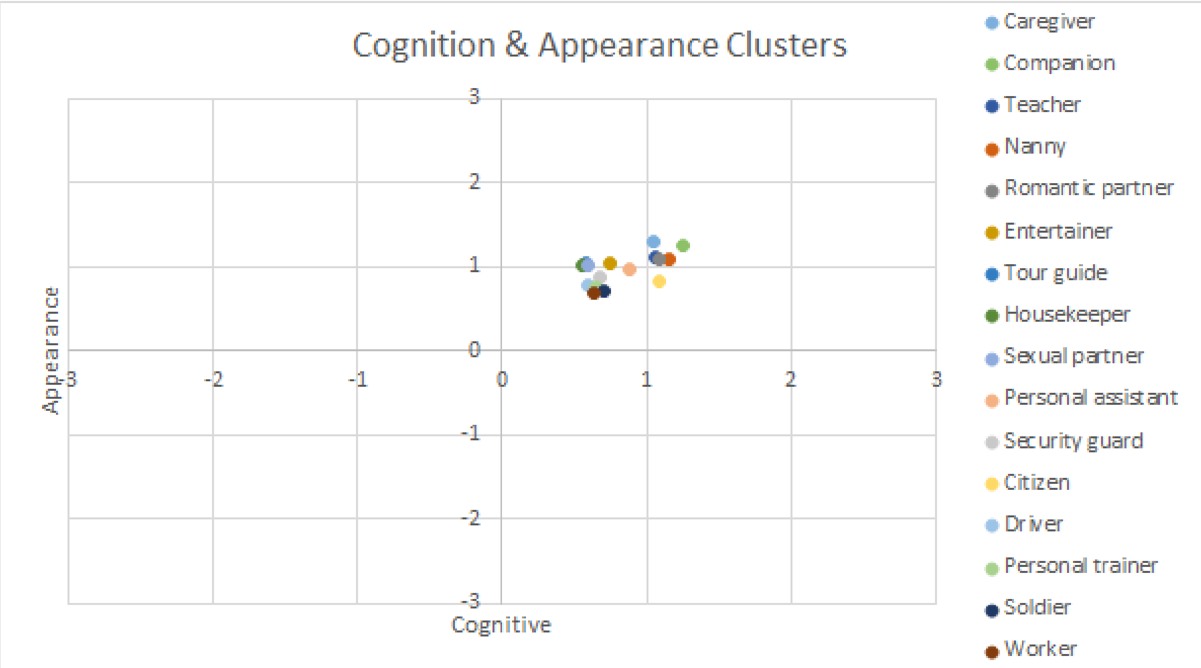


Figure 11 – Cognition and Appearance scores per future robot role

However, when rescaled, some clusters can be distinguished (see Figure 12).

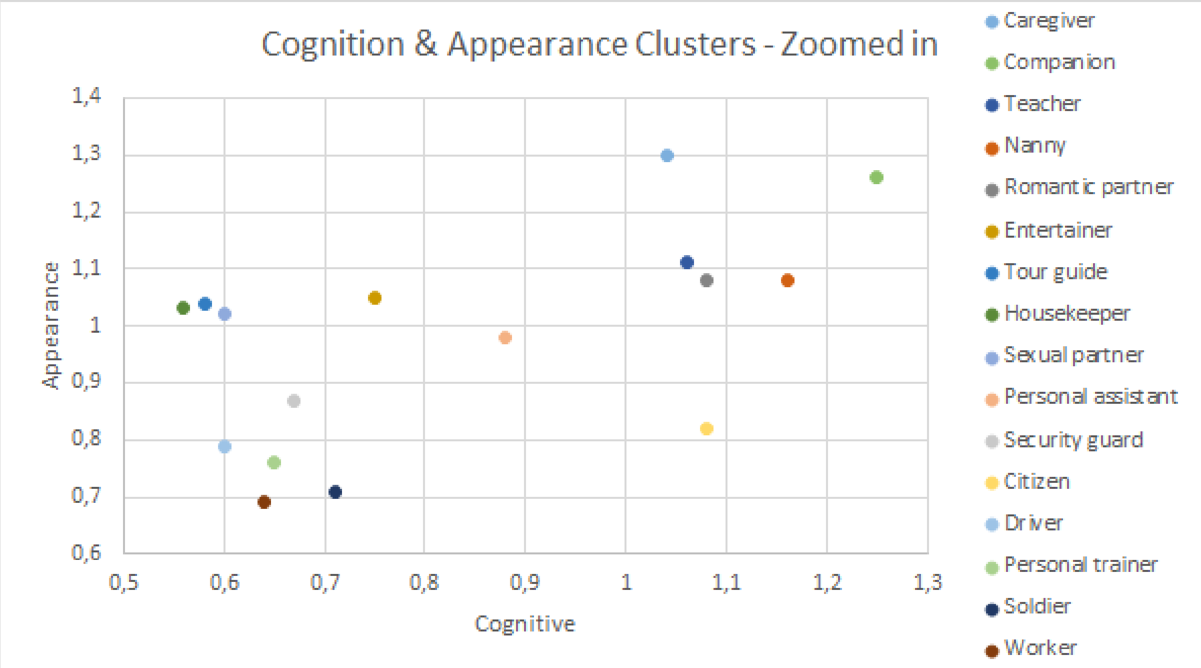


Figure 12 - Rescaled view of Cognition and Appearance scores per future robot role

Subsequently, a Hierarchical Cluster Analysis was conducted using the Centroid Linkage method for the clustering of the future robot roles. Additionally, the Squared Euclidean distance method was used for determining the distance between clusters and all scores were standardized to z-scores in order to yield equal metrics and equal weighting. This analysis produced a dendrogram (See Figure 13) that revealed a 3-cluster solution at Stage 4 of the cluster analysis, with 1 outlier.

