How to Make Your Robot Cute

Examining and Categorizing the Effects of a Robot's Appearance on Human Likeability

Babs van der Hoeven

SNR: 1269662

Master's Thesis Communication and Information Sciences

Track New Media Design

School of Humanities and Digital Sciences

Tilburg University, Tilburg

Supervisor: Jan de Wit

Second Supervisor: prof. dr. Emiel Krahmer

Second Reader: prof. dr. Marc Swerts

January 12th, 2021

Table of contents

Abstract

A large amount of research has been done on the design and functionality of robots in the field of human-robot interaction. However, previous research is conducted in a way where it is difficult to compare studies: separate studies examine different study designs, factors and circumstances, while covering a broad subject. This systematic literature review aims to structurize previous research on robot appearances while specifying the effect robot appearance has on a human's likeability rating of that robot. A concise definition for likeability is proposed, which is used to collect relevant literature on the subject. By using this definition in an explorative systematic search, n = 42 articles are summarized and presented, and an additional image database is made from the robots used in the relevant studies. The period of time covered in this literature review is from 2007 until March 2020. Next, all robot designs are categorized on both domain of use and appearance traits, in order to examine which effect each specific trait has on likeability, regarding each domain. To do so, a classification of human-like, animal-like and machine-like traits is made and presented, to apply to the robot image database. As a result, each appearance trait's occurrence and effectiveness is reported, followed by assumptions and conclusions. Finally, an overview of suggestive guidelines is presented for further research, along with additional suggestions for future research. The general conclusion is that all examined human-like, animal-like and machine-like traits all have different strengths and weaknesses, dependent on personal differences, context and domain of use.

Keywords: robot design, robot appearance, human-robot interaction, HRI, human-like, animal-like, machine-like, likeability

1. Introduction

1.1 The rise of human-robot interactions and the current state of affairs

While exploring the world of scientific research, one could note a continuous growing interest in robots. Since several decades, the field of robotics is a popular field, producing modern robots bearing new functions, new designs and new technology. The production phase in robotics is often taken for granted by anyone but programming scientists, and the amount of elaborate work that goes into developing robots is often underestimated. Even more so, commercially available robots get used for many different purposes, while there might be a better fitting robot for the same goal. An example of this is the robot NAO, a small and basic humanoid robot that is used for research in many departments, as well as in education and in healthcare, for almost every age group and context.

Robots are introduced in increasingly more domains, and they are used more and more in context with humans, or working together with humans. Naturally, the number of interactions between humans and robots increases, calling for the importance of the research field for HRI.

Along with the increase in interactions between humans and robots, people are slowly getting more used to robots. Robots are becoming less intimidating: they are being adjusted to fit the specific user's needs, and consumers are gradually shifting from the idea of "how should *we* change to live together with robots", towards "how should *robots* be shaped to completely fulfill humans' wants and needs" (Eyssel et al., 2011; Sciutti et al., 2018). Since people are getting more used to robots in their daily life, the idea of having robots in one's surroundings is getting normalized (Mende et al., 2019). Because of this, robots gain in interest on the consumer market with both manufacturers and consumers (Mende et al., 2019; Richards & Smart, 2016). More specifically, the marketing industry is examining the potential of house robots, robots that are used in public places, and industrial robots (Karabegović et al., 2011). This leads to almost everyone coming in contact with robots at some point, which is why research on the subject is important.

Accompanying this, it is important to note that every design has different effects and uses for application. This means that robots should have a different appearance for each goal, dependent on their domain of use (Shibata, 2004). For example, a logical assumption could be that robots used in elderly care should have a likeable or even huggable physique, like the seal-like robot PARO (Petersen et al., 2017). On the other hand, robots used in industrial context should look sturdy and polished to perform work efficiently, and their internal computer software is more important than their looks (Kjellsson et al., 2007). Despite the present knowledge that robots should have an appearance that fits their goal, new types of robots are introduced, without thoroughly scientifically supporting every aspect of their design (Caudwell et al., 2019).

Ensuring the development of robots gains adequate attention in science, this work aims to give attention to the thought process and design phase of robots, highlighting the challenges and important domains that are often overlooked. With this goal in mind, this work particularly approaches the effects that the design of robotics has on humans in human-robot interaction (HRI), as different appearance elements can lead to different effects. The importance of HRI is getting noted, as stated by Zheng et al. (2020), the field of HRI is currently giving more attention towards affect and embodiment of robots, concluding with emphasis on the need of more research towards this subject of human emotion. It is important that effects of design are acknowledged in the early stages of robot design, so that the final product is appropriate for the desired goal.

The production of robots designed for HRI is seen as a large, multidisciplinary field with many factors that influence the final product. The complex design of robots calls for an assembly of three disciplines: engineering, design, and psychology (Bartneck et al., 2020). When experts in these disciplines adequately work together, the discipline of engineering covers technical challenges, the discipline of design faces challenges related to appearance and the robots' behavior, while the discipline of psychology accounts for challenges concerning human affect in interactions, cognition and social agents. However, HRI scientists often carry out the role of expert in several, or even all domains of the production process (Bartneck et al., 2020).

Implementing each discipline correctly is important, as producing and formalizing robots is a delicate field since separate design features have different (possibly unwanted) effects. To illustrate the interplay of the three disciplines, Löffler et al. (2020) state that making a robot look like a human can have positive effects such as providing comfort and a pleasant conversation, but it can also cause negative effects, such as the expectation that said robot will behave exactly like a human. There are even already signs of the commercial robot market experiencing a decrease in sales, due to robots setting a certain expectation through their appearance, but not meeting that expectation task-wise (Caudwell et al., 2019; Fernaeus et al., 2010), pointing towards the importance of including all three disciplines of robot design. This example shows an effect of robot design on human psychology, while the practical core remains to be programming and engineering. The more fitting robots can be to a human's expectations, the more interesting social robotics will remain on the consumer market.

Along with these expectations, appearance of robots is important to establish a positive interaction. When humans interact with robots, they base their expectations on the first impression of the appearance; this means people can be steered by a design into expecting relevant things or desired tasks they would like the robot to do (Woods, 2006). Not to be overlooked, humans have a tendency to have strong emotions and attitudes towards robots, based on the robot's appearance (Nomura et al., 2006; Joinson, 2002), indicating that small changes in robot appearance can have a big impact.

The production of robots is mostly done while only using the discipline of engineering, and principles of design and psychology are often overlooked (Bartneck et al., 2020). Regarding the importance of adding psychology in the field of HRI, every human has different expectations and wishes, and responds differently to robots. There is little research principally emphasizing psychology in relation to robots, even though more and more humans interact with robots with these varying expectations on a daily basis (think of robotic vacuum cleaners, Smart Home Assistants, self-driving cars, and virtual trainers in gyms, among many others [Bartneck et al., 2009]). Therefore, how will engineers manage to design an appropriately behaving robot for certain situations, while not taking into account human expectations?

As a designer and as an engineer, a sufficient understanding of human psychology is needed to include and anticipate human responses in a human-robot interaction (Bartneck et al., 2020).The combination of these assumptions (the growing number of different types of robots; and effects of design features on humans regarding a robot's domain of use), forms the basis of the exploratory nature of this work: which is to examine the effects of different physical appearance aspects of robots, on their effectiveness in different domains and how humans will interact with them.

1.2 Relevance to New Media Design

As stated before, the three disciplines in the production of robots, engineering, design and psychology are equally important (Bartneck et al., 2020). This literature review will cover both psychology aspects and design aspects, since I approach the problem in context of the Master's program New Media Design. New Media Design covers both these domains in regard to technology, covering the relationship of humans (users) with modern technology.

Based on these statements and earlier findings, I believe that the domains of psychology and design solicit for more attention. Accordingly supporting this statement, Zheng et al. (2020) imply that the field of HRI is only recently focusing on the effect that robot appearances have on the emotion of the human interacting with the subject. Furthermore, research on appearance specifically is necessary, since it is often not tested alone: in many studies, it is immediately paired with either gestures, or speech, or other types of behaviors. Examples of this are the study by Ludewig et al. (2012) evaluating the robot TOOMAS, testing a verbal robot condition versus a nonverbal robot condition; and the study by Ahmad et al. (2016) evaluating the robot NAO based on gestures.

1.3 Approach to categorizing robot appearance

To properly analyze effects of robot appearances, a certain form of categorization is needed to systematically examine the subject. There are multiple ways that different researchers use to categorize robots: usually categorization is based on looks, or on tasks. First of all, based on looks, there are degrees in life-likeness: Shibata (2004) categorizes in human-likeness and animal-likeness. Human-like robots range from plain humanoid to extremely realistic, and animal-like robots can be based on familiar animals versus unfamiliar animals, ranging from plain to life-like. Aside from life-likeness, Löffler et al. (2020) categorize machine-like robots next to human-like and animal-like robots: machine-like robots are categorized as mechanomorphic, since they primarily look like machines. These robots might have one or a few recognizable features, but they generally still look like an object and are perceived as one. Fong et al. (2003) classify four types of robots: anthropomorphic, zoomorphic, caricatured, and functional.

Based on this literature, I categorize robots by specifying their looks, and zoom in on the appearance features for each group. Regarding terminology, animal-like and zoomorphic robots are considered similar things (Tzafestas, 2016), as well as human-like and humanoid (Mende et al., 2019), and mechanomorphic, functional, and machine-like (Fong et al., 2003; Löffler et al., 2020).

The categorization of what a robot looks like is from here on referred to as appearance, morphology or embodiment, since these terms cover the meaning of a robot’s physical form or body, shape, size and color (Fong et al., 2003; Zheng et al., 2020). To reiterate, this work will focus on the effects that human-like, animal-like and machine-like robot embodiments have on human affect. While keeping in mind these categories and terminology, there is no universal definition of exactly which traits fall under which category. A robot’s appearance is mostly evaluated as a whole, instead of as a combination of traits. As I am interested in the specific effect of individual traits, traits will be specified while using the categorization of human-like, animal-like and machine-like. Appearance traits will be allocated to the congruent category, based on the results. By specifying traits as human-like, animal-like and machine-like, robots can be evaluated on both broad categories, as on specific traits.

Next to this categorization based on looks, robots can be sorted into different domains of use. Following Baraka et al. (2019), a categorization is made to specify these exact domains that robots are used in. The domains used for social robots are healthcare and therapy (children, elderly, or general), industry, education and entertainment, home and workplace, search and rescue, public service, and social sciences. This work aims to include robots of all seven categories, while the overarching categorization will be human-like, animal-like, and machine-like.

Including appearance categories, specific appearance traits, and broad domains of use in this research will provide insights on the effects of appearance traits on multiple levels: in certain situations, it might be useful to examine the effect of human-like robots compared to animal-like robots, while in other situations, individual traits are more important. Adding both levels of appearance can prove meaningful in future research and adds value to this study. Another important note is that this work focuses on full-body robots, in contrast to studies that only measure effects of robot parts, like robot faces, or robot hands.

This literature review aims to categorize existing literature into an easily understandable summary, where effects of appearance traits that have been examined by other scientists are distributed based on a human-like, animal-like, or machine-like categorization. Ideally this summary is to be used as guidelines for further research, so that robot design can be more effective. The research will be done by an exploratory approach, in the form of a systematic literature review. I aim to create guidelines that can help scientists to make deliberate decisions regarding the design and functionality of new robots, where the desired effects are validated by previous research. The theme for this literature review is confined into robot appearance and the relation of robot appearance to human affect/emotions. Since the subject of human emotion is too broad, this is narrowed down to positive affective effects in humans. In the upcoming section, the effect of the design of robots is linked to how much a human likes the robot they are interacting with, how a first impression of said robot is formed, and how a human’s feelings and emotions are affected by a robot’s appearance: in other words, how the design of robots influences its likeability.

2. Theoretical framework

2.1 Approach to defining likeability

As stated before, HRI includes human affect towards robots concerning interactions. Defined as "any experience of feeling or emotion" (American Psychological Association, n.d.), in this context 'affect' is used to describe the phenomenon that humans experience when they are moved by emotions or feelings.

Likeability is used to conceptualize 'human affect', making it a relevant subject of this thesis. As a key item of HRI, likeability is seen as the effect of a robot's appearance on human affect, leading to a positive feeling and positive first impression in said robot (Bartneck et al., 2009; Löffler et al., 2020). Likeability is such an important topic, because high likeability and positive affect lead to a positive overall evaluation of a robot, which in turn leads to various positive outcomes. Some of these desired outcomes as a result of an appropriate level of likeability are an increase of a human's engagement and usage, and higher level of trust, which are the end goal for many robotics designers (Löffler et al., 2020; Robbins & DeNisi, 1994). This literature review will focus on likeability on its own, without looking at indirect effects of likeability, to keep the scope of this work concise.

2.2 Robots and likeability

Next to likeability, robots can elicit positive effects and improve efficiency of tasks. An example of a domain interested in this positive contribution of robots is the military: research is constantly done on improving military robots, as they are useful for transportation and aid. In practice, research has demonstrated that military robots increase efficiency and decrease workload for soldiers (Voth, 2004). Recently, culture and emotions of soldiers are acknowledged often in military studies (Carpenter, 2016; Jentsch, 2016). However, in practice, soldiers often dislike and do not accept military robots, leading to situations that can severely damage both the robot and its environment (Desai, 2009). This introduces the importance of the need of robot acceptance, and more importantly research on how to establish liking and acceptance of robots.

This example emphasizes the usefulness of researching a human's likeability towards robots, for a robot to be truly efficient and effective. Without inducing positive affect in humans, these robots will never be truly beneficial in the domains in which they have potential, because humans will not accept them, or want to interact with them. After all, when robots are designed adequately, they can prove helpful to solve shortage in certain professions. For example, Matsui and Yamada (2019) show that electronic teachers can be effective in several fields: for example, robot teachers are especially effective in teaching a technical domain such as science or programming. Furthermore, Heerink et al. (2010) show that social robots can be a viable addition to elderly care too, provided that the elderly show high technology acceptance. This is important, since the elderly population is growing and a shortage of staff is a problem in several nursing homes (Heerink et al., 2010). Next to robotic companions for the elderly to reduce loneliness (McGlynn et al., 2017; Tan et al., 2013), social robots like the robot nurse RIBA (Dinet & Vivian, 2014) can be a viable addition to nursing homes to reduce pressure from human nurses.

For the domains used in these examples, it is evident that a robot should be likeable to be effective. A robot can be made likeable by changing its appearance: Li et al. (2010) show that a robot's appearance affects its likeability, with strong effects on trust and satisfaction. This positive effect of appearance was found in the educational domain, personal guidance, entertainment and security (Li et al., 2010). Attractive robots are preferred in everyday life and for short interactions: if people like the appearance of a robot, it is more likely that they will engage with the robot and start an interaction (Heuer, 2019).

2.3 Sources to form the construct likeability

In general, the use of everyday life robots is intended to be a positive experience, creating positive affect in humans. Likeability is relevant to both design and psychology, since it is about the effect of a robot's appearance on a human's state of mind. However, there is no universal definition for likeability. To refine likeability, definitions and previous uses in existing literature are brought in.

To measure likeability, several measurements and scales are used by a variety of researchers, resulting in likeability consisting of several subscales. These subscales are not always exactly the same; in this section the most commonly used scales are named and combined into one overarching subset to define likeability. Furthermore, the appointed measurement instruments and their relevant items are used to determine if a study properly studied the construct of likeability that is used in this work.

The first reference for defining likeability is the validated scale used by Bartneck et al. (2009): the Godspeed scale. This scale includes measurements to determine the level of anthropomorphism, animacy, likeability, perceived intelligence and perceived safety in a robot. It is suggested by Bartneck et al. (2009) to use the Godspeed scale when assessing and supporting design choices in robots with a psychological nature. The Godspeed scale has a subset for likeability with relevant items for this thesis, so studies using this scale as a measurement tool will be included in this study. Likeability is measured in the Godspeed scale with liking, friendliness, kindness, pleasantness, niceness, and relaxedness. These six items are measured on a 5-point semantic differential scale, and their scores are taken into account when determining the likeability of the robot used in relevant studies.

A second commonly used scale to measure liking and disliking towards robots, is the Negative Attitudes towards Robots Scale (NARS). This scale, first introduced by Nomura et al. (2006) and supported by other researchers (e.g. Fraune et al., 2015; Halpern & Katz, 2012; Syrdal et al., 2009), measures people's negative thoughts, perceptions and expectations they have of robots for short-term, or long-term interaction. Found in several studies is a negative relationship between anxiety and likeability, as well as anxiety and ease of use (Heerink, 2010; Louie et al., 2014), thus anxiety is considered as predictor for likeability. The NARS consists of three subscales: a subscale measuring negative attitudes towards interaction with robots; a subscale measuring anxiety towards a robot's social influence; and a subscale measuring emotion. The items used in the subscale 'emotion' are feelings of relaxation, anxiety, friendliness and wanting to befriend (companionship). Since these items for emotion concern affect, and semantically overlap with the items and subscales for other scales that measure likeability, they are included in this work's definition for likeability. Studies using scales or subscales from the NARS, including companionship, friendliness, relaxedness and anxiety, are included in this study. The scores for these subscales are then taken into account to create a measure for likeability.

The third source consulted for defining likeability is the Technology Acceptance Model ([TAM], first introduced by Davis [1989]). The TAM suggests that there are four factors that determine technology use (where technology can be interpreted as robots): perceived usefulness, perceived ease of use, attitude toward technology and intention to use. Perceived ease of use is the relevant construct for likeability: Heerink et al. (2010) present subscales for this construct as anxiety, perceived enjoyment, and perceived sociability. The specific items in questionnaires for the subscale perceived enjoyment are enjoyment of the interaction, fascination, and boredom (Heerink et al., 2008). These items are grouped together as enjoyment for this study since they measure the same concept. The subscale perceived sociability is measured by pleasantness of the interaction, the perception of a robot as a companion or friend, and niceness (Heerink et al., 2010). In conclusion, this means that studies using the TAM are included, with measurements for enjoyment, and perceived sociability, with the measurement items being enjoyment, companionship, pleasantness, and niceness.