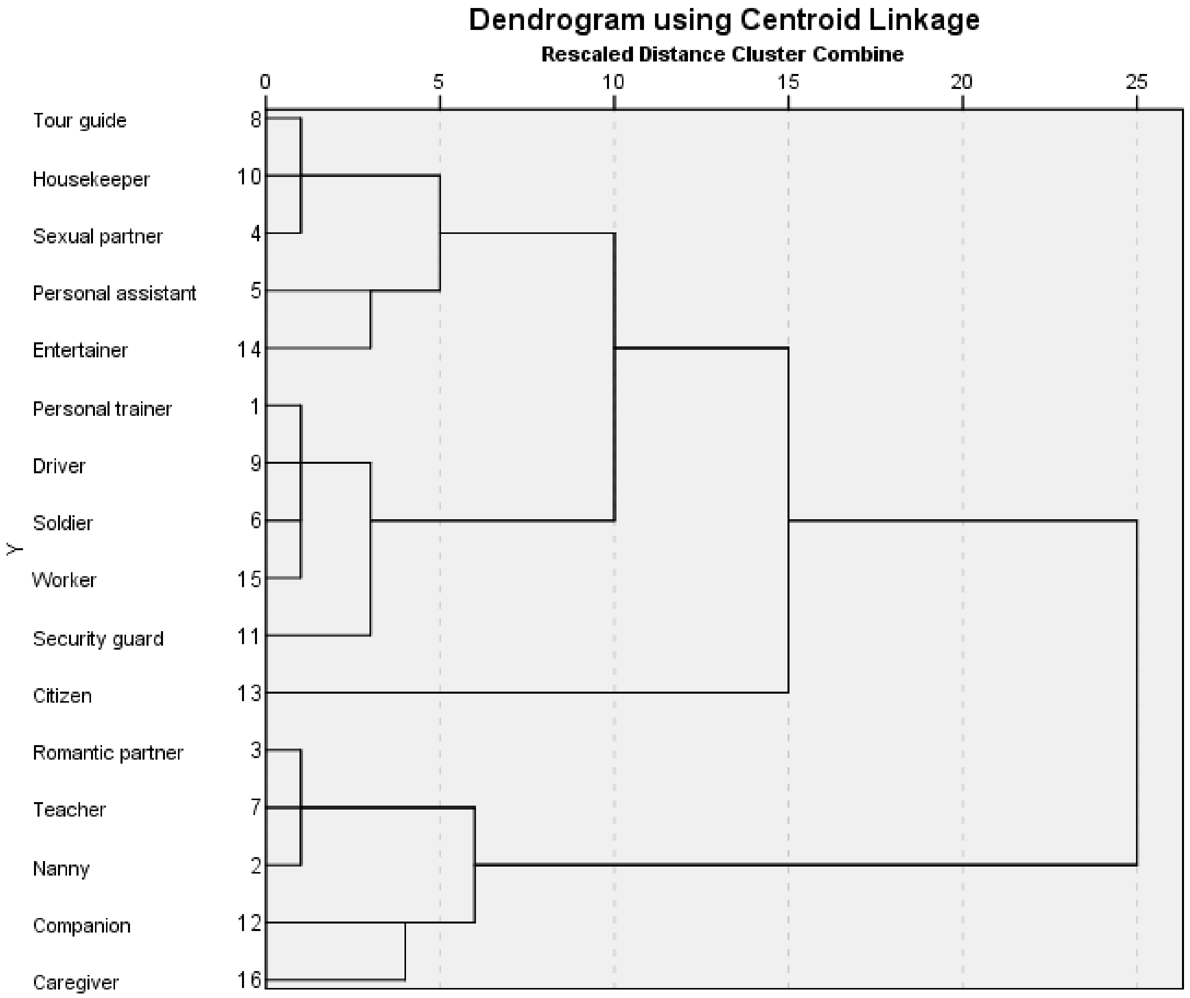


Figure 13 - Hierarchical Cluster Analysis using the Cognition and Appearance scores of each future robot role

1. Cluster A (Low Cognition score. High Appearance score)
    1. Tour guide
    2. Housekeeper
    3. Sexual partner
    4. Personal assistant
    5. Entertainer
2. Cluster B (Low Cognition score. Low Appearance score)
    1. Personal trainer
    2. Driver
    3. Soldier
    4. Worker
    5. Security guard
3. Cluster C (High Cognition score. High Appearance score)
    1. Romantic partner
    2. Teacher
    3. Nanny
    4. Companion
    5. Caregiver
4. Outlier (High Cognition score. Low Appearance score)
    1. Citizen

Cluster A consists of future robot roles that received relatively low Cognition scores and relatively high Appearance scores. Meaning the participants deem it relatively important that the appearance of these robots is humanlike and relatively unimportant that the cognition of these robots is humanlike. Cluster B consists of future robot roles that received relatively low Appearance and Cognition scores. Meaning the participants deem it relatively unimportant that the appearance and cognition of these robots are humanlike. Cluster C consists of future robot roles that received relatively high Appearance and Cognition scores. Meaning the participants deem it relatively important that the appearance and cognition of these robots are humanlike. The future robot role 'Citizen' is considered an outlier because it is the only role that received a relatively low Appearance score and relatively high Cognition score. Meaning the participants deem it relatively unimportant that the appearance of such robots is humanlike and relatively important that the cognition of such robots is humanlike

4.6.2 Realistic & Opinion cluster

Less clear clusters were found than initially expected (see Figure 14).

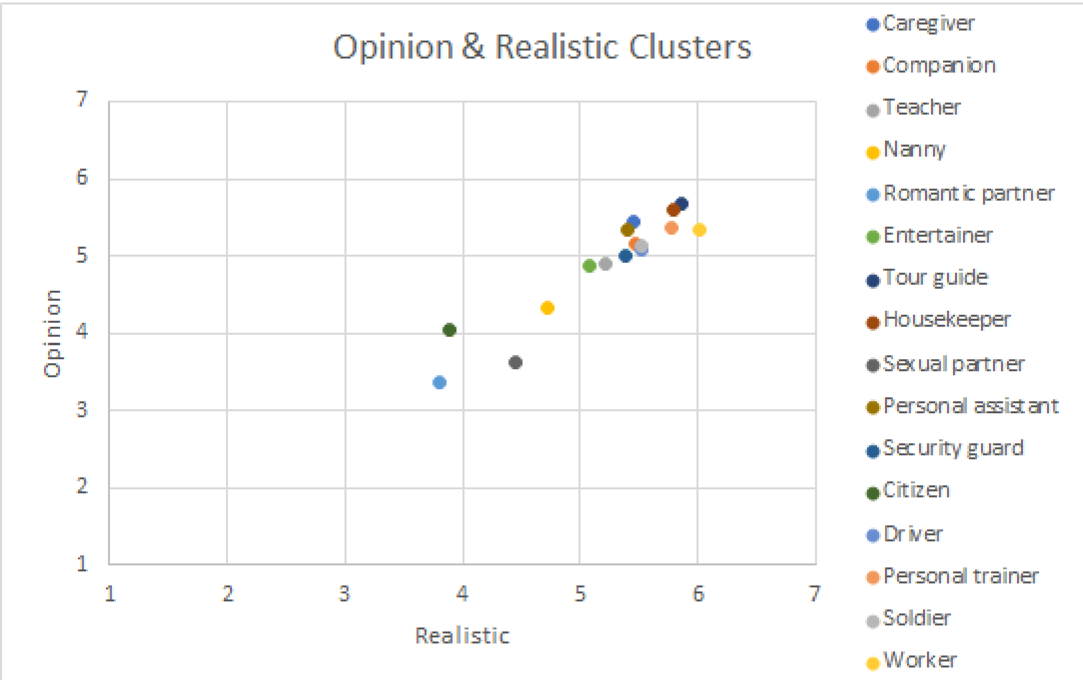


Figure 14 - Opinion and Realistic scores per future robot role

However, when rescaled, some clusters can be distinguished (see Figure 15).

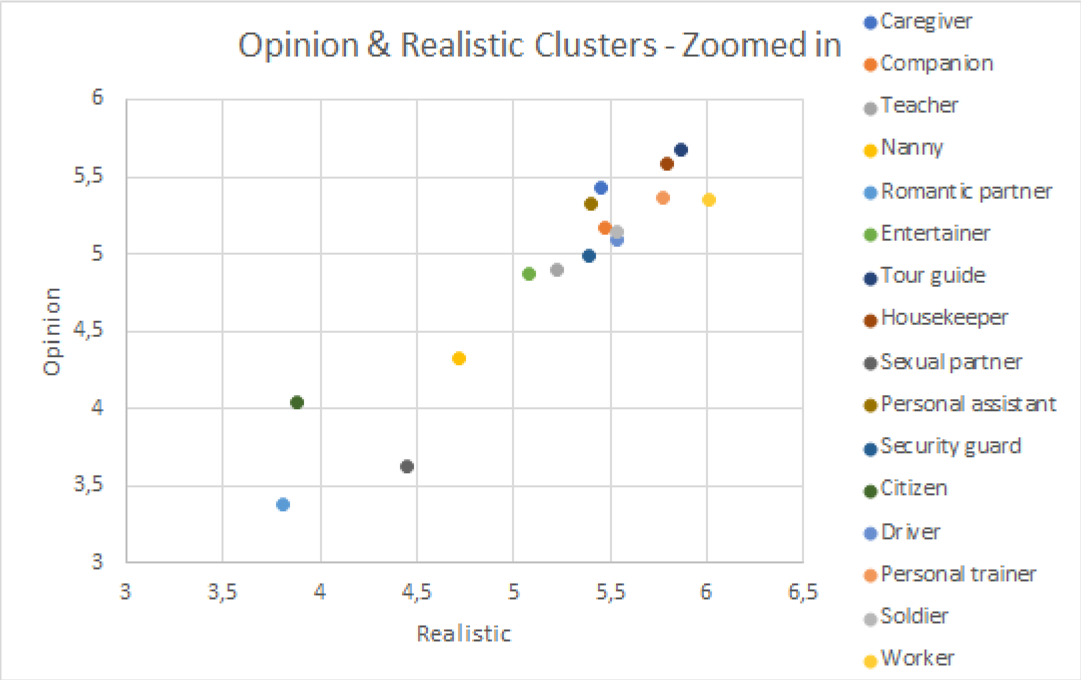


Figure 15 - Rescaled view of Opinion and Realistic scores per future robot role

Subsequently, a Hierarchical Cluster Analysis was conducted using the Centroid Linkage method for the clustering of the future robot roles. Additionally, the Squared Euclidean distance method was used for determining the distance between clusters and all scores were standardized to z-scores in order to yield equal metrics and equal weighting. This analysis produced a dendrogram (See Figure 16) that revealed a 2-cluster solution at Stage 3 of the cluster analysis, with 1 outlier.

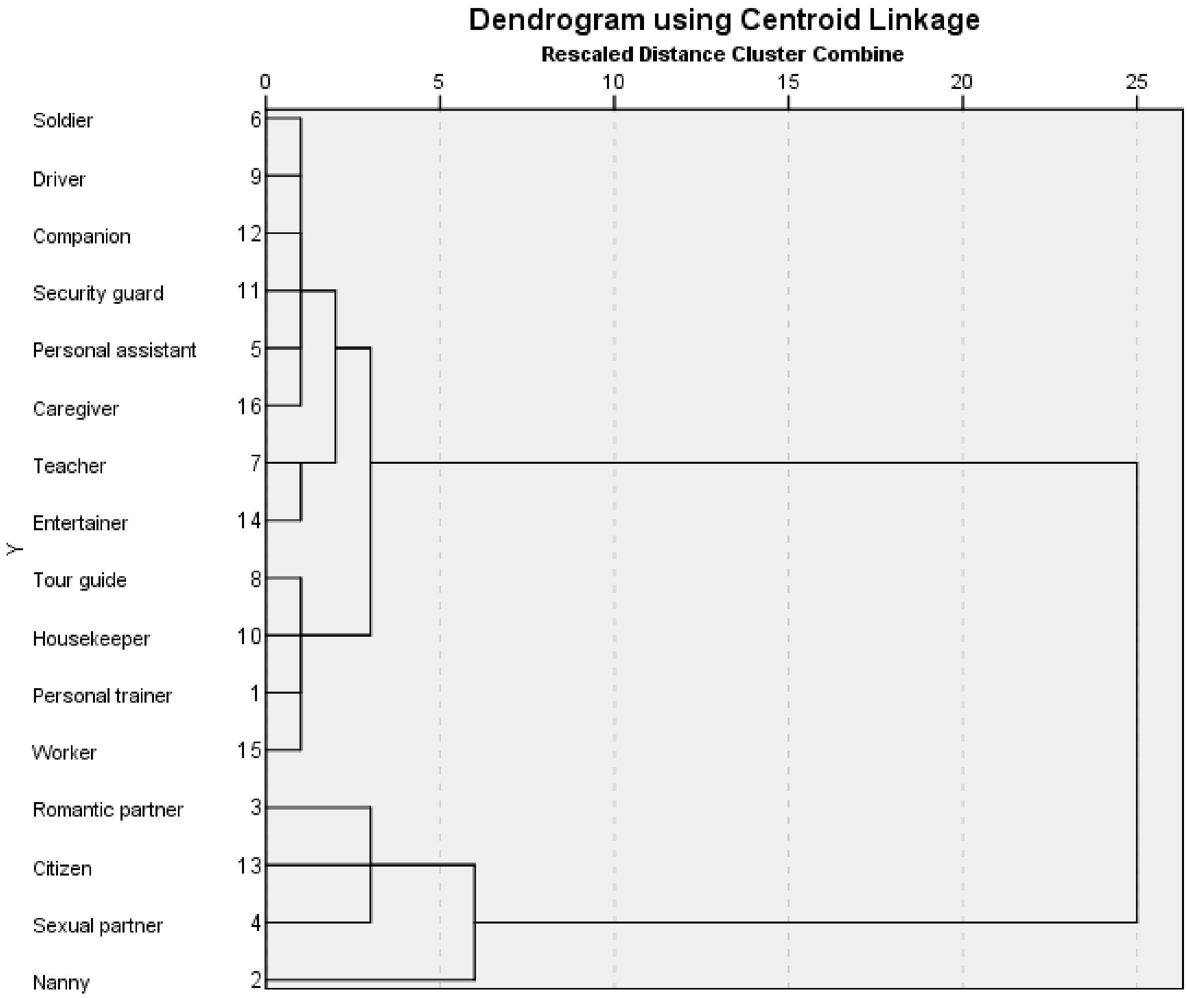


Figure 16 - Hierarchical Cluster Analysis using the Opinion and Realistic scores of each future robot role

1. Cluster D (Positive and realistic)
   1. Soldier
   2. Driver
   3. Companion
   4. Security guard
   5. Personal assistant
   6. Caregiver
   7. Teacher
   8. Entertainer
   9. Tour guide
   10. Housekeeper
   11. Personal trainer
   12. Worker
2. Cluster E (Negative and unrealistic)
   1. Romantic partner
   2. Citizen
   3. Sexual partner
3. Outlier (In between)
   1. Nanny

Cluster D consists of future robot roles that received relatively high Opinion and Realistic scores. Meaning the participants view these robots relatively positive and realistic. Cluster E consists of future robot roles that received relatively low Opinion and Realistic scores. Meaning the participants view these robots relatively negative and unrealistic. The future robot role 'Nanny' is considered an outlier because it is the only role that received neither low or high Opinion and Realistic scores.

4.7 Interactions

4.7.1 Interactions between gender and nationality

Four two-way ANOVA's were conducted that examined the effect of the participant's gender and nationality on the 'Cognition score'(1), 'Appearance score'(2), 'Realistic score'(3), and 'Opinion score'(4). There were no statistically significant interactions between the effects of gender and nationality on how the future robot roles were rated. This shows that the interaction between the gender and nationality of the participants does not significantly influence how humanlike the participants want a particular future robot role to be in terms of cognition or appearance, and that the interaction between the gender and nationality of the participants does not significantly influence participant's attitude towards a particular future robot role.

4.7.2 Interactions between gender and age

Four two-way ANOVA's were conducted that examined the effect of the participant's gender and age on the 'Cognition score'(1), 'Appearance score'(2), 'Realistic score'(3), and 'Opinion score'(4). There was a statistically significant interaction between the effects of gender and age on the 'Cognition score' at the $p<.05$ level [$F(3, 1777) = 5.21$, $p = 0.001$], and between the effects of gender and age on the 'Appearance score' at the $p<.05$ level [$F(3, 1777) = 5.08$, $p = 0.002$]. However, only the age group 45-60 had a significant difference between males and females in the 'Cognition score' ($p = 0.002$) and 'Appearance score' ($p = 0.003$). Males and females in the other age groups did not differ significantly from each other. Furthermore, there were no statistically significant interactions between the effects of gender and nationality on the 'Realistic score' and 'Opinion score'. This shows that the interaction between the gender and age of the participants influence how humanlike the participants want a particular future robot role to be in terms of cognition or appearance, but does not significantly influence participant's attitude towards a particular future robot role.