A fourth validated scale to measure likeability, is the Reysen likeability scale (Reysen, 2005). This scale measures likeability in general, and it includes questions regarding friendliness, warmth, approachability, liking, companionship, attractiveness, and similarity (Reysen, 2005). The overlapping and similar items for likeability with other scales are used for the definition of likeability in this thesis, which are friendliness, liking, companionship, and attractiveness. Studies using the Reysen likeability scale are considered for this literature review, and the relevant items are included in the results.

The fifth validated scale to measure affect, is the Positive Affect and Negative Affect Scale (PANAS), as used by Watson et al. (1988). Positive affect and negative affect have various different effects on a human's state of mind, and account for impressions and evaluations of humans (Watson et al., 1988). Since this is overlapping with the concept of likeability, studies using the PANAS as a measurement instrument are included in this literature review. The relevant items to define likeability from positive affect in the PANAS is measured with enjoyment and niceness, negative affect is measured with anxiety.

2.4 Additional sources

Resulting from the Godspeed scale, NARS, TAM, Reysen likeability scale and PANAS, the concept of likeability used in this thesis consists of liking, friendliness, kindness, pleasantness, niceness, relaxedness, companionship, attractiveness, anxiety and enjoyment. To further support the inclusion of these specific items in the definition for likeability, other sources using a similar questionnaire for measuring likeability are presented here. The items discussed here, have overlap with, or are seen as a subcategory for the ten items that construct likeability, to support and validate the use of these items. A final model to illustrate the concept of likeability which includes all relevant items as discussed, is presented at the end of this chapter in Figure 1.

Salem et al. (2013) state that there are two validated items for likeability: politeness and sympathy. Studies using these items are included in this review, as these items are similar in meaning to kindness and friendliness respectively. Löffler et al. (2020) base their ratings of likeability on wanting to be the object’s friend, friendliness, and an absence of creepiness, which was based on work by Shibata (2004). Absence of creepiness is regarded similar to low levels of anxiety, and thus studies using this term are included too. Shibata (2004) used questionnaires regarding social robotics with items measuring subjective evaluation, including cuteness, friendliness, pleasantness, and liking. These items are similar to the items of the NARS, where cuteness is regarded similar to attractiveness (Kato, 2006). Furthermore, certain items from the “human nature” ratings by Haslam et al. (2008) are seen as relevant to define likeability. They define positive traits when attributing likeability to others, including warmth, friendliness, sociableness, and fun, similar to the definition of pleasantness, friendliness, companionship, and enjoyment respectively, as used by Eyssel et al. (2011).

2.5 Acceptance correlating with likeability

One sidenote regarding likeability is the concept of acceptance. Technology acceptance is an important determinant of product usage: in order for consumers and experts to use robotics in any domain or any context, acceptance should be established (McCloskey, 2006). After all, acceptance of the technology in question is a significant factor in determining its success or failure (Davis, 1991). To induce acceptance and to increase intention of use of robots, positive feelings in humans must be evoked through a robot's design (Davis, 1991; Teo, 2011). It is seen here that establishing acceptance is similar to establishing likeability, illustrating that likeability and acceptance are very closely related. Even more so, content likeability can lead to an increase in product acceptance, for example with robots for use at home, and robots used in exercise therapies for the elderly (e.g. Cruz-Sandoval et al., 2018; de Graaf et al., 2019; Lai & Liu, 2020). The correlation between likeability and acceptance is seen in multiple age groups: increased likeability can lead to better first impressions, intention of use and ultimately acceptance for the elderly (Prakash et al., 2014); and to more fun, acceptance and an increased motivation to play again for children (Logan et al., 2019).

Acceptance is necessary and beneficial to a robot's efficiency: introduced in the work of Ventre-Dominey et al. (2019) and Louie et al. (2014), is the Robot Acceptance Questionnaire (RAQ), measuring respondents' positive attitudes, and levels of interest, pleasantness, niceness, trust and anxiety towards a robot. The RAQ thus measures positive affect, as a predictor of robot acceptance. These items are similar to items that measure likeability. The other way around, most widely used scales for likeability also include acceptance, further supporting the claim that likeability and acceptance are very closely linked (Bartneck et al., 2009; Heerink et al., 2008; Heerink et al., 2010; Nomura et al., 2006).

Finally, de Graaf et al. (2019), conducted a large-scale questionnaire (N = 1162) measuring acceptance, where the questionnaire was based multiple validated scales: relevant terms used were enjoyment, attractiveness, and companionship. These terms are derived are proven to be effective in measuring acceptance and in its turn likeability, as the items are the same.

Consequently, the relation between acceptance and likeability is too apparent to ignore: the constructs are overlapping in such a way, that both constructs sometimes get measured by the same items. While not focusing on it, acceptance is a big construct that not only could be a consequence of likeability, but also an influence of it and a precedent for it. Studies measuring both concepts might be taken into account in this study, but studies measuring only acceptance with the items discussed are not taken into account, since the focus will remain on likeability.

2.6 Conclusion definition likeability

In conclusion, for this thesis the effect of a robot's appearance on likeability is measured: the ability of a robot to evoke positive emotion or affect in a human, which ultimately directly influences the evaluations of the robots. As found in literature stated above, likeability is correlated with and influenced by: liking, friendliness, kindness, relaxedness, niceness, companionship (likely to/wanting to befriend), pleasantness, enjoyment, attractiveness, and anxiety. So again, studies that measure one of these items are added into the literature review, where they will be used as categorization for likeability traits and effects.

Furthermore, since acceptance is so closely intertwined with likeability, a positive correlation cannot be ignored (Ventre-Dominey et al., 2019): acceptance is influenced by and correlates with companionship, pleasantness, enjoyment, attractiveness, and anxiety. Robot acceptance is not further examined; however, some overlap can be found in the relevant studies considered for likeability.

2.7 Research questions

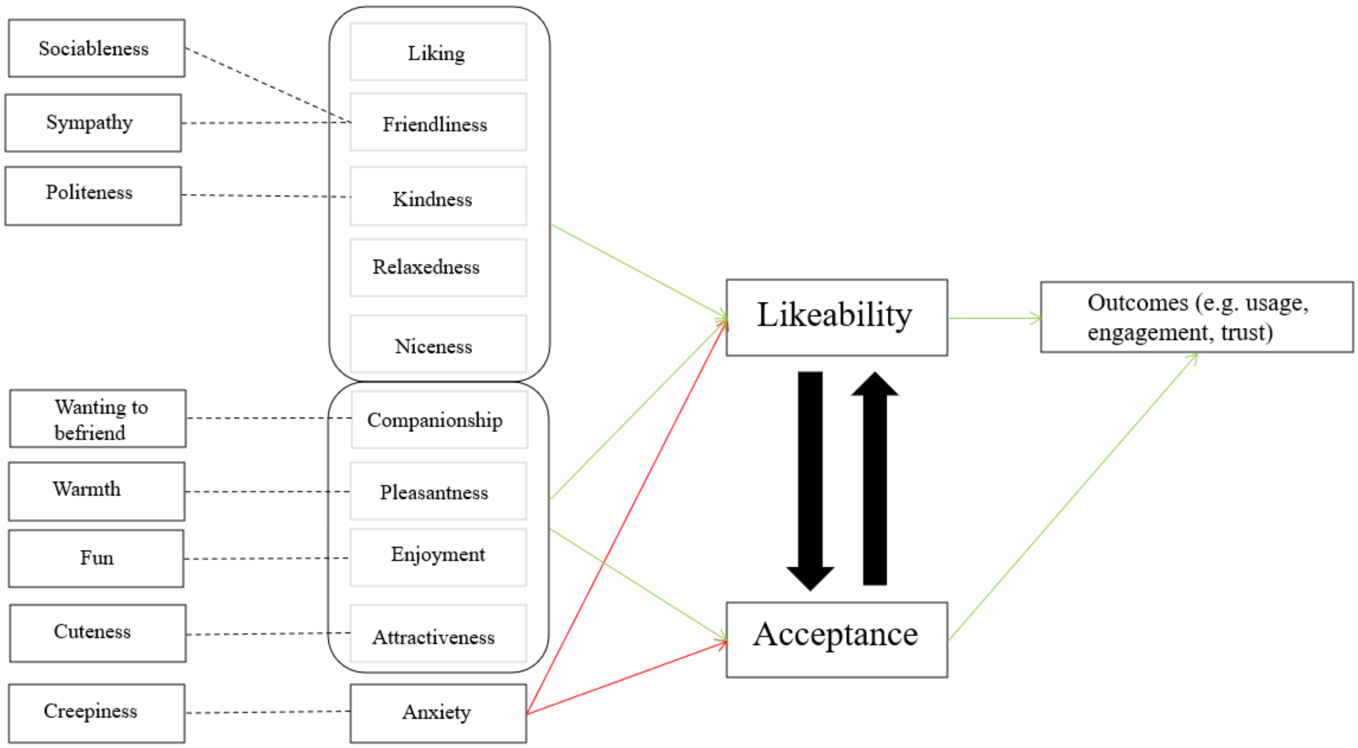
To conclude: likeability and acceptance in the context of technology are positively correlated in a two-way relationship, where high scores lead to higher positive outcomes (Figure 1), and likeability is constructed by ten items. With this theoretical base, I present the research questions:

RQ1: Which appearance traits are specific for human-like, animal-like and machine-like robots?

RQ2: What specific effect does each of these traits have on the overall likeability rating of a robot's design?

Figure 1

Model of the effect and inclusion of correlations of likeability



Note. Each line represents a relation to another concept. Green lines indicate positive correlations and red lines indicate negative correlations. Items on the far left are regarded as similar items as the subscales they connect with (dotted lines).

3. Method

3.1 Design

To gather relevant research regarding the effect of appearance on likeability, a systematic approach was applied, using several databases. The structure of this section is based on the Preferred Reporting Items for Systematic Reviews checklist (Liberati et al., 2009). To reiterate the inclusion criteria and state the boundaries of the literature assessed in this work, the previously determined requirements are studies including at least one human-like, animal-like and/or machine-like robot; studies that examine affective effects of the robots' full-body appearances; studies that use the discussed measurements or that include items from those measurements; and studies written in English.

3.2 Procedure

First, a list of search queries has been set up (Table 1), to search for literature that should be included in this work. Searched for are the studies regarding the relevant categories of robots (human-like, animal-like, machine-like) and studies examining robots from any of the domains healthcare and therapy, industry, education and entertainment, home and workplace, search and rescue, public service, and social sciences. To keep the scope concise, the search criteria include the terms likeability, liking and affect (affection), to ensure that the focus lies on these terms. Search results are later assessed on relevance.

No distinction is made in demographics (age groups, gender, or country of origin) in advance of the literature selection process, since the aim of this study is to explore, collect and summarize affective effects in every demographic and domain that has been studied, to map all relevant previously found effects. Following the exploratory nature of this thesis, these subcriteria can lead to different effects, regarding certain contexts, so the relevant effects will be presented after data collection. Furthermore, all years until March 2020 have been considered, since the field of robotics has been rapidly growing for the past few decades, so I aim to visualize that growth and the shift in studies including likeability, design and psychology into robotics.

3.3 Information sources

The search queries from Table 1 have been applied in the following databases: WorldCat, ACM DL, Web of Science and PsycINFO (Ebscohost). Search queries with no results have been deleted from this initial list, meaning the literature used has been found with one of these search terms. Mendeley is used to import the data as a tool to optimize the selection process. Through reading the items' titles, a number of irrelevant items are deleted. The deleted items contain errata, prefaces, reviews, incomplete objects, and other miscellaneous irrelevant items. The subsequent step is to select articles on relevancy through examining the abstracts of the items. Thereafter, articles are scanned on relevancy to likeability and appearance. If the remaining dataset is n > 100, another narrower search will be done with more strict inclusion criteria.

Table 1

Search queries used in the databases

Category	Search query
Likeability	Robot AND human-like AND likeability
	Robot AND humanoid AND likeability
	Robot AND animal AND likeability
	Robot AND animal-like AND likeability
	Robot AND zoomorphic AND likeability
	Robot AND mechanical AND likeability
	Robot AND mechanical-looking AND likeability
	Robot AND industrial AND likeability
Affect	Robot AND human-like AND affect
	Robot AND humanoid AND affect
	Robot AND animal AND affect
	Robot AND animal-like AND affect
	Robot AND mechanical-looking AND affect
	Robot AND mechanical AND affect
	Robot AND industrial AND affect
Liking	Robot AND human-like AND liking
	Robot AND humanoid AND liking
	Robot AND animal AND liking
	Robot AND zoomorphic AND liking
	Robot AND mechanical AND liking
	Robot AND mechanical-looking AND liking
	Robot AND industrial AND liking
Human-robot [interaction]	Human-robot AND likeability
	Human-robot AND liking
	Human-robot AND affect
Morphology	Robot AND morphology AND likeability
	Robot AND morphology AND liking
	Robot AND appearance AND likeability
	Robot AND appearance AND liking
	Robot AND design AND likeability
	Robot AND design AND liking
HRI and social robotics	HRI AND morphology AND likeability
	HRI AND animal AND likeability
	HRI AND design AND likeability
	Social AND robotics AND likeability
	Social AND robotics AND liking
Affection	Robot AND human-like AND affection
	Robot AND humanoid AND affection
	Robot AND animal AND affection
	Robot AND mechanical AND affection

Note. Search queries resulting in no results have been removed from this list.

3.4 Data structurization

Included in the dataset are studies and models using measurements scale for likeability, such as the Godspeed scale, NARS, TAM, Reysen's scale and PANAS. Results are summarized in the results section on effect: which effect is found in what context. The context is divided in the categories age, country, domain, and gender. Furthermore, in order to base guidelines for future research on the found results, the robots used in relevant studies are sorted and coded according to their appearance, so that certain deductions can be made based on specific exterior traits.

4. Results

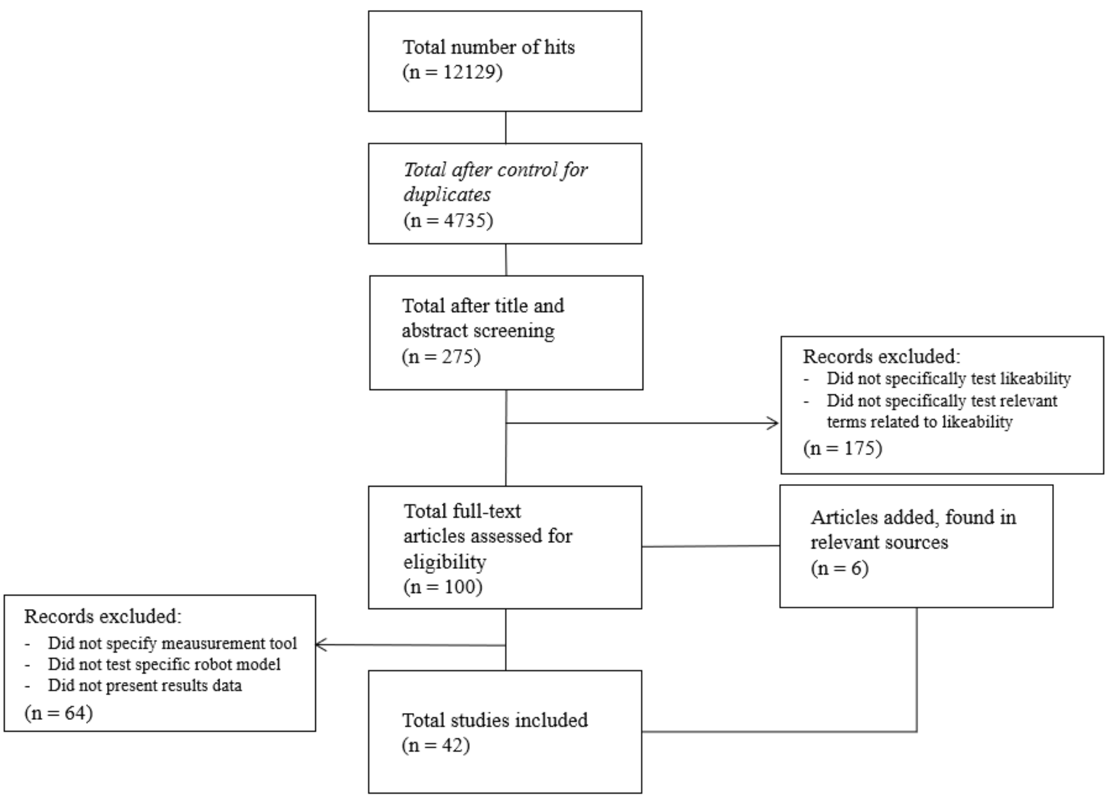
4.1 Results of search queries

A search through the databases resulted in n = 12129 records. Subsequently, duplicates were removed, leaving a total amount of n = 4735 articles. Worth noting, the fact that after controlling for these redundant articles, roughly two-thirds of the articles initially found are omitted, signals that there was a considerable amount of overlap in the databases used. Next, scanning through the content, a number of n = 294 irrelevant hits were deleted, being several objects that did not comply to the search criteria, such as articles written in a language other than English, prefaces, corrections and errata. After deleting the irrelevant items, the total number of articles to scan through resulted in n = 4441 articles. The subsequent step was to scan the articles on relevancy through examining the abstracts of the items.