4.7.3 Interactions between gender and the level of education

Four two-way ANOVA's were conducted that examined the effect of the participant's gender and level of education on the 'Cognition score'(1), 'Appearance score'(2), 'Realistic score'(3), and 'Opinion score'(4). There were no statistically significant interactions between the effects of gender and level of education on how the future robot roles were rated. This shows that the interaction between the gender and level of education of the participants does not significantly influence how humanlike the participants want a particular future robot role to be in terms of cognition or appearance, and that the interaction between the gender and level of education of the participants does not significantly influence participant's attitude towards a particular future robot role.

4.7.4 Interactions between gender and future robot role

Four two-way ANOVA's were conducted that examined the effect of the participant's gender and what particular role was rated on the 'Cognition score'(1), 'Appearance score'(2), 'Realistic score'(3), and 'Opinion score'(4). There were no statistically significant interactions between the effects of gender and what particular role was rated on how the future robot roles were rated. This shows that the interaction between the gender and what particular role was rated does not significantly influence how humanlike the participants want a particular future robot role to be in terms of cognition or appearance, and that the interaction between the gender and what particular role was rated does not significantly influence participant's attitude towards a particular future robot role.

4.7.5 Interactions between age and future robot role

Four two-way ANOVA's were conducted that examined the effect of the participant's age and what particular role was rated on the 'Cognition score'(1), 'Appearance score'(2), 'Realistic score'(3), and 'Opinion score'(4). There was a statistically significant interaction between the effects of the participant's age and what particular role was rated on the 'Opinion score' at the $p<.05$ level [$F(45, 1721) = 1.66$, $p = 0.004$]. Between the age groups 18-29 and 45-60, the difference was significant for four future robot roles: 'Romantic partner' ($p < 0.001$), 'Sexual partner' ($p = 0.021$), 'Teacher' ($p = 0.018$), and 'Citizen' ($p = 0.002$). Between the age groups 30-44 and 60+, the difference was significant for four future robot roles: 'Sexual partner' ($p = 0.020$), 'Teacher' ($p = 0.041$), 'Companion' ($p = 0.006$), and 'Citizen' ($p = 0.008$). And between the age groups 18-29 and 60+, the difference was significant for four future robot roles: 'Soldier' ($p = 0.032$), 'Teacher' ($p = 0.011$), 'Security guard' ($p = 0.026$), and 'Citizen' ($p < 0.001$). Future robot roles in the other age groups did not differ significantly from each other. Furthermore, there were no statistically significant interactions between the effects of the participant's age and what particular role was rated on the 'Cognition score', 'Appearance score', and 'Realistic score'. This shows that the interaction between the age of the participants and what particular role was rated influences how negative or positive the participants view a particular future robot role to be, but does not significantly influence how humanlike the participants want a particular future robot role to be in terms of cognition or appearance, or how realistic the participants view a particular future robot role to be.

4.7.6 Interactions between the level of education and future robot role

Four two-way ANOVA's were conducted that examined the effect of the participant's level of education and what particular role was rated on the 'Cognition score'(1), 'Appearance score'(2), 'Realistic score'(3), and 'Opinion score'(4). There was a statistically significant interaction between the effects of the participant's level of education and what particular role was rated on the 'Realistic score' at the $p<.05$ level [$F(15, 1748) = 2.30$, $p = 0.003$]. However, only the future robot roles 'Romantic partner' ($p = 0.002$), 'Teacher' ($p = 0.010$), and 'Driver' ($p = 0.002$) had a significant difference in the 'Realistic score' between the 'Medium' and 'High' education level. Furthermore, there were no statistically significant interactions between the effects of the participant's level of education and what particular role was rated on the 'Cognition score', 'Appearance score', and 'Opinion score'. This shows that the interaction between the level of education of the participants and what particular role was rated, influences how realistic the participants view a particular future robot role to be, but does not significantly influence how humanlike the participants want a particular future robot role to be in terms of cognition or appearance, or how negative or positive the participants view a particular future robot role to be.

5 Design guidelines

5.1 Clusters

A subgoal of this study was to compose guidelines for the design of future robots, as this paves the way to guidelines that will inform technology design. First, a literature review was conducted in order to compose design guidelines, which are presented in '5.2 Guidelines'. It should be noted that these are not comprehensive guidelines for the design of future robots, but rather preliminary guidelines meant to propel the development of future robots in a desirable direction. Subsequently, the general public's desires of, and attitudes towards, the potential future robot roles in our society were investigated in order to establish a better understanding of how to make robots successful artificial companions. This study clustered robot roles together based on what appearance features and cognitive capacities were deemed necessary by the participants of this study. Table 6 and Table 7 in '5.2 Guidelines' indicate for which cluster each design guideline is relevant.

Cluster A: Tour guide, Housekeeper, Sexual partner, Personal assistant, and Entertainer.
The future robot roles in this cluster received a low Cognition score and a high Appearance score. These ratings show that the designers of these robots should focus on making the appearance of the robot humanlike rather than making the robots humanlike in terms of cognition.

Cluster B: Personal trainer, Driver, Soldier, Worker, and Security guard.
The future robot roles in this cluster received a low Cognition score and a low Appearance score.

Cluster C: Romantic partner, Teacher, Nanny, Companion, and Caregiver.
The future robot roles in this cluster received a high Cognition score and a high Appearance score. These ratings show that the designers of these robots should focus on making the appearance of the robot humanlike as well as making the robots humanlike in terms of cognition.

5.2 Guidelines

5.2.1 Appearance guidelines

Guideline	Cluster
The robot's role (intended use) should be taken into account while designing the appearance of the robot in order to enable intuitive human-robot interaction (Haring et al., 2016).	A, B & C
If the robot should gain the cooperation of people, the robot's appearance should match the seriousness of the task (Goetz et al., 2003).	A, B & C
If the robot will operate in a context where it will need to make moral decisions, the robot's human likeness in terms of appearance should match its capability to make moral decisions (Malle et al., 2016).	A, B & C
If people should empathize with the robot, the robot should have a humanlike appearance (Riek et al., 2009).	A & C
If people should provide a lot of information and explain themselves to the robot, the robot should violate the stereotype associated with the social role it fulfills (Powers et al., 2005).	A & C
If the robot should be perceived as humanlike as possible, the many details concerning its facial aspects should be considered and a mismatch among elements (e.g. humanlike eyes combined with an absence of skin) should be avoided (Ho et al., 2008).	A & C
If people should delegate responsibility to the robots or when the task is too demanding for people to do, the robot should have a humanlike appearance (Hinds et al., 2004).	A & C
If the robot's role will require a significant amount of social interaction, the robot should have a playful childlike appearance (Złotowski, Khalil, & Abdallah, 2019).	A & C
If the robot should have a minimal and efficient conversation with people, the robot should fit the stereotype associated with the social role it fulfills (Powers et al., 2005).	B
If people should feel responsible when successfully completing a task with a robot, the robot should have a mechanical appearance (Hinds et al., 2004).	B
If the robot's role belongs to the dull, dangerous and dirty category, and it will not have much social communication with humans, the robot should have a machine-like appearance (Złotowski et al., 2019).	B

Table 6 - Guidelines for the design of the appearance of future robots

5.2.2 Cognition guidelines

Guideline	Cluster
The robot's role (intended use) should be taken into account while developing the cognition of the robot in order to enable intuitive human-robot interaction (Haring et al., 2016).	A, B & C
If people should be able to interact efficient and intuitively with the robot, it should communicate its internal state through visual cues or continual feedback (Breazeal & Brooks, 2005).	A, B & C
If people expect to spend their free time with the robot, the robot should behave unpredictable (Eyssel et al., 2001).	A, B & C
If the robot will interact verbally with people, the many details concerning the robot's vocal content and paralinguistic cues such as volume and speech rate should be considered (Nass & Lee, 2001).	A, B & C
If the robot will interact with introverted individuals, the robot should avoid invading the individual's personal space (Tapus et al., 2008).	A, B & C
If the robot should be perceived as humanlike as possible, the robot should convey emotional reactions to the human interaction partner (Eyssel et al., 2010).	A & C

Table 7 - Guidelines for the design of the cognition of future robots

6 Discussion

6.1 Findings

6.1.1 How society views robots

Enz et al. (2011) found that people are skeptical towards robots in areas where they will act as social role models such as caretaking, education, and public security. In the context of the general public's attitude towards future robot roles, the findings of this study do not agree with those of Enz et al. (2011), but do agree with the findings of de Graaf and Allouch (2016) in the sense that robot roles that challenge social norms, such as robots as romantic partners, are viewed negatively. Moreover, the findings of this study agree with de Graaf and Allouch (2016) on the premise that future robot roles that ensure efficiency, convenience, and a decrease of casualties, are viewed positively.

Several studies have reported that people associate future robots with the negative prospect of the loss of jobs (Takayama et al., 2008; Eurobarometer, 2012; de Graaf and Allouch, 2016). Interestingly, the participants of this study rated the 'Worker' role positively. One possible explanation for this is that the role description this study used for the 'Worker' role did not explicitly mention that this robot role could potentially cause humans to lose their job due to robots replacing them. Another possible explanation is that the participants of this study were aware that, as with earlier major revolutions in our society, the introduction of new technologies created jobs elsewhere (Atkeson & Kehoe, 2001). Similarly, the findings of this study agree with Takayama et al. (2008) that people prefer robots for tasks that require memorization, keen perceptual skills, and service-orientation.

6.1.2 Different views within society

The findings of this study agree with those of Enz et al. (2011), who found that males are generally more positive towards future robots than females are. Moreover, this study shows that males also view future robots to be more realistic compared to female participants.

Scopelliti, Giuliani and Fornara (2005) found that older people are more mistrustful and fearful of new technology than younger people. They also found that older people are less confident in the abilities of robots. These findings are in line with the findings of this study that younger people view future robots more positively and to be more realistic compared to older people. Moreover, this study shows that younger people want future robots to be more humanlike in terms of appearance and cognition compared to older people.

Eurobarometer (2012) found that, in general, the higher the number of years EU citizens stayed in full-time education, the more positive their view of robots were. In a similar vein, the findings of this study agree with those of Eurobarometer (2012) by suggesting that people with a 'high' education level (bachelor's, master's, or doctoral degree) want future robots to be more humanlike in terms of cognition compared to people with a 'medium' education level (high school). One could question the validity of this finding due to a possible difference between the age of the people with a medium and high level of education. However, people with a high education level were on average older compared to people with a medium education level. As described above, younger people want future robots to be more humanlike in terms of cognition compared to older people. Meaning people with a high education level want robots to be more humanlike in terms of cognition compared to people with a medium education level despite being older than people with a medium education level.

The findings of this study also suggest that Asian people want future robots to be more humanlike in terms of cognition and appearance compared to people from other continents. Moreover, the results suggest that Asian people view future robots more positively and to be more realistic compared to people from other continents.

6.2 Limitations

The results of this study can prove useful for future research into the design of future robots. However, the research method that was used could have some potential drawbacks. First, on the one hand, the use of text-based application scenarios (written future robot role descriptions) is believed to be an appropriate research method for this study's goal (Young, Hawkins, Sharlin, & Igarashi, 2009). On the other hand, it is expected that real interactions between participants and robots will produce different results due to the larger role emotions play in real life interactions compared to the use of written role descriptions (Hwang, Park, & Hwang, 2013). Second, 93% of the participants of this study that were categorized as 'Asian' were people from India. It should be noted that Indian culture differs from other Asian cultures. However, the same holds for European and North-American participants and this is an inevitable consequence of the decision to divide the world's population in 6 continents. Third, the study included participants that were almost exclusively recruited through the MTurk platform. On the one hand, MTurk participants can originate from all over the world, on the other hand, 91% of MTurk participants originate from India and North America (Difallah, et al., 2018). This could reduce the generalizability of this study's findings because people's nationality affects their evaluation of robots (Eurobarometer, 2012) as well as several cultural factors (Lee & Sabanović, 2014). It is therefore important to verify the current findings with the opinion of different nationalities and cultures. A third limitation is that no justifications for the participant's ratings were assessed. de Graaf and Allouch (2016) did investigate people's justifications for their evaluation of potential future robot applications. However, the study conducted by de Graaf and Allouch (2016) included only participants from The Netherlands. Meaning future research should point out how other nationalities and cultures evaluate potential future robot applications.

7 Conclusion

With robots becoming popular in a variety of areas of life and an expected growth of robots in our future society, the impact of robotics on society is likely to increase (Enz et al., 2011). To make robots successful artificial companions, the concerns and attitudes among potential human users need to be addressed. This is why the evaluation of potential future robot roles in our society was investigated from the perspective of the general public. A robot's appearance and its behavior shape people's attitudes towards the robot (DiSalvo et al., 2002; Haring et al., 2013; Phillip et al., 2018). Moreover, the more appearance features a robot has, the more intelligent it is perceived (Haring et al., 2016). This plays a crucial part in HRI because a mismatch between the robot's appearance and its behavior (cognition) could result in a rejection of this new technology. This means that the appearance features and cognitive capacities of a robot should take the intended use of the robot into account to prevent such a mismatch and enable an intuitive interaction between human and robot (Haring et al., 2016). Hence, this study conducted an online questionnaire on Amazon Turk (n= 510) where participants evaluated potential future robot application scenarios in order to map how humanlike people desire the future robots in our society to be in terms of appearance and cognition, and to examine people's attitudes towards potential future roles of robots.

The research reported here is explicitly explorative and the findings should be considered as a starting point for future research ideas and robot designs. This helps pave the way to formulating guidelines that will inform technology design with the goal of developing acceptable robots. The results of this study showed that people's nationality and age influence how humanlike they desire future robots to be in terms of appearance. Specifically, the results suggest that people from Asia desire future robots to appear more humanlike compared to people from Europe and North America, and people between the age of 18 and 30 desire future robots to appear more humanlike compared to older people. Furthermore, this study showed that people's nationality, age, and level of education influence how humanlike they desire future robots to be in terms of cognition. Specifically, the results suggest that people from Asia desire future robots to be cognitively more humanlike compared to people from Europe and North America, people between the age of 18 and 30 desire future robots to be cognitively more humanlike compared to older people, and people with a bachelor's degree or higher desire future robots to be cognitively more humanlike compared to people who do not hold a college degree. Meaning researchers and developers in robotics should not only focus on the technological part of the design of robots, but should also take the demographics of the people who will interact with the robot into account. Goetz et al. (2003) reached a similar conclusion by stating that basing a robot's appearance on the robot's role is a necessary, but not a sufficient aspect of designing acceptable social robots. In a similar vein, Złotowski et al. (2019) state that the design of the appearance of social robots should meet people's personal and cultural expectations in order to improve how suitable people perceive the robot to be for its task.

Lastly, this study showed that the nationality, age, and gender of people influence how realistic and how positive or negative they view the roles of future robots in our society. Specifically, the results suggest that people from Asia view the roles of future robots more positively and to be more realistic compared to people from Europe and North America, people between the age of 18 and 30 view the roles of future robots more positively and to be more realistic compared to older people, and males view the roles of future robots more positively and to be more realistic compared to females. The difference in how positive males view the roles of future robots and how positive females view the roles of future robots is in line with the findings of Enz et al. (2011). Meaning people's opinions should also be taken into account by researchers and developers in robotics because the general public's attitude plays a role in how successful new technology is adopted (Cowan, 1983) and because new technologies become more usable if people's expectations are taken into account (Norman, 1988).

References

Ajzen, I. (2001). Nature and operation of attitudes. Annual review of psychology, 52(1), 27-58.