Abstracts were selected on having found any kind of affect, likeability or acceptance effect directly related to a physical trait. By personal judgment, this led to a new set of relevant articles, with a total number of n = 275. As this indicates a too broad research area, the inclusion criteria were adjusted to any (positive/negative) affective likeability effect involving emotions or feelings, directly induced by an appearance/physical trait, leaving out articles on acceptance, and articles that only measure affect independently from likeability. Having in mind these new inclusion criteria, the remaining n = 275 abstracts were judged more thoroughly on their relevance. This sorting method resulted in n = 100 articles. Since this is yet too broad, a final selection has been done by carefully reading articles to critically leave out studies that included gaze or behavior on top of appearance, as well as strictly reviewing methods, measurements, and results, ensuring any likeability trait is specifically tested on a specific robot appearance with a validated measurement, with minimal confounding variables. This left out articles that, on closer inspection, did not comply with the inclusion criteria because of confounding variables for appearance traits, and studies not testing appearance separately but only in combination with behavior. This final process resulted in a final number of n = 42 of articles, which are presented thematically in Appendix A. The implemented search method is shown in Figure 2.

Figure 2

Flowchart of study selection and inclusion criteria process



Appendix A includes all relevant information, with likeability categorized by the subscales presented in Figure 1, and categorizing robots used by human-like, animal-like or machine-like. As an additional mention, some studies use a single question in their measurements, for example “how likeable is this robot?”. For this reason, likeability itself is added in Appendix A as well.

4.2 Categorization of results

To properly analyze and study the results and effects found in the literature, a certain form of categorization is needed. The goal of this work is, after all, to present useful guidelines for future research, where categorization is based on previously validated measurements. As stated in the introduction, the categorization used to specify robots, is determined as human-like, animal-like and machine-like. To appropriately implement the results found, the categories are divided into specific appearance traits, to further identify which trait is correlated with which effect.

To specify which traits fall under which categories, previously published literature is consulted. A validated measure for determining human-like traits in robots, is the ABOT Database (Phillips et al., 2018). The study prior to this database presents a standardized measure on human-like traits, and some determinants for nonhuman-likeness. The greatest significant appearance determinants for human-likeness are eyelids, head hair, skin (e.g. rubber-like, plastic), genderedness, a nose, eyebrows, and apparel. Research by Fong et al. (2003) supports the use of these traits, presenting a mouth and eyes as significant for life-like traits, and eyebrows for human-like traits, as significant features.

Features that make a robot animal-like, are based on Löffler et al. (2020). The highest significant predicting traits for life-likeness (human-likeness and animal-likeness combined) are a face, eyes, a head, a mouth, pupils, and non-childlike characteristics (natural proportions). Traits determined as animal-specific are a snout, a tail, ears, familiarity (based on animals or well-known fantasy creatures [e.g. dragons]), legs/paws, flexibility and claws (Löffler et al., 2020). Familiarity is seen as readily available knowledge of a robot resembling a specific animal, for example a seal pup has fur while an adult seal does not have fur (Löffler et al., 2020; Shibata, 2004). This is supported by Nakata et al. (1999), stating that the biggest predictor of an animal-like robot is its resemblance with a real animal. Noteworthy is that the trait fur is not a predictor of animal-likeness, since there are of course animals that do not have fur. Instead of fur, some form of skin appears to be more important for life-likeness.

Following the ABOT database for nonhuman-like traits, the greatest significant determinants for machine-likeness are wheels and treads/tracks (Phillips et al., 2018). Extending these traits with the study by Löffler et al. (2020), are traits that are negative predictors for human-likeness and animal-likeness. These negative predictors define robots as more like machines, than living creatures: these traits signal a lifelessness, meaning they can be put directly against life-likeness (Löffler et al., 2020). These lifeless traits are considered open parts, absence of head, absence of mouth, and disproportionate body parts, which are all included in the machine-like category. When a robot includes more life-like traits than lifeless traits, it is considered more of a living creature than an inanimate object, and vice versa. In conclusion, traits that predict machine-likeness, are considered wheels, treads/tracks (moving bands for locomotion), open parts (absence of skin, open technical parts, no smooth surface), absence of a head, absence of mouth, disproportionate body parts, and metal surface.

The appearance features for each category are presented in Table 2. These traits are applied to the robots found in the relevant literature, to categorize the robot's morphology. To do so, an image database is composed from each robot model found in the literature in Appendix A. This image database is presented in Appendix B, and trait codes have been assigned to each robot, based on their appearance. The robots used in the studies presented in Appendix A, are then allocated to the corresponding category for morphology, after establishing their traits based on Table 2.

Table 2

Human-like, animal-like, machine-like coding system to apply on robots found in literature

Life-likeness predictors	Code	Human-like predictors	Code	Animal-like predictors	Code	Machine-like predictors	Code
Head	LL1	Eyelids	HL1	Snout	AL1	Wheels	ML1
Mouth	LL2	Head hair	HL2	Tail	AL2	Treads/tracks	ML2
Pupils	LL3	Skin	HL3	External ears	AL3	Open parts	ML3
Proportionate body parts	LL4	Genderedness	HL4	Familiarity	AL4	Absence of head	ML4
		Nose	HL5	Legs/paws	AL5	Absence of mouth	ML5
		Eyebrows	HL6	Claws/front paws	AL6	Disproportionate body parts	ML6
		Apparel	HL7	Flexibility	AL7	Metal surface	ML7

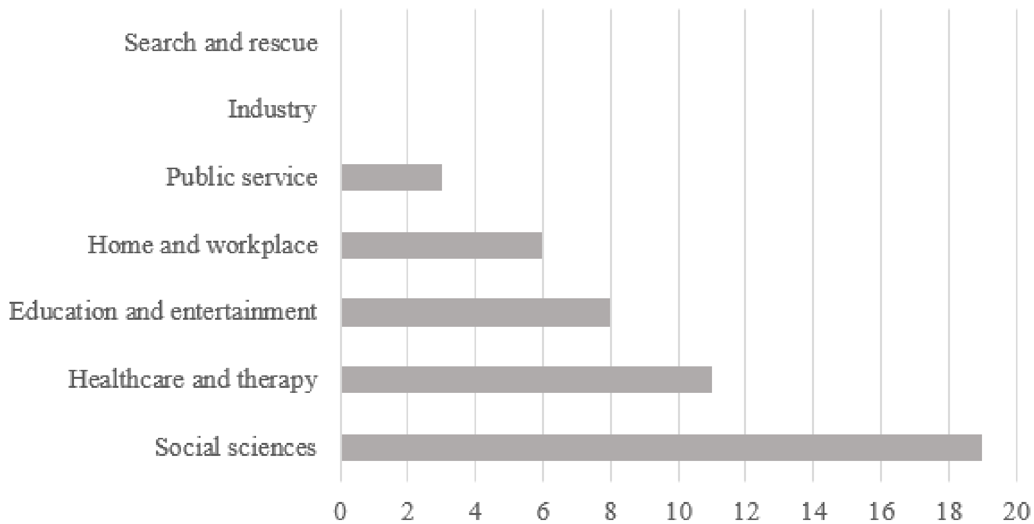
Note. All definitions for body parts and traits are based on Phillips et al. (2018); Baraka et al. (2019); and Löffler et al. (2020).

As noted in the introduction, the categorization of the domains for the social robots in Appendix A is based on Baraka et al. (2019). To recapitulate, the domains used for social robots are healthcare and therapy (children, elderly, or general) (H&T); industry (I); education and entertainment (E&E); home and workplace (H&W); search and rescue (S&R); public service (PS); and social sciences (lab-setting and online questionnaires) (SS). These domains can overlap, in which case all relevant domains are specified, and some studies fall under two domains.

In Figure 3, the total number of articles that studied likeability is listed, sorted on the domain of use. Since no articles on likeability in the domains search and rescue and industry have been found, they are omitted from the following analyzation of results.

Figure 3

Distribution of articles that studied likeability per domain



4.3 Analysis of results

Following Appendix A and Appendix B, for each robot appearance, applicable appearance trait codes are allocated to categorize and code the robot's morphology, so that specific appearance effects can be derived. This information will be used to analyze the results found.

4.3.1 Demographics and statistics

In Figure 4 a distribution is made of the number of articles by publication year. The year 2020 is not included in the graph, since only articles until March 2020 are added into the literature review, making 2020 incomplete as a year of relevant articles. In Figure 4, an increase can be seen of articles published on robot appearance in relation to likeability.

In Figure 5, the distribution of published articles around the world is presented. Noteworthy is that the topic is being studied worldwide, however not yet in every continent, but merely in the northern hemisphere, with peaks in North America and Germany.

Figure 4

Distribution of articles by publication year

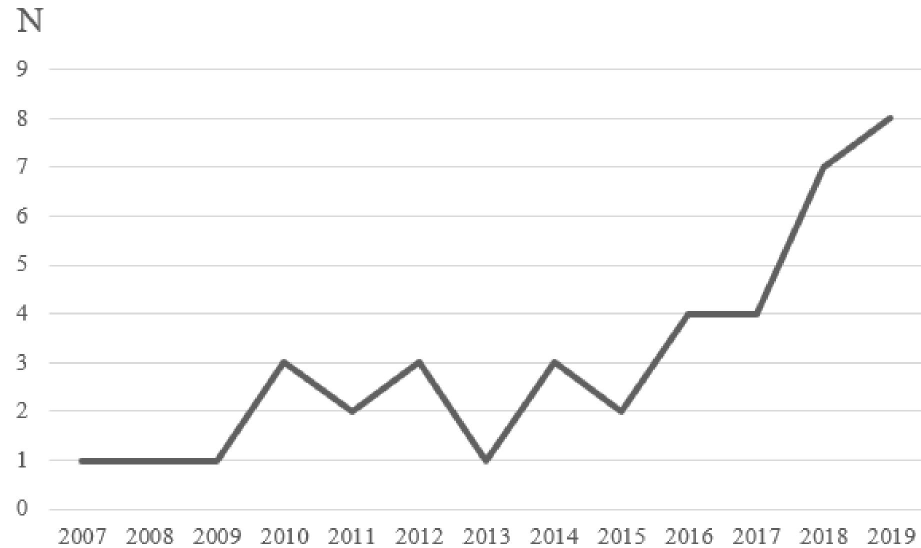
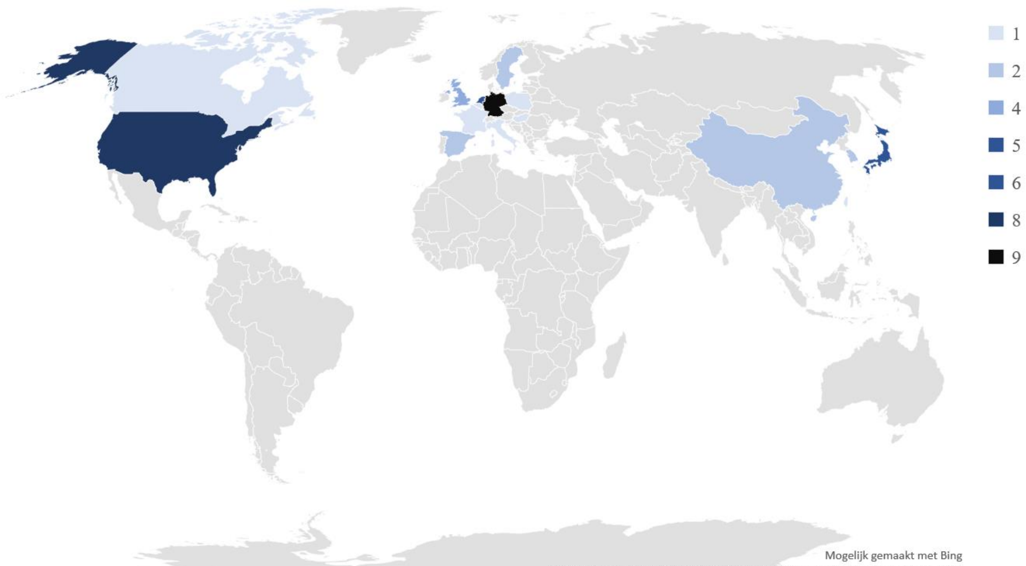


Figure 5

Worldwide distribution of publications



4.3.2 Likeability effects

With the categorization of robots by appearance traits in Appendix B, the results from this literature review can be analyzed more in-depth. To do so, an overview is made of each occurrence of specific appearance traits, classified as significant or not significant. This distribution is presented in Table 3. To elaborate on how Table 3 can be interpreted, the trait ‘snout’ is presented as an example. The trait ‘snout’ (AL1) has been tested five times in total in the domain education and entertainment, where in three of the five occurrences, ‘snout’ had a significantly positive effect on likeability. For the domain education and entertainment, 24 occurrences of traits had a positive effect on likeability in total, of which three were of the trait ‘snout’. This means that, for the domain education and entertainment, ‘snout’ has a relatively big impact on a robot’s likeability, since the majority of occurrences had a positive effect on likeability, and it has been tested several times in multiple studies. Each of the appearance traits can be interpreted in that manner. This illustrates which traits were most effective in increasing a robot’s likeability and which traits are best avoided in certain domains, relative to the studies that have already been conducted.

Important to note is that many of the appearance traits appear in certain fixed combinations, for example, a considerable amount of robots had a life-like appearance with a head, mouth, pupils and proportionate body parts. Since the life-like traits are tested in plain robots such as NAO, and they turned out to have a positive effect on likeability, these traits are considered likeable by themselves. For this reason, not significant appearance traits for life-likeness are not seen as negatively influencing likeability. This results in life-like traits not being added to the total amount of traits tested per domain in the bottom row of Table 3. By adding up the significant and not significant human-like, animal-like and machine-like traits, Table 3 gives an accurate view of the number of times a certain appearance trait has been tested and in how many cases it had a significantly positive effect on likeability.

Using the information from Table 3, Figure 6 shows a distribution of the percentages of traits that have found a positive effect in research of appearance on likeability. Important to note here is that the division and appearance of robots per domain is skewed, since some domains are examined more often than others. The robots and the appearance trait categories have not all been studied in the same amount, meaning this figure has to be interpreted with caution. To elucidate the effect of the three categories of appearance traits, Figure 7 is presented to accompany Figure 6, which states the total number of studies on likeability, with both significant and not significant effects. To interpret both graphs next to each other, it can be seen that, for example in the domain of education and entertainment, animal-like traits are tested less often than human-like traits and machine-like traits (Figure 7), but animal-like traits relatively lead to the most positive effects (Figure 6). Animal-like traits seem to be effective in general, notably compared to the other traits.

Figure 6

Justification of human-like, animal-like, and machine-like traits leading to a positive likeability effect per domain

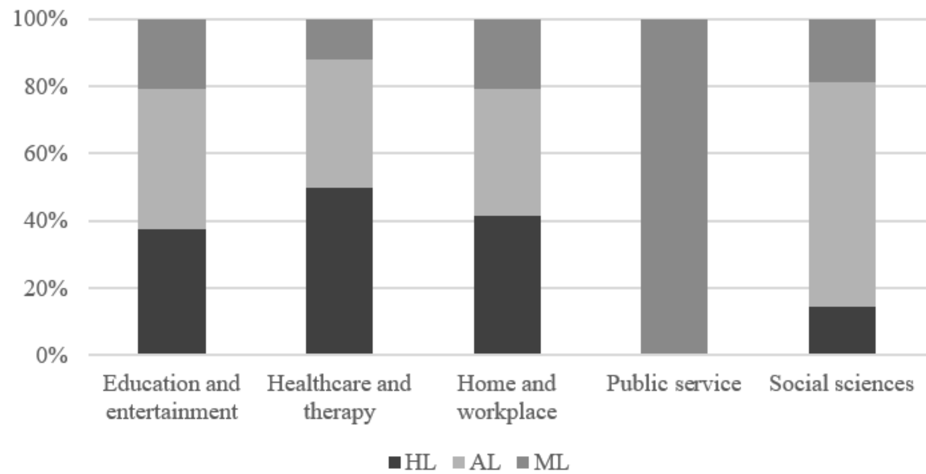


Figure 7

The percentage of traits that have been tested, regardless of the effect (positive, negative, no effect) on likeability per domain.

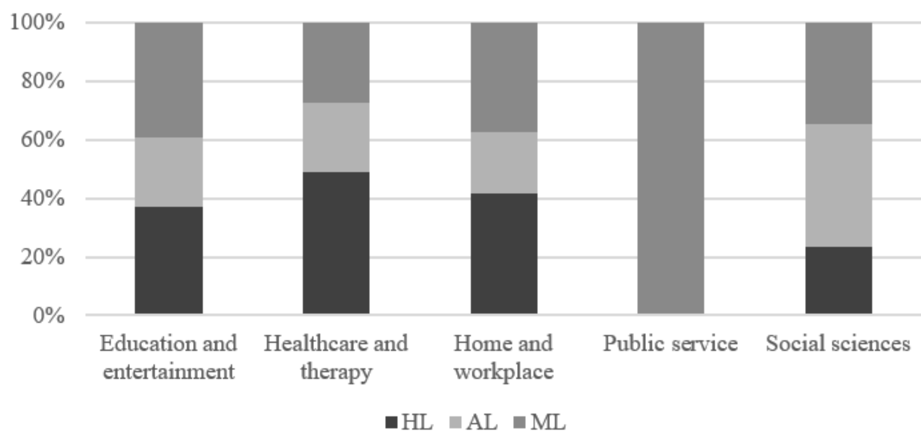


Table 3

Distribution of significant and not significant traits per domain

		Significantly positive					Not significant					Total				
	Domain	E&E	H&T	H&W	PS	SS	E&E	H&T	H&W	PS	SS	E&E	H&T	H&W	PS	SS
Trait	Code															
Head	LL1	9	13	5	3	37	7	10	3	0	40	16	23	8	3	77
Mouth	LL2	7	11	4	3	21	6	7	2	0	31	13	18	6	3	52
Pupils	LL3	8	13	4	3	17	5	5	2	0	26	13	18	6	3	43
Proportionate body	LL4	9	10	3	3	32	6	10	2	0	33	15	20	5	3	65
Eyelids	HL1	2	4	2	0	4	4	4	2	0	13	6	8	4	0	17
Head hair	HL2	1	2	1	0	2	3	3	1	0	6	4	5	2	0	8
Skin	HL3	1	3	3	0	4	4	4	1	0	9	5	7	4	0	13
Gender	HL4	1	4	1	0	3	2	3	1	0	6	3	7	2	0	9
Nose	HL5	1	3	1	0	2	5	4	1	0	8	6	7	2	0	10
Eyebrows	HL6	1	4	1	0	2	4	3	1	0	11	5	7	2	0	13
Apparel	HL7	2	5	1	0	6	2	4	1	0	8	4	9	2	0	14
Snout	AL1	3	4	1	0	19	2	1	0	0	9	5	5	1	0	28
Tail	AL2	1	2	1	0	17	2	1	0	0	5	3	3	1	0	22
External ears	AL3	2	3	1	0	16	2	0	0	0	8	4	3	1	0	24
Familiarity	AL4	1	3	2	0	21	2	1	0	0	7	3	4	2	0	28
Legs/paws	AL5	2	3	2	0	19	1	1	0	0	8	3	4	2	0	27
Claws	AL6	0	1	0	0	1	0	0	0	0	2	0	1	0	0	3
Flexibility	AL7	1	3	2	0	15	2	1	0	0	5	3	4	2	0	20
Wheels	ML1	0	2	2	2	7	1	4	2	0	10	1	6	4	2	17
Treads/tracks	ML2	1	1	0	0	0	2	1	0	0	6	3	2	0	0	6
Open parts	ML3	1	1	1	0	1	4	4	3	0	20	5	5	4	0	21
Absence of head	ML4	0	0	0	0	1	4	2	1	0	10	4	2	1	0	11
Absence of mouth	ML5	0	0	1	0	5	5	4	2	0	18	5	4	3	0	23
Disproportionate body	ML6	0	1	0	0	4	5	2	1	0	12	5	3	1	0	16
Metal surface	ML7	3	1	1	0	12	9	5	2	0	18	12	6	3	1	30
Total mentions per domain (excluding LL)		24	50	24	2	161	65	52	19	0	199	89	102	43	3	360

Figure 6 and Figure 7 should not be compared directly to each other, as both graphs are interpreted differently. Figure 6 illustrates which traits correlate with positive likeability effects per domain, but it does not mean that less common traits are not likeable, since it could also mean that a certain category is not studied enough. Figure 7 shows which traits have been studied at all on likeability, regardless of a positive, negative or no effect. Ideally, the distribution of Figure 7 should be equally divided, for every category should be studied equivalently. If from a graph like Figure 7 the categories appear to be studied an equal amount, then a graph like Figure 6 would provide an accurate explanation of appearance effects for human-like, animal-like and machine-like traits.

Again, the 10 subscales for likeability together (liking, friendliness, kindness, pleasantness, niceness, relaxedness, companionship, attractiveness, low anxiety, and enjoyment), are seen as total likeability. Next, an overview is made of the number of times that each subscale for likeability is measured per domain, to illustrate which subscales appear to be more important for certain domains. This information is presented in Table 4. For example, this table shows that for the domain of healthcare and therapy, liking and companionship seem to be important traits that were focused on in previous studies.

Table 4

Distribution of articles per domain per subscale of likeability

Subscale	Number of occurrences per domain				
	Education and entertainment	Healthcare and therapy	Home and workplace	Public service	Social sciences
Liking	4	9	5	2	12
Friendliness	2	2	1	1	6
Kindness	0	0	0	0	1
Pleasantness	0	2	2	2	2
Niceness	0	1	0	1	0
Relaxedness	0	2	1	0	3
Companionship	2	4	1	0	4
Attractiveness	1	1	1	0	2
Anxiety	0	0	0	1	3
Enjoyment	2	3	1	1	4
Likeability	1	1	0	0	3

Finally, most traits appear in groups: for example, the robot PARO is assumably effective because of the combination of all specific traits, while the specific traits on their own might have other effects. For this reason, the most effective and the least effective robot per domain are stated in Table 5, based on the information in Appendix A, combined with the most effective traits in Table 3. No experimental studies have been done to compare the exact effect of these robots; the results are based on the robots that are said to be most likeable by the participants, relative to each other and to the studies done in each domain, which means that the assumptions have been made from logical deduction. Traits that overlap in the robots that are said to be the most likeable and least likeable, cancel each other out (meaning that the appearance effect is regarded as neutral) and are removed from Table 5.

Table 5

Most and least effective robot plus overlapping traits per domain, based on assumptions

Domain	Most effective robots	Corresponding and overlapping traits	Least effective robots	Corresponding and overlapping traits
Education and entertainment	NAO with clothes, Huggable	Apparel Eyelids Snout Flexibility	Teksta, Nexi	Nose Eyebrows Open parts Metal surface
Health and therapy	CuDDler, NAO with clothes	Eyelids Apparel Snout Eyebrows Genderedness	Aethon TUG altered version, Justin	Absence of mouth Metal surface Wheels Open parts Absence of head
Home and workplace	Pleo (pink girl), Pepper	Genderedness Skin Familiarity Wheels	Ethon 2, Meka	Open parts Metal surface Absence of mouth
Public service	NAO, Pepper, RobotMan	Proportionate body parts Wheels	N/A	N/A
Social sciences	Animatronic Male Wolf, PARO	Snout Tail Familiarity Legs/paws Flexibility	Ethon 2, HRP-4	Wheels Absence of mouth Disproportionate body parts Open parts

Note. For the domain public service, not enough studies have been found to make comparisons on the least effective robots.

4.3.3 Specific results for domains

For the healthcare and therapy domain, likeable objects or companions can be helpful in elderly care, but due to some restrictions in elderly institutes, pets might not be allowed: here robots can be a valuable addition in companionship for elderly people (McGlynn et al., 2017; Tan et al., 2013). Applying the results from Appendix A, this could be interesting, not only for the elderly, but also for children: Logan et al. (2019) found that children report more positive affect after interacting with a Huggable robot, than with a tablet or a regular plush animal. The Huggable was rated high on likeability, and has shown positive feedback regarding enjoyment and friendliness, and the children even reported lower levels of pain (Logan et al., 2019). For elderly people in nursing homes, the CuDDler robot proved to be an effective option, with high likeability ratings, and higher enjoyment (Tan et al. 2013). Found by McGlynn et al. (2017), the cuddly robot seal PARO is also an effective companion for elderly people in therapy. PARO, Huggable and CuDDler all have a soft fur and provide comfort, so adding a fur might be efficient for therapy care social robotics. These robots are also high in familiarity, which seems significantly important.

Research of public service robots and likeability is scarce, only three articles regarding public service robots are found in this review. One example of a security robot is RobotMan: as found by Trovato et al. (2019), RobotMan could be a viable option for the security domain for robots. The security robot received similar likeability ratings as a human security guard: participants reported mixed positive and negative attitudes for both the robot guard and the human guard, which might suggest that a security guard does not have to be likeable to be effective. Furthermore, some participants did not engage with the security robot because they thought of the robot as an unfamiliar, unexpected object. When expectations are managed and it is made known that a security robot is patrolling the halls, the robot might be more effective (Trovato et al., 2019).

In the public service domain, machine-like traits do not seem to lower effectiveness: for the security robot RobotMan, which has life-like traits and machine-like traits, the likeability and success rate were similar to that of a human security agent (Trovato et al., 2019). This suggests that likeable appearance traits are not necessary for a security robot to be effective in a security position, but future research is necessary to validate this effect, since only three relevant articles were found for the public service domain.

Concerning the domain of home and workplace, de Graaf et al. (2016) present in a long-term study that robots intended for use at home should look nice, safe and friendly. On top of this, small humanoids are seen as more friendly for home use (Dinet & Vivian, 2014), and elderly people think that robots in their homes should not be too big (Frennert et al., 2012; Prakash et al.,2014).

4.3.4 Results regarding time effects

The majority of studies found in Appendix A is short-term. From the long-term studies, several general suggestions can be formed: de Graaf et al. (2016) suggest that because of mismatched expectations, the robot Karotz became boring over time, and evaluations of liking decreased. The same goes for the robot Pleo, who was expected to evolve, grow and learn the longer it was interacted with, but since it could only do the same few things, it lead to disappointment in adults (Fernaeus et al., 2010). These results indicate effects that occur over time, where likeability ratings decrease over time, even with the same robot. As a solution for this, Shibata (2004) states that a robot's appearance is most important for short-term interactions, while robot learning might be most important for long-term interactions. For this reason, long-term HRI should perhaps be viewed differently from short-term HRI.

4.3.5 Noteworthy appearance traits

Some combinations of machine-like traits can have a drastically negative effect on likeability as well. The human-like robot BARTHOC has life-like and human-like traits, but also open body parts, revealing metal structures and wires. As a result, BARTHOC receives mostly negative evaluations, being unlikeable, scary and looking "like it came from a horror movie" (Lohse et al., 2007). It seems that the trait 'open body parts' is a notable predictor for low levels of likeability, compared to the other machine-like traits. The same result is found by Overgoor and Funk (2018): the furry robot Idlebot received higher ratings of likeability than the Idlebot prototype, where the latter had open body parts, no legs, and no skin or fur. A third example is found in the study by von der Pütten and Krämer (2012), where the least likeable robots all had open parts. These results suggest that open parts that reveal internal structures and technological parts decrease likeability: covering up open parts of robots so that the technical parts are hidden, might be beneficial to the overall likeability score.

Noteworthy for animal-like traits is that the trait ‘fur’ is not a significant predictor of animal-likeness, since there are animals that do not have fur. Instead of fur, some form of skin appears to be more important for life-likeness, and familiarity is a higher predictor for animal-likeness than fur (Löffler et al., 2020). This is reasonable since not all animals have fur, and the degree of familiarity makes it appear more natural than just fur would do (Löffler et al., 2020). However, robots with fur that resemble an animal with fur and are thus familiar, seem to be more likeable than robots without fur: as found by Schwind et al. (2018), a cat robot with fur and whiskers is more likeable than one missing these features. So, fur is not a determinant for animal-likeness, but when familiar, it can lead to more likeability in an animal-like robot.

5. Discussion

5.1 Answering the research questions

As stated in the introduction, the research questions for this thesis are:

RQ1: Which appearance traits are specific for human-like, animal-like, and machine-like robots?

RQ2: What specific effect does each of these traits have on the overall likeability rating of a robot’s design?

Based on the information in the literature found by the systematic approach, I aim to answer these questions. The first research question is answered based on the results section: the specific traits can be found in Table 2. As seen in the Appendix B, the appearance codes from Table 2 are applied to the robots found in the literature, leading to Table 3. Here, the human-like, animal-like and machine-like traits seem to be coherent and relevant for the robots found in the selection process: robots that fall into one of these three categories seem to possess the same appearance traits.

Regarding the appearance categories in RQ1, the category for life-likeness was not accounted for at the start of the selection process, but regarding the selected literature it appears to be an important separate category. A robot does not have to resemble a specific animal, or a realistic human, to be likeable. These life-like traits seem to have a considerably positive effect by themselves. Examples of this positive effect of these basic traits are found in studies using NAO and Pepper, who both only consist of life-like traits: NAO is a valuable addition for engagement and guidance in museums due to its high likeability (Pitsch et al., 2011), and Pepper is liked more than Erica and Sophia, which are both human-like (Esposito et al., 2019). Even more so, NAO and Pepper both have the four basic life-like traits, but Pepper has one additional machine-like trait: as a result, NAO is liked more than Pepper (Thunberg et al., 2017), suggesting that life-like traits are more likeable than machine-like traits. In general, machine-like traits seem to stand directly opposed to life-like traits, as supported by Phillips et al. (2018), Löffler et al. (2020) and Nakata et al. (1999). Machine-like traits are seen as least likeable considering Table 3, thus replacing machine-like traits with life-like, human-like or animal-like traits could account for an increase in likeability, with a few exceptions.

The second research question goes more in-depth regarding the effect on likeability, evoked by appearance traits. The overall score for likeability consists of the items liking, friendliness, kindness, relaxedness, niceness, companionship, pleasantness, enjoyment, attractiveness, and anxiety (Figure 1). Each of these items influences likeability, and some items might have a different effect than others. An important note for the interpretation of the results, is that I will make some assumptions based on interconnections and combined results from the literature found. This is done with caution, since the studies done were not all similar in design, which means comparing them is not always accurate. To completely validate these findings, future research is necessary, which will be addressed later.

To study the effect that each individual appearance has on each subscale for likeability, Appendix C is presented. To interpret the findings in Appendix C, the 10 items that form the overall score for likeability, are looked into. When leaving out the category for life-likeness (since these traits are considered likeable by themselves), the human-like trait 'skin' (HL3) seems to be the human-like trait that infers the most liking, and the animal-like traits 'familiarity' (AL4) and 'snout' (AL1) seem to infer the highest liking for animal-like traits. The animal-like trait 'legs/paws' (AL5) seems effective for friendliness and the animal-like traits of 'snout' (AL1) and 'external ears' (AL3) seem to be effective for higher enjoyment. 'Familiarity' (AL4) and 'legs/paws' (AL5) seem to have a positive effect on relaxedness and might help people relax during the interaction with a robot that possesses these traits. These traits might be effective because of the direct association they have with pets: Kaminski et al. (2002) found that pets (more specifically, dogs) have a positive effect on perceived support, increased relaxation and nurturing responses in child-therapy. These effects might be applicable to animal-like robots as well. For the other items, there are only n = 3 or fewer occurrences, so comparisons are difficult to make here.

Based on the relevant literature, it seems most important that a robot has a head, a mouth, pupils, and proportionate body parts, to be seen as likeable. This does not imply that a machine-like robot cannot be likeable, but from the literature found on the subject of likeability, machine-like robots seem less likeable than human-like and animal-like robots. Adding skin (rather than open parts for a body), a snout, ears, paws, and making the robot resemble an existing animal or a fantasy creature, could be significantly beneficial to increase likeability ratings.

5.2 Feedback to the theory

Next to answering the research questions, in relation to the articles found in Appendix A, some information is worth mentioning. There is some feedback to the theory stated in the introduction and regarding previous expectations, on two effects that were not expected in advance: the uncanny valley effect and the effect of expectations.

5.2.1 Uncanny valley effect: android robots

Previously it has been stated that adding human-like or animal-like effects to a robot's design is beneficial for likeability in general. However, one exception on this suggestion is found in android robots. Android robots are robots that look extremely realistic to humans. Examples are Geminoid HI-1/HI-4, Erica and Sophia (Esposito et al., 2019; Schweinberger et al., 2020). A curious phenomenon occurs with extremely human-like robots, called the uncanny valley effect: originally proposed by Mori (1970), the uncanny valley effect explains the occurrence that highly human-like, but imperfect robots are deemed as less likeable, creepy and eerie, in comparison to medium human-like and low human-like robots, showing a cubic function (Löffler et al., 2020; Mori et al., 2012). A too high level of human-likeness can create an eerie response in humans, where a robot is not entirely seen as a robot anymore, but also not yet a living thing; which has serious negative effects on human affect (Dautenhahn et al., 2005; Tschöpe et al., 2017). Because of the uncanny valley effect, android robots like Geminoid HI-1/HI-4, Repliee Q2 and Erica score very low on likeability, despite having appearance traits that could be likeable in other combinations (Hegel et al., 2009; Kim et al., 2019; Schweinberger et al., 2020).

Next to human-like robots, the uncanny valley effect has also been found in animal-like robots. This effect is, however, different from the effect found in human-like robots: where the uncanny valley lies in extremely realistic robots for humans, and likeability only goes up again with real humans, animal-like robots are liked least when they are partly stylized and partly realistic, showing a quadratic (U-shaped) function rather than a cubic function (Löffler et al., 2020; Schwind et al., 2018). Therefore, it is advisable for robot designers to avoid extreme human-like embodiments when designing a robot, while animal-like robots can be made either stylized or realistic, since a robot's physical appearance is the biggest predictor of the uncanny valley effect, rather than behavior and speech (Dautenhahn, 2004; Dautenhahn et al., 2005).

A very important implication for the result section and Table 3, based on the evaluation of androids, is the note that the not significant human-like appearance traits provide an erroneous image of effectiveness. Due to the too high amount of realism, some human-like traits are seen as not significant, while in other combinations they would have been effective. Table 3 makes it seem as if human-like traits are relatively not effective, while another image would have been presented if androids were left out. For this reason, it is important that androids are considered as a different category for future research.

5.2.2 Effect of expectations

Next to the degree of life-likeness, expectations based on appearance seem to be important as well. The phenomenon that occurs when people have expectations, assumptions, and biases of attributing human traits to robots, based on a robot's or computer's appearance, text or behavior is known as the Eliza Effect, as proposed by Ekbia (2008). This can lead to a misinterpretation of a robot's capabilities and goals, such as expectations of the ability to perform highly complicated tasks (Dillon, 2020).

The Eliza Effect is found in several items in this review. First of all, Kim et al. (2019) found that consumers attribute more human psychological traits to robots with more human-like appearance traits, creating the expectation that a robot can speak and think like a real human. These psychological traits are often unrealistic for a machine, resulting in an unmet expectation, ultimately leading to a decrease in attitudes and likeability (Kim et al., 2019). Austermann et al. (2010) found the same effect, showing that participants had higher expectations from ASIMO (human-like) than from AIBO (animal-like), suggesting that human-like robots elicit unrealistic human behavioral expectations. These findings can suggest that a robot should not look too life-like, since more life-like traits can infer unrealistic expectations.