Anderson, S. L. (2008). Asimov's "three laws of robotics" and machine metaethics. Ai & Society, 22(4), 477-493.

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. Journal of personality and social psychology, 37(5), 715.

Atkeson, A., & Kehoe, P. J. (2001). The transition to a new economy after the second industrial revolution (No. w8676). National Bureau of Economic Research.

Arkin, R. (2009). Governing lethal behavior in autonomous robots. Chapman and Hall/CRC.

Arras, K. O., & Cerqui, D. (2005). Do we want to share our lives and bodies with robots? A 2000 people survey: A 2000-people survey. Technical report, 605.

Asimov, I. (1964). The rest of the robots. Doubleday.

Bartneck, C., & Forlizzi, J. (2004, April). Shaping human-robot interaction: understanding the social aspects of intelligent robotic products. In CHI'04 Extended Abstracts on Human Factors in Computing Systems (pp. 1731-1732). ACM.

Beer, J. M., Smarr, C. A., Chen, T. L., Prakash, A., Mitzner, T. L., Kemp, C. C., & Rogers, W. A. (2012, March). The domesticated robot: design guidelines for assisting older adults to age in place. In Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction (pp. 335-342). ACM.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. Political analysis, 20(3), 351-368.

Bizer, G. Y., Tormala, Z. L., Rucker, D. D., & Petty, R. E. (2006). Memory-based versus on-line processing: Implications for attitude strength. Journal of Experimental Social Psychology, 42(5), 646-653.

Breazeal, C. (2003). Emotion and sociable humanoid robots. International journal of human-computer studies, 59(1-2), 119-155.

Breazeal, C., & Brooks, R. (2005). Robot emotion: A functional perspective. Who needs emotions, 271-310.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?. Perspectives on psychological science, 6(1), 3-5.

Chang, W. L., Šabanovic, S., & Huber, L. (2013, March). Use of seal-like robot PARO in sensory group therapy for older adults with dementia. In 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 101-102). IEEE.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. Perspectives on socially shared cognition, 13(1991), 127-149.

Cowan, R. W. More work for mother. Basic Books, New York, NY, USA, 1983.

Dautenhahn, K., Woods, S., Kaouri, C., Walters, M. L., Koay, K. L., & Werry, I. (2005, August). What is a robot companion-friend, assistant or butler?. In 2005 IEEE/RSJ international conference on intelligent robots and systems (pp. 1192-1197). IEEE

de Graaf, M. M., & Allouch, S. B. (2016, August). Anticipating our future robot society: The evaluation of future robot applications from a user's perspective. In 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) (pp. 755-762). IEEE.

Difallah, D., Filatova, E., & Ipeirotis, P. (2018, February). Demographics and dynamics of mechanical Turk workers. In Proceedings of the eleventh acm international conference on web search and data mining (pp. 135-143). ACM.

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

Dragan, A. D., Bauman, S., Forlizzi, J., & Srinivasa, S. S. (2015, March). Effects of robot motion on human-robot collaboration. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (pp. 51-58). ACM.

Eagly A, Chaiken S (1993) Psychology of attitudes. Harcourt Brace Jovanovich, New York

Eagly, A. H., & Chaiken, S. (2005). Attitude Research in the 21st Century: The Current State of Knowledge.

Enz, S., Diruf, M., Spielhagen, C., Zoll, C., & Vargas, P. A. (2011). The social role of robots in the future - explorative measurement of hopes and fears. International Journal of Social Robotics, 3(3), 263.

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

Eurobarometer, S. (2012). Public attitudes towards robots. Special Eurobarometer 382. European Commission.

Eyssel, F., De Ruiter, L., Kuchenbrandt, D., Bobinger, S., & Hegel, F. (2012, March). ‘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 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 125-126). IEEE.

Eyssel, F., Hegel, F., Horstmann, G., & Wagner, C. (2010, September). Anthropomorphic inferences from emotional nonverbal cues: A case study. In 19th international symposium in robot and human interactive communication (pp. 646-651). IEEE.

Eyssel, F., Kuchenbrandt, D., & Bobinger, S. (2011, March). Effects of anticipated human-robot interaction and predictability of robot behavior on perceptions of anthropomorphism. In Proceedings of the 6th international conference on Human-robot interaction (pp. 61-68). ACM.

Goetz, J., Kiesler, S., & Powers, A. (2003, October). Matching robot appearance and behavior to tasks to improve human-robot cooperation. In The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. (pp. 55-60). Ieee.

Goldman, A. I. (2006). Simulating minds: The philosophy, psychology, and neuroscience of mindreading. Oxford University Press.
Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. science, 315(5812), 619-619.

Guo S, Zhang G (2009) Robot rights. Science 323:876

Hara, K., Adams, A., Milland, K., Savage, S., Hanrahan, B. V., Bigham, J. P., & Callison-Burch, C. (2019, April). Worker Demographics and Earnings on Amazon Mechanical Turk: An Exploratory Analysis. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (p. LBW1217). ACM.

Haring, K. S., Silvera-Tawil, D., Takahashi, T., Watanabe, K., & Velonaki, M. (2016, February). How people perceive different robot types: A direct comparison of an android, humanoid, and non-biomimetic robot. In 2016 8th International Conference on Knowledge and Smart Technology (KST) (pp. 265-270). IEEE.

Haring, K. S., Watanabe, K., & Mougenot, C. (2013, March). The influence of robot appearance on assessment. In 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 131-132). IEEE

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

Ho, C. C., MacDorman, K. F., & Pramono, Z. D. (2008, March). Human emotion and the uncanny valley: a GLM, MDS, and Isomap analysis of robot video ratings. In 2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 169-176). IEEE.

Hwang, J., Park, T., & Hwang, W. (2013). The effects of overall robot shape on the emotions invoked in users and the perceived personalities of robot. Applied ergonomics, 44(3), 459-471.

I. Asimov, The Rest of the Robots (a collection of short stories originally published between 1941 and 1957), Grafton Books, London, 1968.

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

Lee, H. R., & Sabanović, S. (2014, March). Culturally variable preferences for robot design and use in South Korea, Turkey, and the United States. In Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (pp. 17-24). ACM.

M. Mori, "Bukimi no tami", Energy, vol. 7, no. 4, pp. 33-35, 1970.

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. Interaction Studies, 7(3), 297-337.

Malle, B. F. (2019). How many dimensions of mind perception really are there? In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), Proceedings of the 41st Annual Meeting of the Cognitive Science Society (pp. 2268-2274). Montreal, QB: Cognitive Science Society.

Malle, B. F., & Thapa Magar, S. (2017, March). What kind of mind do I want in my robot?: Developing a measure of desired mental capacities in social robots. In Proceedings of the companion of the 2017 ACM/IEEE international conference on human-robot interaction (pp. 195-196). ACM.

Malle, B. F., Scheutz, M., Forlizzi, J., & Voiklis, J. (2016, March). Which robot am I thinking about?: The impact of action and appearance on people's evaluations of a moral robot. In The Eleventh ACM/IEEE International Conference on Human Robot Interaction (pp. 125-132). IEEE Press.

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. Cognition, 146, 22-32.

Moran, M. (2008). Three laws of robotics and surgery. Journal of endourology, 22(8), 1557-1560.

Morris, A., Donamukkala, R., Kapuria, A., Steinfeld, A., Matthews, J. T., Dunbar-Jacob, J., & Thrun, S. (2003, September). A robotic walker that provides guidance. In 2003 IEEE International Conference on Robotics and Automation (Cat. No. 03CH37422) (Vol. 1, pp. 25-30). IEEE.

Mukai, T., Hirano, S., Nakashima, H., Kato, Y., Sakaida, Y., Guo, S., & Hosoe, S. (2010, October). Development of a nursing-care assistant robot RIBA that can lift a human in its arms. In 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 5996-6001). IEEE.