It appears the preferable appearance for a robot is stylized rather than realistic. However, robots should have a minimum ability to perform simple tasks, too, since De Graaf et al. (2017) make a similar finding of a minimal expectation of behavior, when participants reported that the animal-like robot Karotz 'did not do so much' and its appearance suggested that it could perform more varieties of tasks than it could perform in reality. In line with this suggestion is the study by Hegel et al. (2009), further showing that a robot's appearance determines expectations of what domain it is operating in. For example, as rated by participants, ASIMO (human-like) is expected to be suitable for security work, research, healthcare and as assistant, and PARO (animal-like) is expected to be suitable as toy or as pet. These examples show a different level of expected intelligence based on their human-like or animal-like appearance: a human-like robot is expected to be more intelligent and to do more difficult tasks. This effect is also found by Nomura et al. (2008), suggesting that human-like robots are viewed as more suitable for social work, and animal-like robots are viewed more suitable as pets or toy; these assumptions are validated for cross-cultural participants. Supporting this is the study by Collins et al. (2015), claiming that animal-like robots set a lower expectation of task-performance and intelligence, than human-like robots. If the robot then does not meet these expectations, its likeability goes down (Caudwell et al., 2019; de Graaf et al., 2017; Hegel et al., 2009; Kim et al., 2019). Another example of the expectation effect, as seen in the results, is the previously discussed security robot RoboGuard: by managing expectations about a robot monitoring people's surroundings, the robot might be more successful in its task (Trovato et al., 2019).

As stated in the introduction, a negative result of these unmet expectations is that evaluation of the robot will disappoint the consumer, eventually leading to unwanted negative outcomes in the usage of the robot (Fernaeus et al., 2010). These negative outcomes might be partly avoided by adequately adjusting a robot's appearance to calculate the effects its appearance will evoke, based on the guidelines presented in the following text. On top of this, Löffler et al. (2020) suggest that, to overcome mismatched expectations, a robot's appearance should be based on an uncommon, but familiar animal rather than a common animal. This is in line with the findings from Appendix A and Table 3, where familiarity appears to be an important appearance trait: note that familiarity relates to a known mental model of an animal, which can be a common animal, uncommon animal, or a known fantasy creature. Löffler et al. (2020) state that an uncommon animal such as a seal creates lower and less intelligent expectations than a common animal such as a dog. Even more important seems to be the fact that a robot's appearance should match its goal or task (de Graaf et al., 2016).

5.3 Cross-connections with appearance

Some relevant concepts are noteworthy, that follow from the results section. The concepts anthropomorphism and robot genderedness seem to be narrowly intertwined with robot appearance and human affect: these concepts were not accounted for when searching for relevant literature, but they surfaced because of the results from several studies .

5.3.1 Anthropomorphism

Relevant of the uncanny valley effect is anthropomorphism, defined as the attribution of human traits to inanimate objects, to make them more desirable (Han et al., 2019). As suggested by Bartneck et al. (2009), an anthropomorphic robot might be perceived as more likeable. This means that anthropomorphism seems to be highly correlational with likeability: in Barco et al. (2020), NAO is seen as the robot with the highest level of anthropomorphism, compared to Cozmo (machine-like) and Pleo (animal-like), and NAO is seen as the most likeable, friendly, and sociable robot. Like Barco et al. (2020), 10 of the 42 articles found in Appendix A, measure anthropomorphism directly next to likeability.

An example of how anthropomorphism can be used to induce positive effects, is the study by Osawa et al. (2008): when human parts are attached to printers in a working environment, ease of use is rated higher, human attributes are assigned to the printer (giving it a nickname instead of naming the model), functions are recognized sooner, and the printer itself is noticed sooner. However, too much anthropomorphism to the point where an inanimate object seems like a realistic human, induces the negative uncanny valley effect: too high levels of anthropomorphism increase psychological warmth, but decrease likeability (Kim et al., 2019). The same is found by Tung (2016), showing that children like a robot less, when anthropomorphism is too high. However, if by increasing anthropomorphism, machine-likeness drops, medium levels of anthropomorphism in appearance traits can increase likeability: Chung and Shin (2015) show that anthropomorphism increases life-likeness, which in turn increases likeability, for same-gendered anthropomorphized Pleos.

5.3.2 Genderedness

These likeable same-gendered Pleos (Chung & Shin, 2015) introduce the importance of genderedness: assigning a gender to a robot can be seen as a specific form of anthropomorphism. The robot's gender seems to be important; several studies state an effect of genderedness. Genderedness is categorized as a human-like appearance trait in Table 2, and it can be applied on other categories of robots too (like any other trait). Independently of being human-like, female gendered robots tend to be associated as friendlier, are more easily accepted in the home and tend to reduce anxiety and eeriness (Esposito et al., 2019; Ladwig & Ferstl, 2018; Woods, 2006). Also, female animal robots are viewed as more likeable (Chung & Shin, 2015), and female robots are more effective and likeable for children (Woods, 2006).

Though female genderedness can be positive in general, same-genderedness might be even more effective: Cameron et al. (2018) found that same-gendered robots as the human interacting with them, are more effective than cross-gendered robots: boys had more fun with a boy robot, and girls had more fun with a girl robot. Zhumabekova (2018) supports this result, with findings that suggest that children liked playing with a robot with the same gender as them.

Interestingly, no significant differences for participant's gender have been found. Regardless of gender, evaluations of robots tend to be similar across individuals. Only Dinet and Vivian (2014) found a small, not significant difference, where men evaluated robots higher on likeability then women. The robot's genderedness seems to be more important than the interacting human's gender.

5.4 Generalizability

When determining the generalizability of this study, it is noted that some mixed results are found in the literature in Appendix A. Some of these mixed results could be explained through effects of culture (Bernotat & Eyssel, 2018; Li et al., 2010; Löffler et al., 2020). On top of cultural effects, there will always be personal differences, but generalizable trends can be deduced from the found literature as well. The specific results are dependent on the context that the study is done in, or the domain, target demographic, age, or culture (Bernotat & Eyssel, 2018; Li et al., 2010; Löffler et al., 2020). Trends found in the literature that will be discussed in more detail, are cultural effects, demographic effects of age, and specific diagnoses for therapy.

The effect of likeability of a robot's appearance could be influenced by cultural background of the user. Found in Appendix A are examples of these cultural effects: Katz and Halpern (2014) found that participants with Judeo-Christian beliefs report lower positive attitudes toward the robots Romeo, AIBO and 'the android' than participants with Eastern religious beliefs: Judeo-Christians report lower levels of liking and higher robot anxiety. Li et al. (2010) found similar effects, where German participants reported less positive attitudes towards the LEGO Mindstorm NXT robots, than Korean and Chinese participants. However, Bernotat and Eyssel (2018) demonstrate that Japanese participants reported similar levels of liking and uncanniness towards the robots Floka and Meka, as German participants. Japanese participants even reported lower levels of trust towards these robots than German participants. Again, personal differences apply here, but general suggestions can be derived from previous literature when conducting a similar study with a similar target demographic.

Whereas the articles in Appendix A provide specific effects of specific robots, the literature review by Papadopoulos and Koulouglioti (2018) presents an overview of cultural effects on robot evaluations. The results are fairly broad, but provide valuable information to determine which robot works best for which cultural background. If cultural effects or correlations are expected in a certain robot's design, this article can provide suggestions for either a human-like or an animal-like robot, depending on the context. However, this article does not specify specific appearance traits, and reports general effects on attitudes, and it concludes with the notion that concrete conclusions cannot be made, and more research that is guided by strong theoretical frameworks is needed (Papadopoulos & Koulouglioti, 2018).

Next to cultural effects, trends on age effects follow from the results, as well. Older people are generally expected to have low acceptance of technology, and thus to dislike robots more than younger people (Broadbent et al., 2009). However, in this literature review, findings of the elderly's likeability perception of a robot are mixed: Schweinberger et al. (2020) found that older people reported higher likeability ratings than young adults, for the robots ASIMO, NAO, Ri-man, Wakamaru, Geminoid HI-1/HI-4, HRP-4C, Justin and Robonova II. For these robots, elderly people liked HRP-4C the most, which is very high in human-likeness, with a metal surface. On top of that, a large-scale study by Gnambs and Appel (2019) found no significant negative influence of age on robot likeability. Interestingly, Giuliani et al. (2005) show that "younger-old" adults (63-75 years) have higher technology acceptance than "older-old" adults (>75 years). These results combined, when taking the publication date for these studies into account, could suggest that there is an increase in acceptance and likeability for elderly people over the years, and the elderly could become more and more used to social robotics as time goes on.

Supporting the idea that the elderly do indeed find robots likeable, is the study by Esposito et al. (2019), showing that elderly people perceive the robots Pepper, Erica and Sophia as likeable, and are eager and curious to interact with them. Along with human-like robots, the animal-like robot PARO is also seen as likeable, relaxing and friendly by the elderly (McGlynn et al., 2017). However, Chu et al. (2019) found no significant effect for likeability of human-like or animal-like robots for the elderly, demonstrating that the results of the elderly and likeability of robots are mixed. Even though results are mixed, there seems to be a positive trend for robots regarding elderly people: it is important that likeability for the elderly is studied further, since robots are considered a promising technology that can assist and prolong independent living among older adults (Khosravi & Ghapanchi, 2016).

Next to effects for the elderly, some findings have emerged about children as well: children seem to have an aversion to android robots, as they might be scary (Feng et al., 2018). Furthermore, children often evaluate human-like and animal-like robots to be happy, and machine-like robots to be sad, signaling a preference for human-like and animal-like robots over machine-like robots (Woods, 2006). In line with this, for children a robot is more likeable when it has legs to move rather than treads or tracks; when it has facial features and when it is gendered (Woods, 2006). As suggested by Tung (2016), human-like traits are beneficial for likeability, when the robot interaction concerns children: human-like traits, but not necessarily a complete human form increase positive emotions in children. In line with the uncanny valley effect, robots for children should not be too realistic, because children are nowadays exposed to robots in such an extent that the concept of robot is not novel to them anymore (Tung, 2016; Woods, 2006). To not blur the line between humans and robots, robots for children could have cartoon-like features and bright colors to infer likeability (Woods, 2006). Even younger children might prefer cute, cuddly robots with fur, and again stylized and cartoon-like features, as for ages 3 to 5, a robot with a round, furry body shape and large eyes, was liked much more than a comparable robot with a metal surface (Howard & Vick, 2010).

Regarding neurological effects, one study reported results on elderly people with a mild cognitive impairment: elderly people without cognitive impairment tend to like the machine-like robot RAMCIP more than elderly people with mild cognitive impairment (Gerłowska et al., 2018). It might be beneficial to research the effect of machine-like robots versus human-like and animal-like robots for elderly people with a cognitive impairment, to study which embodiment is the most effective for this target demographic.

Another neurological effect can be found in participants diagnosed with autism spectrum disorder (ASD). People diagnosed with ASD seem to have a preference for human-like robots over animal-like or machine-like robots. Within the participants with ASD there seem to be personal differences: younger individuals with ASD prefer plain human-like robots, while older children with ASD slightly prefer realistic androids (Kumazaki, 2017). This is in line with Robins et al. (2006) who found that robots used in child therapy related to ASD are most effective when a plain, humanoid robot is used. This could be the case since younger children in general could be frightened by extremely life-like robots, such as ACTROID-F (Kumazaki et al. 2017). Interestingly and contrarily to Kumazaki et al. (2017), the uncanny valley effect discussed previously, might be weak in individuals with ASD (Feng et al., 2018). In research with children, aged 5 to 7, participants with ASD showed no decrease in liking for either stylizing faces and increasing realism in human faces, while participants without ASD reported a decrease in liking when presented with these modifications (Feng et al., 2018).

These mixed results call for caution when studying effects in individuals with ASD, because of the great personal differences and results on ASD, which are difficult to generalize. The occurrence that does arise from multiple studies, is the fact that human-like robots are preferred over animal-like and machine-like robots for participants diagnosed with ASD.

5.5 Theoretical implications

The results of this study introduce some implications and learning points: they prove the value of researching likeability and appearance effects and researchers will benefit from adding these components when producing robots. Table 3 and Table 5 provide insight on the effects that specific traits have on likeability, while Figure 3, Figure 6, Figure 7 and Table 4 offer insight on which domains and categories deserve more attention in research. By doing more studies based on the categories presented here, robots can be optimized to fully connect to the intended goal for each domain. When applying these results on further research, it is important to also take into account other factors that might influence likeability, being expectations, cultural effects and age effects, and the uncanny valley effect (with different effects for human-like robots than animal-like robots). On a larger scale, these results could be used in relation to research on acceptance of robots, or even further on intention of use for robots, trust in robots and automation, and robot engagement.

Even though not much research is found on likeability of robots in the public service sections, robots are already employed as security robots in the public service domain here and there. An interesting example for public service robots is Boston Dynamic’s robotic dog SPOT, which is deployed in a public park to remind people to keep a safe distance from each other during the COVID-19 pandemic (Vincent, May 2020). Furthermore, Weiss et al. (2014) found that a luggage carrying robot is seen as purely task-oriented, and participants evaluating its design reported that its likeability was less important than its task efficiency. To improve its design, participants suggested adding gaze or gesture movements, but a human-like appearance was not necessary (Weiss et al., 2014), signaling that robots in the public service domain are evaluated differently and calling for future research on the matter.

Regarding Figure 3, it should be noted that, from the established domains of social robotics, there were two domains for which no research is found: search and rescue, and industry. This is remarkable, since the military already uses robots for search and rescue missions (Carpenter, 2016; Jentsch, 2016). Research on military robots' likeability is therefore meaningful. For the domain of industry, it is beneficial to examine likeability effects too, even if at first glance it does not seem relevant. For example, Bortot et al. (2013) suggest that HRI in industrial robotics grows in interest, making design and psychology relevant here as well. Robots that make non-predictable motions lead to reduced well-being and performance in staff (Bortot et al., 2013). Introducing anthropomorphism or likeability to create attention for the robots might help, making research on likeability is necessary in the industry domain.

An interesting follow up for these domains could be anthropomorphism of objects and robots, as it is discussed that robots with human-like traits could lead to more effective task completion in the workplace, as suggested by the anthropomorphized printers discussed before (Osawa et al., 2008).

5.6 Practical implications

The results discussed in this discussion section, are summarized and presented here as general guidelines for robot design, intended for researchers and employers active in any of the domains for social robotics. Importantly, communicating tasks and capabilities of the robots is necessary to ensure there are no mismatched expectations that can decrease likeability.

A life-like base with the traits head, mouth, pupils and proportionate body parts is suggested as starting point. Also, open parts should be covered to hide mechanical and technical parts. Furthermore, specific appearance traits seem to be desired in general, in certain domains, presented in Table 6. Additionally to Table 6, when cultural effects are expected and have to be taken into account, it is suggested to consult the literature review by Papadopoulos and Koulouglioti (2018), discussed previously.

Table 6

General guidelines for specific domains, to be used as base for robots that should be likeable to humans interacting with them

When doing research in…	Start with…	Then optionally specify with…	And stay away from…
Education and entertainment	Plain human-like robots with apparel or smooth exterior	Add bright colors, legs instead of treads for walking, and stylization (medium realism) for animal-like and human-like robots	Androids
Healthcare and therapy	Huggable, cuddly or soft robot	Elderly without cognitive impairment: human-like, animal-like, and machine-like robots can all be effective Elderly with cognitive impairment: human-like and especially cuddly animal-like robots are highly preferred over machine-like robots ASD therapy: high functioning individuals might benefit from androids; low functioning individuals might prefer plain human-like robots	Machine-like robots without head and mouth, open parts, wheels, and a metal surface
Home and workplace	Nice, safe, small and friendly robot. Animal-like robots seem to be preferred, but human-like and machine-like robots can be effective too, provided they are nice, safe, small and friendly	Robots for the workplace should be task-oriented. A nice, calming color could have a positive effect	Open parts, big robots, and metal surfaces
Public service	Life-likeness is most important with head, mouth, pupils and proportionate body parts. No preference for human-like, animal-like or machine-like	For security: likeability is less important than expectation management	
Social sciences	Highly realistic animal-like robots are most effective, medium realistic human-like and animal-like robots are effective too. Metal surface is fine, too, as long as there are no open parts	Adults show slight preference for a small human-like robot over big robots and animal-like robots. This might be because small human-like robots are unobtrusive	Avoid machine-like and highly realistic human-like robots

Table 6 is intended to be used as a starting point when designing a robot's morphology from scratch, or to choose a readily available robot that already possesses the traits that match with the suggestions in this table. As an interesting idea, a robot with an adjustable appearance could be developed, where traits from Table 2 can be added or removed to make it more versatile. A customizable robot would fit the needs of multiple people with varying personal differences, as suggested by McGlynn et al. (2017), Trovato et al. (2019), and Fujita (2004). Appearance and these guidelines are mostly effective for short-term interactions; for the long-term, learning is more important. For each domain it applies that mismatched expectations should be avoided, so that disappointment will not negatively influence human affect. Since Table 6 is based on the information found in Table 5 and Table 3, it appears that there is no robot that performs well in all the different domains, at least for now. The wishes of the target audience should be taken seriously in order to design an effective robot, and this literature review illustrates reasons why.

5.7 Suggestions for further research

Some suggestions for further research can be made, based on this literature review. First of all, as noted in Appendix A and in this discussion section, androids can induce negative affective feelings, such as fear and uncanniness. For this reason, research on the expectations of android robots can give a direction of the use for android robots. For all domains found in the literature in Appendix A, androids seem to be best avoided for now. Secondly, for likeability specifically, more research is necessary on robots that are currently being used in the domains public service; search and rescue; and industry, since not enough articles have been found in these domains.