Murphy, R., & Woods, D. D. (2009). Beyond Asimov: the three laws of responsible robotics. IEEE Intelligent Systems, 24(4), 14-20.

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. Journal of experimental psychology: applied, 7(3), 171.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. Journal of social issues, 56(1), 81-103.

Norman, D.A. The Psychology of Everyday Things. Basic Books, New York, NY, USA, 1988.

Norman, D. (2001). How might humans interact with robots, keynote address to the DARPA/NSF workshop on Human-Robot Interaction, San Luis Obispo, CA.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. Judgment and Decision making, 5(5), 411-419.

Phillips, E., Zhao, X., Ullman, D., & Malle, B. F. (2018, February). What is human-like?: Decomposing robots' human-like appearance using the anthropomorphic robot (abot) database. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (pp. 105-113). ACM.

Pittam, J. (1994). Voice in social interaction (Vol. 5). Sage.

Powers, A., Kramer, A. D., Lim, S., Kuo, J., Lee, S. L., & Kiesler, S. (2005, August). Eliciting information from people with a gendered humanoid robot. In ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005. (pp. 158-163). IEEE.

RAE (2009). Electronic report of the royal academy of engineering on “autonomous systems: social, legal and ethical issues”.

Rand, D. G. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. Journal of theoretical biology, 299, 172-179.

Ray, C., Mondada, F., & Siegwart, R. (2008, September). What do people expect from robots?. In 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 3816-3821). IEEE.
Reeves, B., & Nass, C. I. (1996). The media equation: How people treat computers, television, and new media like real people and places. Cambridge university press.

Riek, L. D., Rabinowitch, T. C., Chakrabarti, B., & Robinson, P. (2009, September). Empathizing with robots: Fellow feeling along the anthropomorphic spectrum. In 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops (pp. 1-6). IEEE.

Royakkers, L., & van Est, R. (2015). A literature review on new robotics: automation from love to war. International journal of social robotics, 7(5), 549-570.

Scopelliti, M., Giuliani, M. V., & Fornara, F. (2005). Robots in a domestic setting: a psychological approach. Universal access in the information society, 4(2), 146-155.

Seyama, J. I., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. Presence: Teleoperators and virtual environments, 16(4), 337-351.

Sloman, A. (2006). Why Asimov's three laws of robotics are unethical. Retrieved June, 9, 2006.

Stafford, R. Q., MacDonald, B. A., Jayawardena, C., Wegner, D. M., & Broadbent, E. (2014). Does the robot have a mind? Mind perception and attitudes towards robots predict use of an eldercare robot. International journal of social robotics, 6(1), 17-32.

Sytsma, J. (2014). The robots of the dawn of experimental philosophy. Current controversies in experimental philosophy. E. Machery & E. O'Neill, eds. Routledge. 48–64.

Takayama, L., Ju, W., & Nass, C. (2008, March). Beyond dirty, dangerous and dull: what everyday people think robots should do. In 2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 25-32). IEEE.

Tapus, A., Ţăpuş, C., & Matarić, M. J. (2008). User-robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. Intelligent Service Robotics, 1(2), 169.

Tusing, K. J., & Dillard, J. P. (2000). The sounds of dominance. Vocal precursors of perceived dominance during interpersonal influence. Human Communication Research, 26(1), 148-171.

United Nations & The International Federation of Robotics, World Robotics. Statistics, Market Analysis, Forecasts, Case Studies and Profitability of Robot Investment, New York and Geneva, 2006.

Vargas, P. A., Ho, W. C., Lim, M., Enz, S., & Aylett, R. (2009). To forget or not to forget: towards a roboethical memory control. Killer Robots or Friendly Fridges: the Social Understanding of Artificial Intelligence, AISB, 9, 18-23.

Veruggio, G., & Operto, F. (2006). Roboethics: a bottom-up interdisciplinary discourse in the field of applied ethics in robotics. International review of information ethics, 6(12), 2-8.

Vriens, M., Oppewal, H., & Wedel, M. (1998). Ratings-based versus choice-based latent class conjoint models. Market Research Society. Journal., 40(3), 1-11.

Wada, K., & Shibata, T. (2007, August). Robot therapy in a care house-Change of relationship among the residents and seal robot during a 2-month long study. In RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication (pp. 107-112). IEEE.

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. Trends in cognitive sciences, 14(8), 383-388.

Wohlin, C. (2014, May). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In Proceedings of the 18th international conference on evaluation and assessment in software engineering (p. 38). ACM.

Woods, D. D., & Hollnagel, E. (2006). Joint cognitive systems: Patterns in cognitive systems engineering. CRC Press.

Wu, X., Gong, H., Chen, P., Zhi, Z., & Xu, Y. (2009, February). Intelligent household surveillance robot. In 2008 IEEE International Conference on Robotics and Biomimetics (pp. 1734-1739). IEEE.

Young, J. E., Hawkins, R., Sharlin, E., & Igarashi, T. (2009). Toward acceptable domestic robots: Applying insights from social psychology. International Journal of Social Robotics, 1(1), 95.

Złotowski, J., Khalil, A., & Abdallah, S. (2019). One robot doesn't fit all: aligning social robot appearance and job suitability from a Middle Eastern perspective. AI & SOCIETY, 1-16.

Appendices

Appendix A - Future robot role descriptions

1. Personal assistant

Robots can perform their tasks independently, will follow any order without question, and will always do their work enthusiastically. Therefore, robots can serve as a personal assistant. They will look up information for you, will manage your agenda and remind you of appointments, and support you in organizing parties or meetings.

2. Nanny

Robots are on stand-by day and night, are able to read social cues, and can easily adjust to our personal needs and preferences. Therefore, robots could serve as a nanny. They will help your child with bathing, grooming, and dressing, provide first aid and call in the emergency services if needed, and entertain your child with educational games.

3. Romantic partner

Robots can read our social cues, will respond in a social way, and can easily adjust to our personal needs and preferences. Therefore, robots could serve as a romantic partner. They will match your ideal image of a partner, provide emotional support, and be reliable and loyal to you.

4. Sexual partner

Robots are on stand-by day and night, will follow any order without question, and will respond in a social way. Therefore, robots could serve as a sexual partner. They will satisfy your sexual needs, always respond to your sexual advances, and will explore sexual fantasies together.

5. Personal trainer

Robots are on stand-by day and night, can read our social cues, and can easily adjust to our personal needs and preferences. Therefore, robots could serve as a personal trainer. They will help making healthy choices in your eating pattern, encourage you to exercise sufficiently, and guide you during (sports)exercises.

6. Soldier
Robots are on standby day and night, follow any order without question, and can perform their tasks independently. Therefore, robots could serve as a soldier. They will defuse bombs, recognize and kill enemies, and bring civilians to safety.

7. Teacher
Robots will always do their work enthusiastically, can easily adjust to our personal needs and preferences, and will respond in a social way. Therefore, robots could serve as a teacher. They will transfer knowledge, provide personalized instructions, and assess students' performances.

8. Tour guide
Robots are on stand-by day and night, will always do their work enthusiastically, and can easily adjust to our personal needs and preferences. Therefore, robots can be used as a tour guide. They will show you around in museums, answer all your questions, and only provide background information about the things that interest you.

9. Driver
Robots can focus their undivided attention on their duties, will follow any order without question, and can easily adjust to our personal needs and preferences. Therefore, robots could serve as a driver. They will reach your destination in a timely and safe matter, will always have a clean driving record, and will be courteous to other road users.

10. Housekeeper
Robots are on stand-by day and night, will follow any order without question, and can easily adjust to our personal needs and preferences. Therefore, robots could serve as a housekeeper. They will clean our house, do our groceries, and maintain our garden.

11. Security guard
Robots can focus their undivided attention on their duties, are on stand-by day and night, and can perform their tasks independently. Therefore, robots could serve as a security guard. They will enforce the rules of the law, apprehend or evict violators from premises, and act quickly and appropriately during emergencies.

12. Companion
Robots can focus their undivided attention on us, can read our social cues, and will respond in a social way. Therefore, robots could serve as a companion. They will watch television with us, play a game, or join us in a conversation.

13. Citizen
Robots can perform their tasks independently, have their own preferences, and possess super intelligence. Therefore, robots could hold basic rights. They will have the right not to be mistreated, may say whatever they want whenever they want to, and will be held responsible when something goes wrong.

14. Entertainer
Robots can focus their undivided attention on us, will always do their work enthusiastically, and can easily adapt to our personal needs and preferences. Therefore, robots could serve as an entertainer. They will be able to tell fascinating stories, play our favorite music, or do magic tricks.

15. Worker
Robots are strong machines, are on stand-by day and night, and will follow any order without question. Therefore, robots could serve as a worker. They will perform assembly line work, move heavy materials, and perform tasks with hazardous substances or work in danger zones.

16. Caregiver

Robots can read social cues, are on stand-by day and night, and can easily adapt to our personal needs and preferences. Therefore, robots could serve as a caregiver. They will support people with a physical disability in daily activities, aid the elderly with bathing, grooming, and dressing, and help people with chronic diseases cope with their condition.

Appendix B - Screenshot of question about cognitive capacities

Universiteit Utrecht

Role description:

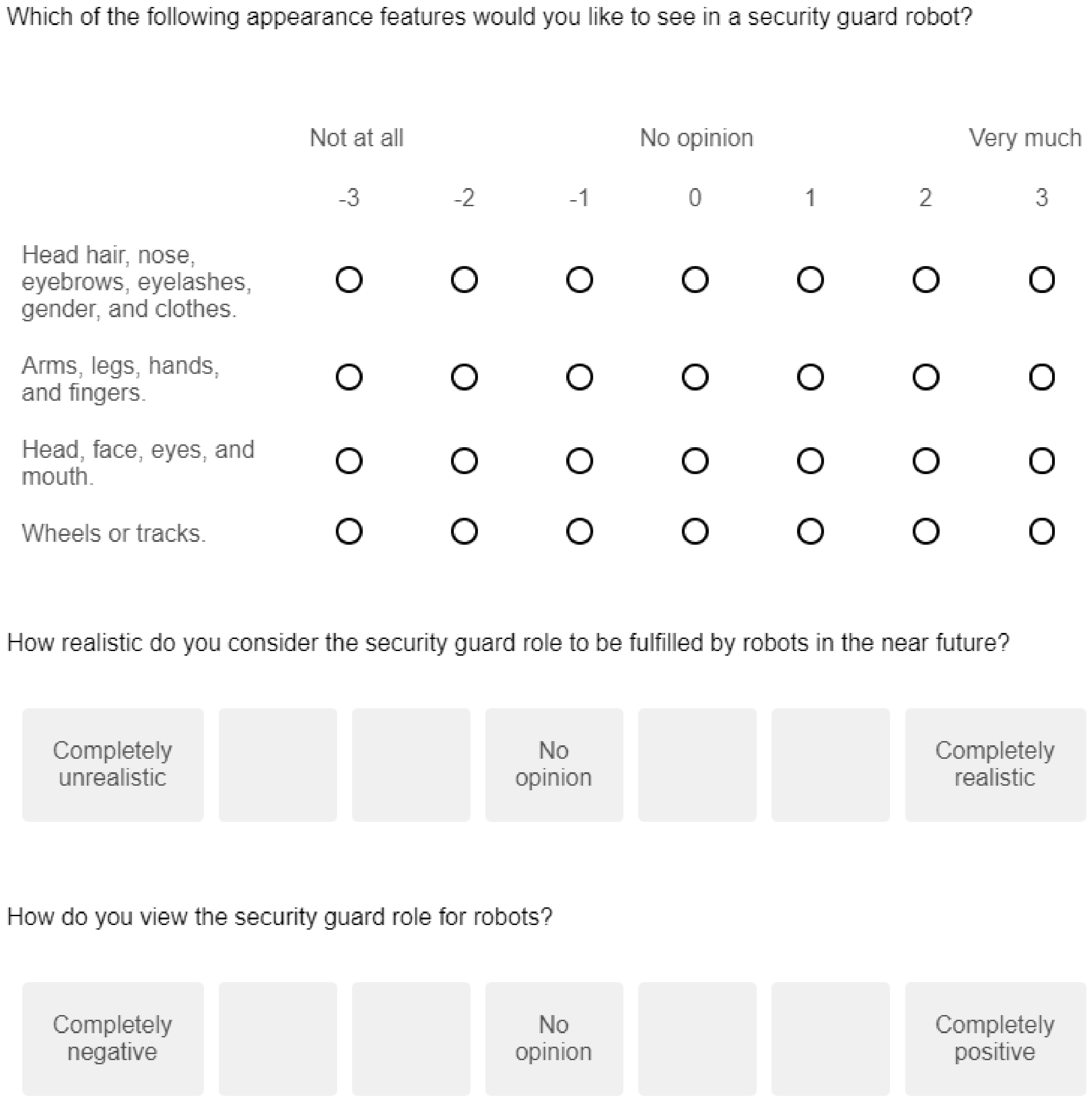
Robots can focus their undivided attention on their duties, are on stand-by day and night, and can perform their tasks independently. Therefore, robots could serve as a security guard. They will enforce the rules of the law, apprehend or evict violators from premises, and act quickly and appropriately during emergencies.

Questions:

Which of the following capacities would you like to see in a security guard robot?

	Not at all			No opinion			Very much
	-3	-2	-1	0	1	2	3
Feeling pleasure, feeling happy, experiencing gratitude, and loving specific people or things.	○	○	○	○	○	○	○
Feeling pain, feeling stressed, feeling tired, and experiencing fear.	○	○	○	○	○	○	○
Telling right from wrong, disapproving immoral actions, upholding moral values, and praising moral actions.	○	○	○	○	○	○	○
Inferring a person's' thinking, understanding others' minds, planning for the future, setting goals.	○	○	○	○	○	○	○
Communicating verbally, seeing and hearing the world, learning from instruction, and moving on their own.	○	○	○	○	○	○	○

Appendix C - Screenshot of question about appearance features



Which of the following appearance features would you like to see in a security guard robot?

	Not at all			No opinion			Very much
	-3	-2	-1	0	1	2	3
Head hair, nose, eyebrows, eyelashes, gender, and clothes.	○	○	○	○	○	○	○
Arms, legs, hands, and fingers.	○	○	○	○	○	○	○
Head, face, eyes, and mouth.	○	○	○	○	○	○	○
Wheels or tracks.	○	○	○	○	○	○	○

How realistic do you consider the security guard role to be fulfilled by robots in the near future?

Completely unrealistic			No opinion			Completely realistic

How do you view the security guard role for robots?

Completely negative			No opinion			Completely positive

Appendix D - Screenshot of questions about participant's attitude

Which of the following appearance features would you like to see in a security guard robot?

	Not at all			No opinion			Very much
	-3	-2	-1	0	1	2	3
Head hair, nose, eyebrows, eyelashes, gender, and clothes.	O	O	O	O	O	O	O
Arms, legs, hands, and fingers.	O	O	O	O	O	O	O
Head, face, eyes, and mouth.	O	O	O	O	O	O	O
Wheels or tracks.	O	O	O	O	O	O	O

How realistic do you consider the security guard role to be fulfilled by robots in the near future?

Completely unrealistic			No opinion			Completely realistic

How do you view the security guard role for robots?

Completely negative			No opinion			Completely positive