Another, more in-depth suggestion for future research, is to conduct studies with experimental measures on appearance traits in specific domains and situations. Ideally, each specific trait noted in Table 2 should be experimentally tested directly against a robot that does not possess the trait, to fully comprehend the specific effect each trait has on likeability. The appearance traits could be measured on likeability by a modified scale based on the scales and models used to define likeability (Figure 1). This way, each trait can be tested on specific effects, which can result in a framework of appearance traits, where researchers and practitioners can add appearance features to a robot, matching its intended use. This way, relevant appearance traits can be utilized, until a robot is constructed that fits the intended goal best. Even in this case, specific combinations of traits can lead to different effects, but the generalizability will be higher if each trait is studied separately.

An interesting suggestion for a particular follow-up study could be an online survey, to measure and determine the effects that individual traits have by themselves. By presenting participants with an image of a robot model that is easily (digitally) adjustable, likeability could be measured for individual appearance traits, based on the proposed items in Figure 1. By adding and removing individual traits to a (digital) robot image, effects can be tested separately to accurately measure its effects. This way, the specific effects of each of the traits presented in Table 2 can be examined in a relatively accessible way.

5.8 Strengths and limitations

This systematic literature review gives a clear view of which effects of a robot's appearance on human likeability have been found in previous research. For this review, only appearance has been taken into account, leaving out movements, gestures and gaze in the measurements used. However, it is difficult to completely rule out movement from this study, since the research done was almost never with static robots.

The categorization for this literature review is based on appearance traits of robots. This is a strong point, because this study was done using an exploratory approach, making it widely inclusive of multiple domains. Through categorizing appearance traits, some generalizations and deductions have been made in the large amount of research already done on robots. Again, the importance of the interplay between engineering, design and psychology cannot be overlooked, and I believe this study illustrates why.

What could be seen as a limitation of this work, is that the domains used might be rather broad: for Table 4, robots in domains are compared on appearance traits, while within these domains there are still several differences. For example, Kumazaki et al. (2017) performed a study on the effect of likeability of robots on children diagnosed with ASD, and McGlynn et al. (2017) studied the effect of likeability of the robot PARO on elderly people: both of these studies fall under the domain for health and education, while there are considerable contextual differences. To solve this, it is attempted to address these personal differences, and the guidelines presented are to be used as general inspirational basis rather than actual strict rules. Another limitation lies in the fact that only studies that measured appearance were included, while there is also a fair amount of research on appearance in combination with gestures and/or speech, which both have a positive impact on likeability as well. For example, Kim et al. (2013) show a positive correlation between a robot using gestures and an increase in enjoyment and engagement. Furthermore, when familiarity increases in a robot's design, participants felt more strong social interactions with the robot, where the effect was even bigger with a robot with gestures (Kim et al., 2013). This signals a need for combining robot appearance with gestures and speech, but to keep this literature review achievable and concrete, the focus was only on appearance.

A final note on this study is that indirect effects of likeability were not taken into account, to keep the scope of this work concise. Indirect effects appear to be important, for example the effects of acceptance and anthropomorphism are closely related to likeability, and thus require attention as well. These subjects are noted as suggestions for future research, to enrich the results found in this thesis.

6. Conclusion

Through a systematic literature review, determinants and expectations have come to light, regarding the appearance of robots on human affect. Since interactions are personal and personal preferences are present when coming in contact with robots, it would be ideal to discuss the wishes of the target audience, to perfect or optimize robot appearances for a specific goal. It would be even better to produce a customizable robot to fit the needs of multiple people with varying personal differences (McGlynn et al., 2017; Trovato et al., 2019; and Fujita, 2004).

The results of this systematic approach are presented in a way that clearly demonstrates the specific effects that are examined, while noting the specific domain. Again, the results presented in Table 6 are intended to provide general guidelines to use as a basis for determining which robot to use in which situation, and what details to keep in mind. Researchers and practitioners can alter the base appearance to fit specific needs. The main goal of this work is to make researchers more aware of the effect that human affect can have on various wanted and unwanted outcomes of studies with robots, and that many of these effects have an origin in a robot's appearance.

To conclude: more attention should be given to the design aspect and the psychological aspect of the production of robots. Ideally, every type of robot suggested in each domain, should be experimentally studied to validate its effects. The guidelines presented in this study are based on previous findings and have been transformed into useful assumptions to use in further research. Finally, if the design of robots is accommodated to the user's wishes and it completely fits the image of the intended use, humans might be more likely to embrace robots in their direct environment, which might even lead to more meaningful interactions in their lives.

Appendix A: *summary of literature*

Author	Year	N	Pop'n	Domain	Term:	Robot	Morphology	Relevancy	Relevant conclusion
Austermann et al.	2010	16	JP, adults	Social sciences (lab-setting)	Short	ASIMO AIBO	Human-like Animal-like	Enjoyment	Enjoyment is higher for AIBO (4/5) than ASIMO (3.5/5). First impression is based on performance rather than appearance.
Barco et al.	2020	35	UK, children	Education and entertainment	Short	NAO Pleo Cozmo	Human-like Animal-like	Companionship, friendliness	NAO's perceived social presence (companionship) is higher than Pleo. NAO and Cozmo's friendliness are higher than Pleo.
Bernotat and Eyssel	2018	102	JP, DE, adults and elderly	Home and workplace	Short	Floka Meka	Human-like Human-like	Liking	Likeability scores: in Japan, Floka (M=3.14) is liked more than Meka (M=2.79). In Germany, Floka (M=3.14) is liked more than Meka (M=2.91). Germans like Meka (M=2.91) more than Japanese (M=2.79).
Cameron et al.	2018	59	UK, children	Education and entertainment	Short	Zeno	Human-like	Enjoyment, liking	Male participants show more liking response to a robot with face with facial expressions, than without. Both females and males had greater enjoyment with face.
Castro-González et al.	2016	56	SP, adults	Social sciences (lab-setting)	Short	Baxter Baxter (arm-only)	Human-like Human-like	Likeability	No main effects of bodily appearance. No significant effect on animacy, likeability, trustworthiness, and unpleasantness
Chu et al.	2019	33	TW, elderly	Healthcare and therapy	Short	PARO Zenbo	Animal-like Machine-like	Liking	Attitude towards technology (liking) was significantly higher for Zenbo than PARO.
Chung and Shin	2015	129	KR, adults	Home and workplace	Short	Pleo	Animal-like	Liking	The feminine male is disliked significantly more than the rest. The girl-labeled (gendered) pink Pleo has the highest likeability.
de Gauquier et al.	2018	307	BE, adults	Public service	Short	Pepper	Human-like	Anxiety, enjoyment, pleasantness, niceness	Having the questionnaire done by a human-like robot has a positive impact on shopping experience and also invokes positive affective reactions. Pepper invokes liking, enjoyment, niceness and reduces anxiety.
de Graaf et al.	2016	102	NL, children and adults	Home and workplace	Long	Karotz	Animal-like	Enjoyment Attractiveness Companionship, Liking	Over time the ratings for Karotz: Enjoyment decreased M=4.89 to 3.87 Attractiveness decreased M=5.05 to 3.74 Companionship increased: M=3.31 to 3.25 Liking decreased M=4.79 to M = 4.09
Dinet and Vivian	2014	217	FR, children, teenagers, adults, elderly	Social sciences (lab-setting)	Short	Mahru-II NAO Teddy-Bear ACTROID-F HRP-4 AIBO Ri-man PARO	Human-like Human-like Animal-like Human-like Human-like Animal-like Animal-like Animal-like	Companionship, friendliness, relaxedness, anxiety	The youngest children have the most liking towards robots. The highest scores for males, respectively on age group, are: Teddy-Bear and PARO (M=9.2), Teddy-Bear (M=8.9), Teddy-Bear (M=9.0), and PARO (M=9.0). For women, the highest scores are respectively of age group: PARO (M=9.4), PARO (M=8.3), Teddy-Bear (M=8.7) and PARO (M=8.4). HRP-4 is disliked the most by everyone.
Esposito et al.	2019	51	IT, elderly	Healthcare and therapy	Short	Pepper, Erica, Sophia	Human-like Human-like Human-like	Attractiveness, liking	Elderly have a general tendency to engage with female robots, all robots have high liking. Pepper (Tina) is seen as most attractive (M=18.6), then Sophia (M=17.7), then Erica (M=16.8).
Feingold-Polak et al.	2018	60	ISR elderly, adults	Social sciences (lab-setting)	Short	Pepper	Human-like	Enjoyment, liking	Young and old adults have similar levels of enjoyment during HRI. Older adults liked Pepper more (M=5.0) than YA (M=4.0).

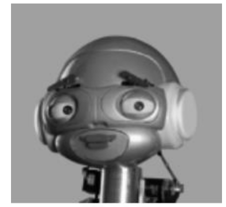
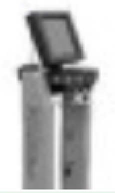
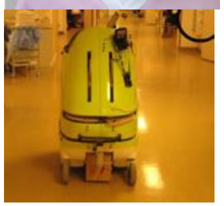
Fraune et al.	2015	127	USA, adults	Social sciences (lab-setting)	Short	NAO Pleo iCreate	Human-like Animal-like Machine-like	Liking, friendliness,	Pleo is as most likeable alone (M=3.6). Pleo is seen as most friendly when alone (M=3.6). Pleo is liked most (M=4.1). iCreate is least anxiety inducing (M=1.7).
Geoffrey Louie et al.	2014	46	CA, elderly	Healthcare and therapy	Short	Brian 2.1	Human-like	Liking, pleasantness	Attitude scored the highest (M=4.22), meaning the general attitude for Brian 2.1 is positive on a 5-point scale: the robot is well received and liked. Open answers of participants, state they "love the idea of the robot", "think the robot is fascinating and interesting".
Gerłowska et al.	2018	17	PO, elderly with and without Alzheimer's	Healthcare and therapy	Short	RAMCIP	Machine-like	Companionship	The robotic assistant is easy to get familiar with in both groups. However, RAMCIP is evaluated differently by people with and without cognitive impairment: people without MCI (dementia) rate the robot higher on all scales.
Ghazali	2019	21	NL, adults	Social sciences (lab-setting)	Short	SociBot	Human-like	Liking	SociBot appearance is a predictor for higher liking. Adding social cues gradually increases likeability.
Heerink et al.	2008	70	NL, elderly	Healthcare and therapy	Short	iCat	Animal-like	Enjoyment, companionship	iCat scores significantly higher than tablet on enjoyment and companionship
Hegel et al.	2009	183	DE, USA, UK, adults	Social sciences	Short	Barthoc iCat AIBO BIRON Keepon Kismet Leonardo Robovie Repliee Q2 ASIMO PARO Pearl	Human-like Animal-like Animal-like Machine-like Animal-like Machine-like Animal-like Machine-like Human-like Human-like Animal-like Machine-like	Attractiveness, enjoyment, liking	ASIMO is most suitable for security work, research, transport, caregiver and health care. Pearl is most suitable as assistant. PARO is most suitable as toy, pet, and companion. Repliee Q2 is most suitable as business or representation, teacher and public assistant. AIBO is most suitable for entertainment. PARO is most likeable. AIBO is most enjoyable. PARO is most attractive.
Howard and Vick	2010	19	UK, children	Education and entertainment	Short	Scoozie Teksta	Animal-like Animal-like	Liking	The mammal-like robot Scoozie is viewed much more positively than Teksta.
Katz and Halpern	2014	873	USA, young adults	Social sciences	Short	Romeo AIBO the Human-like	Human-like Animal-like Human-like	Liking	On the scale Robot-liking, the AIBO scored higher than the Human-like. Judeo-Christian people like robots less than Eastern religions. Romeo is liked most, than AIBO, then the Human-like. Human-like appearance does not affect attitude towards robots: human-likeness does.
Kim et al.	2019	106	USA, adults	Social sciences, home and workplace	Short	Ethon 2 Pepper Erica	Machine-like Human-like Human-like	Pleasantness, liking	Pepper has the highest score for warmth (pleasantness). With Erica, warmth is overtaken by uncanniness. Uncanniness negatively influences liking. Ethon 2 has the lowest score for likeability.
Kipp and Kummert	2016	39	DE, students	Education and entertainment	Long	Flobi	Human-like	Likeability	A strong opponent decreases likeability. Flobi session 1 (M=3.9), and session 4 (M=3.5). Remote human rating session 1 (M=4.1) and session 4 (M=3.8)

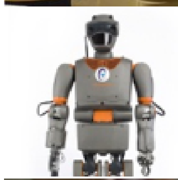
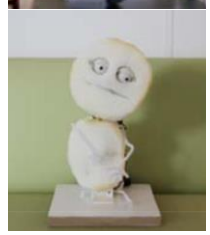
Konok et al.	2018	176	HU, students	Social sciences	Short	ASIMO Custom orb-like robot PeopleBot AIBO	Human-like Machine-like Machine-like Animal-like	Liking	People's attitudes are more negative towards robots, than dogs. The household robots Custom orb and AIBO are liked more than the social companion robots ASIMO and Peoplebot.
Kumazaki et al.	2017	16	JP, children	Healthcare and therapy	Short	ACTROID-F KABO-chan M³-Synchy	Human-like Human-like Human-like	Liking	Appearance liking varies highly between high-functioning ASD individuals. However, in older high trait ASD children, Human-like robots are liked more (ACTROID-F is liked most), which may be scary to younger children. Young children liked KABO-chan most.
Lee et al.	2011	33	SG, students	Social sciences	Short	ASIMO AIBO Pearl Repliee Q2 PaPeRo	Human-like Animal-like Machine-like Human-like Machine-like	Liking	Participants ascribe more warm capabilities to human- and animal-like robots, than machine-like robots. Liking is not influenced by higher warmth. Humanlike robots are not more likeable than nonhuman-like. Likeability is highest for PaPeRo, then ASIMO, then Repliee Q2, then Pearl.
Li et al.	2010	108	DE, KR, CHN, students	Social sciences	Short	LEGO Mindstorm NXT Human-like, Zoomorphic, Machine-like	Human-like Animal-like Machine-like	Friendliness, liking, companionship, attractiveness	Effect of culture on likeability: Chinese likeability is highest (M=4.65), then Korean (M=4.6), then German (M=3.2), for all the robots combined. Human- or animal-likeness both positively affect likeability, animal-likeness has the strongest effect.
Ljungblad et al.	2012	25	SWE, adults	Home and workplace	Long	Aethon TUG, altered version	Machine-like	Liking	Initial responses are more negative: the robot looked unfamiliar, new. It was initially named cute by some, and big and ugly by others. After 13 days, the robot was seen as intelligent, discrete, reliable, cute, cool and clever.
Löffler et al.	2020	187	DE, JP, USA, Adults and elderly	Social sciences	Short	ia. TabbyCat Pleo PARO Parle AIBO Animatronic dog Animatronic Male Wolf	Animal-like Animal-like Animal-like Animal-like Animal-like Animal-like Animal-like	Companionship, friendliness, relaxedness, anxiety, liking	The robots with a medium level of animal-likeness are judged as the least likeable. This supports the claim that realism inconsistency increases creepiness. In Germany and Japan, robots are more common and more likeable than in the US. The Animatronic Male Wolf is the most likeable (M=41), then AIBO (M=27), then Animatronic Dog (M=22).
Logan et al.	2019	54	USA, children	Education and entertainment, healthcare and therapy	Short	Huggable	Animal-like	Liking, enjoyment, friendliness	The positive effects of liking from children in hospitals are bigger with a social robot, than with a plush animal. Joy: Robot (M=0.43), Tablet (M=0.41), Plush M=0.38. Agreeableness: Robot (M=0.6), Tablet (M=0.55), Plushie (M=0.5).
Lohse et al.	2007	113	DE, students	Social sciences	Short	iCat AIBO Biron BARTHOC	Animal-like Animal-like Machine-like Human-like	Enjoyment, liking	Most enjoyable robot: AIBO 57%, then iCat 17%, then Biron 17%, then BARTHOC 9%. Most likeable: AIBO 46%, then iCat 38%, then Biron 9%, then BARTHOC 3%

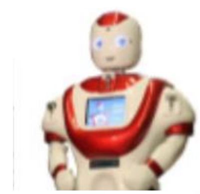
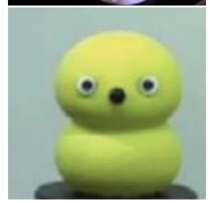
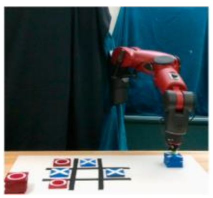
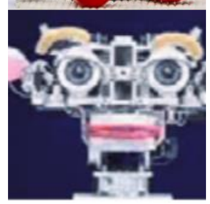
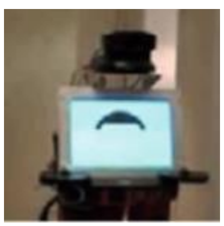
McGlynn et al.	2017	30	USA, elderly	Healthcare and therapy	Short	PARO	Animal-like	Liking, enjoyment, companionship, relaxedness	PARO makes participants feel less anxious, and more calm. It was also seen as cute and friendly. 12% say it has a general use for enjoyment. 25% of the respondents say it could be their friend, showing companionship
Okanda et al.	2019	50	JP, students	Social sciences (lab-setting)	Short	KIROBO AIBO Tapia	Human-like Animal-like Machine-like	Companionship	KIROBO is seen as most alive and psychological with 0.35, then AIBO, then Tapia. KIROBO is seen as best fit to befriend (companionship). Most participants wanted to be friends with all robots.
Overgoor and Funk	2018	21	NL, students	Social sciences (lab-setting), home and workplace	Short	IdleBot prototype IdleBot	Machine-like	Liking, friendliness, pleasantness, relaxedness	IdleBot is seen as positive and liked, making a room less lonely. It helps people destress and calm down. IdleBot scores high on likeability. IdleBot prototype is too shy and neutral: a bright orange fluffy coat increased its likeability.
Pitsch et al.	2011	177	DE, museum visitors	Public service	Short	NAO	Human-like	Friendliness, liking, pleasantness	NAO as a guide is seen as likeable and a positive experience (high ratings of friendliness, available, motivating).
von der Pütten and Krämer	2012	151	DE, adults	Social sciences	Short	Clusters: C1: NAO, C2: Ri-man, C3: Geminoid HI-1, C4: HRP-4C, C5: Justin, C6: Kismet	Human-like Animal-like	Likeability	The cluster with the highest likeability, C4, consists of Geminoid DK & HRP-4C. However, this cluster is also seen as threatening. The second highest likeability and lowest threat is C1, consisting of Robovie, Cosmobot, Autom, PaPeRo, RIBA, NAO, ASIMO, Atom and Leonardo. The lowest likeability scores are for Justin, Robonova II, Robosapien, HRP-4 and REEM-1.
Schweinberger et al.	2020	60	DE, Young adults and elderly	Healthcare and therapy	Short	ASIMO NAO Ri-man Wakamaru Geminoid HI-1/HI-4 HRP-4C Justin Robonova II	Human-like Human-like Human-like Human-like Human-like Human-like Human-like Human-like	Likeability, companionship	In general, people with high autistic traits on the AQ scale, rated the robots more likeable. For both genders together, high AQ view robots as more likeable (M=3.44) than low AQ (M=3.15). By young participants, NAO is most likeable (M=3.08) and for older participants HRP-4C is most likeable (M=4.78). Geminoid HI-1/HI-4 and Justin are least liked by all groups.
Sinatra et al.	2012	111	USA, Young adults	Social sciences (lab-setting)	Short	AIBO Lego Mindstorm NXT Human-like	Animal-like Machine-like	Friendliness Kindness anxiety	The dog was rated highest in all categories, but the AIBO was rated higher than the cat in friendliness.
Tan et al.	2013	59	SG, elderly	Healthcare and therapy	Short	CuDDler	Animal-like	Liking, friendliness, pleasantness, niceness, relaxedness	Among all the Godspeed attributes, likeability attribute is rated highest with an average score of 4.06. Individual scores: liking M=4.14, friendliness M=4.10, kindness M=3.93, pleasantness M=4.08, niceness M=4.02, relaxedness M=4.20.
Thunberg et al.	2017	36	SWE, students	Social sciences (lab-setting)	Short	NAO Pepper	Human-like Human-like	Likeability	NAO is seen as more likeable than Pepper. Also people use more positive words when describing NAO, than Pepper.

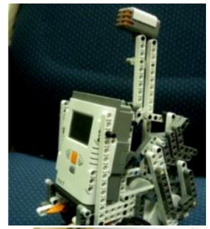
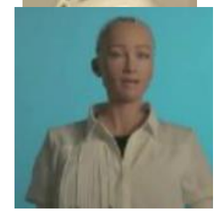
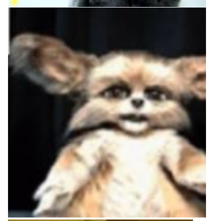
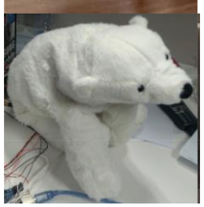
Trovato et al.	2019	102	SP, students	Public service (security)	Short	RobotMan	Human-like	Liking	Most participants found RobotMan positive, some found it negative. Almost all participants engaged with the robot, some ignored it. Liking for RobotMan increased after interaction. It gets similar results on affect as the human.
Tung	2016	578	CHN, children	Education and entertainment	Short	A1, A2, A3, wakamaru, B2, NAO, Nexi, C2, C3, Erica, HRP-4C	Human-like (all)	Companionship, attractiveness.	Children find robots that are low/medium human-like more attractive than robots that are medium/highly human-like. The B group was most attractive. B2 was highest social attraction (companionship), and C3 had the lowest. B2 had the highest physical attraction, C1 the lowest. Subgroup A is least human-like, subgroup D is most human-like.
van Straten et al.	2019	8	NL, children	Education and entertainment, healthcare and therapy	Short	NAO NAO with clothes	Human-like	Liking	Humanizing body appearance leads to more liking and is more interesting: NAO (M=3.4 and M=3.1), vs NAO with clothes (M=4.0 and M=4.1).

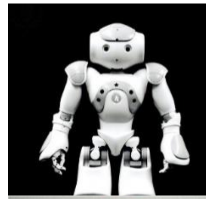
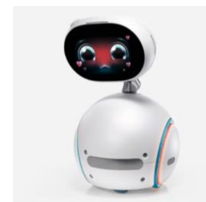
Appendix B: *image database*

Name	Image	Code		Name	Image	Code	Source	Name	Image	Code	Source
A1		ML3, ML4, ML5, ML6, ML7	Tung (2016)	Geminoid DK		LL1, LL2, LL3, LL4, HL1, HL2, HL3, HL4, HL5, HL6, HL7	von der Pütten and Krämer, (2012)	PARO		LL1, LL2, LL3, LL4, AL1, AL2, AL3, AL4, AL5, AL6, AL7	Retrieved from Encyclo-pædia Britannica ImageQuest
A2		ML4, ML5, ML6, ML7	Tung (2016)	Geminoid HI-1/HI-4		LL1, LL2, LL3, LL4, HL1, HL2, HL3, HL4, HL5, HL6, HL7	Schweinberger et al. (2020)	Pearl		LL1, LL2, LL3, HL1, HL6, ML1, ML3, ML7	Prakash and Rogers (2014)
A3		ML3, ML5, ML6, ML7	Tung (2016)	HRP-4		LL1, LL4, ML5, ML7	Dinet and Vivian (2014)	PeopleBot		ML1, ML4, ML5, ML6, ML7	Konok et al. (2018)
ACTROID-F		LL1, LL2, LL3, LL4, HL1, HL2, HL3, HL4, HL5, HL6, HL7	Kumazaki et al. (2017).	HRP-4C		LL1, LL2, LL3, LL4, HL1, HL2, HL3, HL4, HL5, HL6, HL7, ML7	Logan et al. (2019)	Pepper		LL1, LL2, LL3, LL4, ML1	Pandey and Gelin (2018)
Aethon TUG altered version		ML1, ML3, ML4, ML5, ML6, ML7	Ljungblad et al. (2012)	Huggable		LL1, LL3, LL4, HL1, AL1, AL3, AL5, AL7	Lohse et al. (2007)	Pleo		LL1, LL2, LL3, LL4, HL1, HL3, AL1, AL2, AL4, AL5, AL7	Retrieved from Encyclo-pædia Britannica ImageQuest.

AIBO		LL1, LL4, AL1, AL2, AL3, AL4, AL5, AL7, ML7	Lohse et al. (2007)	iCat		LL1, LL2, LL3, HL1, HL6, AL1, AL3, ML2, ML6	Fraune et al. (2015)	RAMCIP		ML1, ML3, ML4, ML5, ML6, ML7	Kostavelis et al. (2018)
Animatronic Dog, Jim Henson company		LL1, LL2, LL3, LL4, AL1, AL2, AL3, AL4, AL5	Löffler et al. (2020)	iCreate		ML1, ML3, ML4, ML5, ML6	Fraune et al. (2015)	REEM-I		LL1, LL2, LL3, LL4, ML3	von der Pütten and Krämer (2012)
Animatronic Male Wolf, Sally Corporation		LL1, LL2, LL3, LL4, AL1, AL2, AL3, AL4, AL5, AL6, AL7,	Löffler et al. (2020)	IdleBot		LL1, LL2, LL3, HL1, HL7, AL4, AL5	Overgoor and Funk (2018)	Repliee Q2		LL1, LL2, LL3, LL4, HL1, HL2, HL3, HL4, HL5, HL6, HL7	Hegel et al. (2009)
ASIMO		LL1, LL2, LL4, HL7	Retrieved from Encyclo-pædia Britannica ImageQuest .	IdleBot prototype		LL1, LL2, LL3, HL1, ML3	Schweinberger et al. (2020)	RIBA		LL1, LL2, LL4, AL1, AL3, AL4, AL5, ML1	Dinet and Vivian (2014)
Atom		LL1, LL2, LL3, LL4	von der Pütten and Krämer (2012)	Justin		LL1, LL4, ML1, ML5, ML7	Kumazaki et al. (2017)	Ri-man		LL1, LL2, LL4, ML1	Schweinberger et al. (2020)
Autom		LL1, LL2, LL3	von der Pütten and Krämer (2012)	KABO-chan (Smile Supplement Robot)		LL1, LL3, HL4, HL7, AL4	De Graaf et al. (2017)	Robonova II		LL1, LL4, ML3, ML5, ML7	Schweinberger et al. (2020)

B2		LL1, LL2, LL3, HL7, ML3, ML6, ML7	Tung (2016)	Karotz		LL1, HL3, AL1, AL3, AL4, ML5	Hegel et al. (2009)	RobotMan		LL1, LL2, LL3, LL4, ML1	Trovato et al. (2019)
BARTHOC		LL1, LL2, LL4, HL3, HL5, ML3	Lohse et al. (2007)	Keepon		LL1, LL2, LL3, AL7	Gasumova and Porter (2019)	Robosapien		LL1, LL2, LL3, ML3, ML6	von der Pütten and Krämer, (2012)
Baxter		LL4, ML1, ML3, ML4, ML5, ML7	Castro-González et al. (2016)	KIROBO mini, Toyota		LL1, LL2, HL4, HL7, ML6, ML7	Hegel et al. (2009)	Robovie I		LL1, ML1, ML3, ML5, ML6, ML7	Hegel et al. (2009)
Baxter (arm-only)		ML3, ML4, ML5, ML6, ML7	Castro-González et al. (2016)	Kismet		LL1, LL2, LL3, LL4, HL1, HL6, AL3, ML3, ML7	Li et al. (2010)	Romeo		LL1, LL2, LL3, LL4	Katz and Halpern (2014)
BIRON		ML2, ML3, ML4, ML5, ML6, ML7	Hegel et al. (2009)	LEGO Mindstorm NXT human-like		LL1, LL4, ML3, ML5	Li et al. (2010)	Scoozie		LL1, LL2, LL3, LL4, AL1, AL2, AL3, AL4	Howard and Vick (2010)

Brian 2.1		LL1, LL2, LL3, LL4, HL3, HL4, HL5, HL6, HL7, ML3	Louie et al. (2014)	LEGO Mindstorm NXT machine-like		ML1, ML2, ML3, ML4, ML5, ML6	Li et al. (2010)	SociBot		LL1, LL2, LL3, ML3, ML7	Ghazali et al. (2018)
C2		LL1, LL2, LL4, HL5, ML3, ML7	Tung (2016)	LEGO Mindstorm NXT animal-like		LL1, HL3, AL3, AL4, ML5, ML6	Hegel et al. (2009)	Sophia (Hanson Robotics)		LL1, LL2, LL3, LL4, HL1, HL3, HL4, HL5, HL6, HL7	Esposito et al. (2019)
C3		LL1, LL2, LL3, LL4, HL1, HL2, HL3, HL4, HL5, HL6, HL7,	Tung (2016)	Leonardo		LL1, LL2, LL3, LL4, HL1, AL1, AL3, AL4, AL5, AL7	Kumazaki et al. (2017)	Tabby cat, Hasbro		LL1, LL2, LL3, LL4, AL1, AL2, AL3, AL4, AL5, AL6	Löffler et al. (2020)
Cosmobot		LL1, LL2, LL3, LL4, ML7	von der Pütten and Krämer (2012)	M³-Synchy		LL1, LL2, LL3, LL4, HL1, HL3, HL5, ML3	Dinet and Vivian (2014)	Tapia, MJI		ML4, ML5, ML6	Tapia. MJI Inc. Retrieved from https://mjirobotics.co.jp/en/
Cozmo (Anki robotics)		ML2, ML4, ML5, ML6, ML7	Ciardo et al. (2018)	MAHRU-II		LL1, LL3, LL4, ML5, ML7	Bernotat and Eyssel (2018)	Teddy-bear (Fujitsu)		LL1, LL4, AL1, AL2, AL3, AL4, AL5	Dinet and Vivian (2014)
CuDDler		LL1, LL2, LL3, LL4, AL1, AL2, AL3, AL4, AL5, AL6, AL7	Tan et al. (2013)	Meka		LL1, LL4, ML1, ML3, ML5, ML7	Retrieved from Encyclo-pædia Britannica ImageQuest.	Teksta		LL1, LL4, AL1, AL2, AL3, AL4, AL5, AL7, ML7	Howard and Vick (2010)

Custom orb-like robot		ML1, ML4, ML5, ML6, ML7	Konok et al. (2018)	NAO		LL1, LL2, LL3, LL4	van Straten et al. (2018)	the android		LL4, ML2, ML4, ML5, ML7	Katz and Halpern (2014)
Erica (Hiroshi Ishiguro and Dylan Glas)		LL1, LL2, LL3, LL4, HL1, HL2, HL3, HL4, HL5, HL6, HL7	Esposito et al. (2018)	NAO with clothes		LL1, LL2, LL3, LL4, HL7	Phillips et al. (2018)	Wakamaru		LL1, LL4, ML2, ML7	Phillips et al. (2018)
Ethon 2		ML1, ML3, ML4, ML5, ML6, ML7	Kim et al. (2019)	Nexi		LL1, LL2, LL3, LL4, HL1, HL5, HL6, ML1, ML3, ML7	Osada et al. (2016)	Zenbo (ASUS)		LL1, LL2, LL3, ML1, ML6, ML7	Chien et al. (2019)
Flobi		LL1, LL2, LL3, HL1, HL2, HL3, HL5, HL6	Phillips et al. (2018)	PaPeRo		LL1, LL4, ML1, ML5, ML7,	Löffler et al. (2020)	Zeno, the Hanson Robokind		LL1, LL2, LL3, LL4, HL1, HL2, HL3, HL4, HL5, HL6, ML7	Cameron et al. (2018)
Floka (Flobi head on Meka torso)		LL1, LL2, LL3, LL4, HL1, HL2, HL3, HL5, HL6, ML1, ML3, ML7	Bernotat and Eyssel (2018)	Parle, MIT		LL1, LL2, LL3, LL4, AL5, AL6	Löffler et al. (2020)				

Note. Majority of images found via source used in literature, others from https://quest.eb.com/ images database provided by Tilburg University

Appendix C: *occurrences of traits by subscale*

	LL 1	LL 2	LL 3	LL 4	HL 1	HL 2	HL 3	HL 4	HL 5	HL 6	HL 7	AL 1	AL 2	AL 3	AL 4	AL 5	AL 6	AL 7	ML 1	ML 2	ML 3	ML 4	ML 5	ML 6	ML 7
Liking	20	14	16	14	5	3	8	6	5	5	5	10	9	9	11	9	3	8	7	0	3	1	2	3	7
Friendliness	5	3	4	5	1	0	1	0	0	0	0	4	3	3	3	5	1	4	0	1	0	1	1	1	2
Kindness	2	0	0	2	0	0	0	0	0	0	0	1	1	1	1	1	0	1	0	0	1	0	1	0	1
Relaxedness	2	2	2	0	2	0	0	0	0	0	2	2	2	2	4	4	2	2	0	0	0	0	0	0	0
Niceness	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0
Companionship	5	4	3	1	1	0	1	1	0	1	2	2	1	3	1	1	1	2	1	1	2	1	2	4	3
Pleasantness	5	5	5	4	1	0	0	0	0	0	1	0	0	0	1	1	0	0	3	0	0	0	0	0	0
Enjoyment	8	4	5	7	3	1	1	1	1	2	0	6	4	6	4	5	1	5	2	1	0	0	0	1	4
Attractiveness	4	3	3	3	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1	2
Anxiety	2	2	2	1	1	0	0	0	0	0	1	0	0	0	1	1	0	0	2	0	1	1	1	1	0
Likeability	9	7	7	7	3	2	3	1	2	2	2	4	4	4	5	5	2	5	1	0	0	0	1	0	3

References

American Psychological Association. (n.d.). Just-world hypothesis. In *APA dictionary of psychology*. Retrieved October 21, 2020, from https://dictionary.apa.org/affect

Austermann, A., Yamada, S., Funakoshi, K., & Nakano, M. (2010). How do users interact with a pet-robot and a humanoid. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems* (pp. 3727-3732).

Baraka, K., Alves-Oliveira, P., & Ribeiro, T. (2020). An extended framework for characterizing social robots. In *Human-Robot Interaction* (pp. 21-64). Springer, Cham.

Barco, A., de Jong, C., Peter, J., Kühne, R., & van Straten, C. L. (2020). Robot Morphology and Children's Perception of Social Robots: An Exploratory Study. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 125-127).

Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M., & Sabanovic, S. (2020). Human-Robot Interaction – An Introduction. Cambridge: *Cambridge University Press*.

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

Bernotat, J., & Eyssel, F. (2018). Can (‘t) Wait to Have a Robot at Home?-Japanese and German Users' Attitudes Toward Service Robots in Smart Homes. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 15-22). IEEE.

Bortot, D., Born, M., & Bengler, K. (2013). Directly or on detours? How should industrial robots approximate humans?. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 89-90). IEEE.

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

Cameron, D., Millings, A., Fernando, S., Collins, E. C., Moore, R., Sharkey, A., ... & Prescott, T. (2018). The effects of robot facial emotional expressions and gender on child–robot interaction in a field study. *Connection science*, 30(4), 343-361.

Carpenter, J. (2016). Culture and human-robot interaction in militarized spaces: A war story. *Routledge*.

Castro-González, Á., Admoni, H., & Scassellati, B. (2016). Effects of form and motion on judgments of social robots ′animacy, likability, trustworthiness and unpleasantness. *International Journal of Human-Computer Studies*, 90, 27-38.

Caudwell, C., Lacey, C., & Sandoval, E. B. (2019). The (Ir) relevance of Robot Cuteness: An Exploratory Study of Emotionally Durable Robot Design. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction* (pp. 64-72).

Chu, L., Chen, H. W., Cheng, P. Y., Ho, P., Weng, I. T., Yang, P. L., ... & Fung, H. H. (2019). Identifying features that enhance older adults' acceptance of robots: A mixed methods study. *Gerontology*, 65(4), 441-450.

Chung, K. M., & Shin, D. H. (2015). How Anthropomorphism Affects Human Perception of Color-Gender-Labeled Pet Robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts* (pp. 75-76).

Collins, E. C., Prescott, T. J., Mitchinson, B., & Conran, S. (2015). MIRO: a versatile biomimetic edutainment robot. In *Proceedings of the 12th International Conference on Advances in Computer Entertainment Technology* (pp. 1-4).

Cruz-Sandoval, D., Penaloza, C. I., Favela, J., & Castro-Coronel, A. P. (2018). Towards social robots that support exercise therapies for persons with dementia. In *Proceedings of the 2018 ACM*

International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (pp. 1729-1734).

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

Dautenhahn, K. (2007). Methodology & themes of human-robot interaction: A growing research field. *International Journal of Advanced Robotic Systems*, 4(1), 15.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 319–339.

Davis, F. D. (1991). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International journal of man-machine studies*, 38(3), 475-487.

Desai, M., Stubbs, K., Steinfeld, A., & Yanco, H. (2009). Creating trustworthy robots: Lessons and inspirations from automated systems.

Dinet, J., & Vivian, R. (2014). Exploratory investigation of attitudes towards assistive robots for future users. *Le travail humain*, 77(2), 105-125.

Ekbia, H. R. (2008). Artificial dreams: the quest for non-biological intelligence (Vol. 200, No. 8). Cambridge: *Cambridge University Press*.

Esposito, A., Amorese, T., Cuciniello, M., Pica, I., Riviello, M. T., Troncone, A., ... & Esposito, A. M. (2019). Elders prefer female robots with a high degree of human likeness. In *2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT)* (pp. 243-246). IEEE.

Eyssel, F., Kuchenbrandt, D., & Bobinger, S. (2011). 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).

Feingold-Polak, R., Elishay, A., Shahar, Y., Stein, M., Edan, Y., & Levy-Tzedek, S. (2018). Differences between young and old users when interacting with a humanoid robot: A qualitative usability study. Paladyn, *Journal of Behavioral Robotics*, 9(1), 183-192.

Feng, S., Wang, X., Wang, Q., Fang, J., Wu, Y., Yi, L., & Wei, K. (2018). The uncanny valley effect in typically developing children and its absence in children with autism spectrum disorders. *PloS* one, 13(11), e0206343.

Fernaeus, Y., Håkansson, M., Jacobsson, M., & Ljungblad, S. (2010). How do you play with a robotic toy animal? A long-term study of Pleo. *In Proceedings of the 9th international Conference on interaction Design and Children* (pp. 39-48).

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

Fraune, M. R., Sherrin, S., Sabanović, S., & Smith, E. R. (2015). Rabble of robots effects: Number and type of robots modulates attitudes, emotions, and stereotypes. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 109-116).

de Gauquier, L., Cao, H. L., Gomez Esteban, P., De Beir, A., van de Sanden, S., Willems, K., ... & Vanderborght, B. (2018). Humanoid robot pepper at a Belgian chocolate shop. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 373-373).

Gerłowska, J., Skrobas, U., Grabowska-Aleksandrowicz, K., Korchut, A., Szklener, S., Szczęśniak-Stańczyk, D., ... & Rejdak, K. (2018). Assessment of perceived attractiveness, usability, and societal impact of a multimodal robotic assistant for aging patients with memory impairments. *Frontiers in neurology*, 9, 392.

Ghazali, A. S., Ham, J., Barakova, E. I., & Markopoulos, P. (2018). Effects of robot facial characteristics and gender in persuasive human-robot interaction. *Frontiers in Robotics and AI*, *5*, 73.

de Graaf, M. M., Allouch, S. B., & van Dijk, J. A. (2016). Long-term evaluation of a social robot in real homes. *Interaction studies*, *17*(3), 462-491.

de Graaf, M., Allouch, S. B., & Van Dijk, J. (2017). Why do they refuse to use my robot?: Reasons for non-use derived from a long-term home study. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction* (HRI (pp. 224-233). IEEE.

de Graaf, M. M., Allouch, S. B., & van Dijk, J. A. (2019). Why would I use this in my home? A model of domestic social robot acceptance. *Human–Computer Interaction*, 34(2), 115-173.

Han, N. R., Baek, T. H., Yoon, S., & Kim, Y. (2019). Is that coffee mug smiling at me? How anthropomorphism impacts the effectiveness of desirability vs. feasibility appeals in sustainability advertising. *Journal of Retailing and Consumer Services*, 51, 352-361.

Haslam, N., Loughnan, S., Kashima, Y., & Bain, P. (2008). Attributing and denying humanness to others. *European review of social psychology*, *19*(1), 55-85.

Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2008). The influence of social presence on acceptance of a companion robot by older people.

Heerink, M., Kröse, B., Wielinga, B., & Evers, V. (2008). Enjoyment intention to use and actual use of a conversational robot by elderly people. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction* (pp. 113-120).

Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2010). Assessing acceptance of assistive social agent technology by older adults: the almere model. *International journal of social robotics*, 2(4), 361-375.

Hegel, F., Lohse, M., & Wrede, B. (2009). Effects of visual appearance on the attribution of applications in social robotics. In *RO-MAN 2009-The 18th IEEE International symposium on robot and human interactive communication* (pp. 64-71). IEEE.

Howard, L., & Vick, S. J. (2010). Does it bite? The role of stimuli characteristics on preschoolers' interactions with robots, insects and a dog. *Anthrozoös*, 23(4), 397-413.

Jentsch, F. (2016). Human-robot interactions in future military operations. *CRC Press*.

Joinson, A. N. (2002). Understanding the Psychology of Internet Behavior: Virtual World, Real Lives, Palgrave Macmillan.

Kaminski, M., Pellino, T., & Wish, J. (2002). Play and pets: The physical and emotional impact of child-life and pet therapy on hospitalized children. *Children's health care*, 31(4), 321-335.

Karabegović, I., Doleček, V., & Husak, E. (2011). Analysis of the industrial robots in various production processes in the world. *International Review of Mechanical Engineering*, 5(7), 1272-1277.

Katz, J. E., & Halpern, D. (2014). Attitudes towards robots suitability for various jobs as affected robot appearance. *Behaviour & Information Technology*, 33(9), 941-953.

Khosravi, P., & Ghapanchi, A. H. (2016). Investigating the effectiveness of technologies applied to assist seniors: A systematic literature review. *International journal of medical informatics*, 85(1), 17-26.

Kim, A., Han, J., Jung, Y., & Lee, K. (2013). The effects of familiarity and robot gesture on user acceptance of information. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 159-160). IEEE.

Kim, S. Y., Schmitt, B. H., & Thalmann, N. M. (2019). Eliza in the uncanny valley: anthropomorphizing consumer robots increases their perceived warmth but decreases liking. *Marketing letters*, *30*(1), 1-12.

Kipp, A., & Kummert, F. (2016). " I know how you performed!" Fostering Engagement in a Gaming Situation Using Memory of Past Interaction. In *Proceedings of the Fourth International Conference on Human Agent Interaction* (pp. 281-288).

Kjellsson, J., Bryne, G., Scheible, G., Frey, J. E., Strand, M., & Gentzell, T. (2007). Tool for an Industrial Robot *U.S. Patent Application* No. 10/583,387.

Konok, V., Korcsok, B., Miklósi, Á., & Gácsi, M. (2018). Should we love robots?–The most liked qualities of companion dogs and how they can be implemented in social robots. *Computers in Human Behavior*, 80, 132-142.

Kumazaki, H., Warren, Z., Muramatsu, T., Yoshikawa, Y., Matsumoto, Y., Miyao, M., ... & Mimura, M. (2017). A pilot study for robot appearance preferences among high-functioning individuals with autism spectrum disorder: Implications for therapeutic use. *PloS one*, 12(10), e0186581.

Ladwig, R. C., & Ferstl, E. C. (2018). What's in a name? an online survey on gender stereotyping of humanoid social robots. In *Proceedings of the 4th Conference on Gender & IT* (pp. 67-69).

Lai, I. K. W., & Liu, Y. (2020). The effects of content likeability, content credibility, and social media engagement on users' acceptance of product placement in mobile social networks. *Journal of Theoretical and Applied Electronic Commerce Research*, 15(3), 1-19.

Lee, S. L., Lau, I. Y. M., & Hong, Y. Y. (2011). Effects of appearance and functions on likability and perceived occupational suitability of robots. *Journal of Cognitive Engineering and Decision Making*, 5(2), 232-250.

Li, D., Rau, P. P., & Li, Y. (2010). A cross-cultural study: Effect of robot appearance and task. *International Journal of Social Robotics*, 2(2), 175-186.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., ... & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of clinical epidemiology*, 62(10), e1-e34.

Ljungblad, S., Kotrbova, J., Jacobsson, M., Cramer, H., & Niechwiadowicz, K. (2012). Hospital robot at work: something alien or an intelligent colleague?. In *Proceedings of the ACM 2012 conference on computer supported cooperative work* (pp. 177-186).

Löffler, D., Dörrenbächer, J., & Hassenzahl, M. (2020). The Uncanny Valley Effect in Zoomorphic Robots: The U-Shaped Relation Between Animal Likeness and Likeability. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 261-270).

Logan, D. E., Breazeal, C., Goodwin, M. S., Jeong, S., O'Connell, B., Smith-Freedman, D., ... & Weinstock, P. (2019). Social robots for hospitalized children. *Pediatrics*, 144(1), e20181511.

Lohse, M., Hegel, F., Swadzba, A., Rohlfing, K., Wachsmuth, S., & Wrede, B. (2007). What can I do for you? Appearance and application of robots. In *Proceedings of AISB* (Vol. 7, pp. 121-126).

Louie, W. Y. G., McColl, D., & Nejat, G. (2014). Acceptance and attitudes toward a human-like socially assistive robot by older adults. *Assistive Technology*, *26*(3), 140-150.

MacDorman, K. F., Minato, T., Shimada, M., Itakura, S., Cowley, S., & Ishiguro, H. (2005). Assessing human likeness by eye contact in an android testbed. In *Proceedings of the XXVII annual meeting of the cognitive science society* (pp. 21-23).

Matsui, T., & Yamada, S. (2019). The Design Method of the Virtual Teacher. In *Proceedings of the 7th International Conference on Human-Agent Interaction* (pp. 97-101).

McCloskey, D. W. (2006). The importance of ease of use, usefulness, and trust to online consumers: An examination of the technology acceptance model with older customers. *Journal of Organizational and End User Computing (JOEUC)*, 18(3), 47-65.

McGlynn, S. A., Kemple, S., Mitzner, T. L., King, C. H. A., & Rogers, W. A. (2017). Understanding the potential of PARO for healthy older adults. *International journal of human-computer studies*, 100, 33-47.

Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses. *Journal of Marketing Research*, 56(4), 535-556.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2), 98-100.

Nomura, T., Suzuki, T., Kanda, T., Han, J., Shin, N., Burke, J., & Kato, K. (2008). What people assume about humanoid and animal-type robots: cross-cultural analysis between Japan, Korea, and the United States. *International Journal of Humanoid Robotics*, 5(01), 25-46.

Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Measurement of negative attitudes toward robots. *Interaction Studies*, 7(3), 437-454.

Okanda, M., Taniguchi, K., & Itakura, S. (2019). The Role of Animism Tendencies and Empathy in Adult Evaluations of Robot. *In Proceedings of the 7th International Conference on Human-Agent Interaction* (pp. 51-58).

Osawa, H., Ohmura, R., & Imai, M. (2009). Using attachable humanoid parts for realizing imaginary intention and body image. *International Journal of Social Robotics*, 1(1), 109-123.

Overgoor, C., & Funk, M. (2018). IdleBot: Exploring the design of serendipitous artifacts. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems* (pp. 105-110).

Papadopoulos, I., & Koulouglioti, C. (2018). The influence of culture on attitudes towards humanoid and animal‐like robots: an integrative review. *Journal of Nursing Scholarship*, 50(6), 653-665.

Park, E., Kong, H., Lim, H. T., Lee, J., You, S., & del Pobil, A. P. (2011). The effect of robot's behavior vs. appearance on communication with humans. In *Proceedings of the 6th international conference on Human-robot interaction* (pp. 219-220).

Phillips, E., Zhao, X., Ullman, D., & Malle, B. F. (2018). 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).

Pitsch, K., Wrede, S., Seele, J. C., & Süssenbach, L. (2011). Attitude of german museum visitors towards an interactive art guide robot. In Proceedings of the 6th international conference on Human-robot interaction (pp. 227-228).

Prakash, A., Kemp, C. C., & Rogers, W. A. (2014). Older adults' reactions to a robot's appearance in the context of home use. In 2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 268-269). IEEE.

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

Richards, N. M., & Smart, W. D. (2016). How should the law think about robots?. In *Robot law*. Edward Elgar Publishing.

Robbins, T. L., & DeNisi, A. S. (1994). A closer look at interpersonal affect as a distinct influence on cognitive processing in performance evaluations. *Journal of Applied Psychology*, 79(3), 341.

Robins, B., Dautenhahn, K., & Dubowski, J. (2006). Does appearance matter in the interaction of children with autism with a humanoid robot?. *Interaction studies*, 7(3), 479-512.

Robinson, H., MacDonald, B., & Broadbent, E. (2014). The role of healthcare robots for older people at home: A review. *International Journal of Social Robotics*, 6(4), 575-591.

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

Schweinberger, S. R., Pohl, M., & Winkler, P. (2020). Autistic traits, personality, and evaluations of humanoid robots by young and older adults. *Computers in Human Behavior*, 106, 106256.

Schwind, V., Leicht, K., Jäger, S., Wolf, K., & Henze, N. (2018). Is there an uncanny valley of virtual animals? A quantitative and qualitative investigation. *International Journal of Human-Computer Studies*, 111, 49-61.

Sciutti, A., Mara, M., Tagliasco, V., & Sandini, G. (2018). Humanizing human-robot interaction: On the importance of mutual understanding. *IEEE Technology and Society Magazine*, 37(1), 22-29.

Shibata, T. (2004). An overview of human interactive robots for psychological enrichment. Proceedings of the IEEE, 92(11), 1749-1758.

Sinatra, A. M., Sims, V. K., Chin, M. G., & Lum, H. C. (2012). If it looks like a dog: The effect of physical appearance on human interaction with robots and animals. *Interaction Studies*, 13(2), 235-262.

van Straten, C. L., Smeekens, I., Barakova, E., Glennon, J., Buitelaar, J., & Chen, A. (2018). Effects of robots' intonation and bodily appearance on robot-mediated communicative treatment outcomes for children with autism spectrum disorder. *Personal and Ubiquitous Computing*, 22(2), 379-390.

Syrdal, D. S., Dautenhahn, K., Koay, K. L., & Walters, M. L. (2009). The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study. Adaptive and emergent behaviour and complex systems.

Tan, Y. K., Wong, A., Wong, A., Dung, T. A., Tay, A., Kumar, D. L., ... & Tay, B. (2013). Evaluation of the pet robot CuDDler using godspeed questionnaire. In *International Conference on Smart Homes and Health Telematics* (pp. 102-109). Springer, Berlin, Heidelberg.

Teo, T. (2011). Factors influencing teachers' intention to use technology: Model development and test. *Computers & Education*, 57(4), 2432-2440.

Thunberg, S., Thellman, S., & Ziemke, T. (2017). Don't Judge a Book by its Cover: A Study of the Social Acceptance of NAO vs. Pepper. In *Proceedings of the 5th International Conference on Human Agent Interaction* (pp. 443-446).

Trovato, G., Lopez, A., Paredes, R., Quiroz, D., & Cuellar, F. (2019). Design and Development of a Security and Guidance Robot for Employment in a Mall. *International Journal of Humanoid Robotics*, 16(05), 1950027.

Tschöpe, N., Reiser, J. E., & Oehl, M. (2017). Exploring the uncanny valley effect in social robotics. In Proceedings of the Companion of the 2017 ACM/IEEE *International Conference on Human-Robot Interaction* (pp. 307-308).

Tung, F. W. (2016). Child perception of humanoid robot appearance and behavior. *International Journal of Human-Computer Interaction*, *32*(6), 493-502.

Tzafestas, S. G. (2016). Zoomorphic sociorobots. In Sociorobot World (pp. 155-173). Springer, Cham.

Ventre-Dominey, J., Gibert, G., Bosse-Platiere, M., Farnè, A., Dominey, P. F., & Pavani, F. (2019). Embodiment into a robot increases its acceptability. *Scientific reports*, 9(1), 1-10.

Vincent, J. (2020, May 8). Spot the robot is reminding parkgoers in Singapore to keep their distance from one another. [Blog post]. Retrieved from https://www.theverge.com/2020/5/8/21251788/spot-boston-dynamics-robot-singapore-park-social-distancing

Voth, D. (2004). A new generation of military robots. *IEEE Intelligent Systems*, *19*(4), 2-3.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, 54(6), 1063.

Weiss, A., Bader, M., Vincze, M., Hasenhütl, G., & Moritsch, S. (2014). Designing a service robot for public space: an" action and experiences"-approach. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 318-319).

Woods, S. (2006). Exploring the design space of robots: Children's perspectives. *Interacting with Computers*, *18*(6), 1390-1418.

Zheng, C. Y., Lacey, C., & Paterson, M. (2020, March). Affect and Embodiment in HRI. In Companion of the 2020 ACM/IEEE *International Conference on Human-Robot Interaction* (pp. 667-668).