8

How People Perceive Robots

What is covered in this chapter:

- What different social science theories say about how people form perceptions about others.
- How we understand anthropomorphism of robots based on prior social science literature.
- How anthropomorphism makes us see robots as uncanny, trustworthy, or likable.

Imagine you enter a university building, a retail store, an elderly care facility or—if you are really daring—a friend's home. A social robot approaches you. How do you feel, and what do you think? Of course, your impression of the robot will depend on the specific context and use case, like the ones we have just mentioned. At the same time, the way you feel and think about the robot also strongly depends on the robot, its features, and its functions. It will also depend on your prior knowledge and experiences that you may associate with the robot—a robot whose body is covered in fuzzy fur might suggest to you it is ready for a pat and a hug, whereas a robot with a chef's hat on may make you think a delicious meal is in the works. From research on human–human impression formation, we know that people form impressions readily and nearly automatically based on a variety of observable cues (Macrae and Quadflieg, 2010).

Earlier work in human–computer interaction (HCI) shows that we seem to form quick first impressions about robots (see Chapter 4 on robot design). As we learned there, we attribute humanlike traits, emotions, presence of mind, and other characteristics to nonhuman entities ranging from computers to virtual agents and social robots.

This chapter discusses how people form an impression of a robot; its paradigm is primarily psychological. Section 8.1 covers the general principles of impression formation; Section 8.2 specifically covers anthropomorphism as a form of impression formation. Section 8.3 discusses the kinds of measurements that have been used to evaluate anthropomorphization. Finally, Section 8.4 covers some of the main consequences of anthropomorphism, such as trust, acceptance, and liking.

This material has been published by Cambridge University Press as Human-Robot Interaction by
Christoph Bartneck, Tony Belpaeime, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Sabanovic.
ISBN: 9781009424233 (https://www.cambridge.org/9781009424233).

8.1 Impression formation

People usually form impressions quickly and automatically—snap judgments about a target can be made within milliseconds. In the following section, we describe in some detail a framework psychologists use to explain how such perceptions are formed, called the *dual-process model of impression formation*.

8.1.1 Dual-process models of impression formation

Scholars theorize that people process information—as required in order to make decisions, form impressions, or guide behavior—in two ways. One way is automatic, intuitive, and quick; the other is more deliberate, conscious, and slow (Evans and Stanovich, 2013; Smith and DeCoster, 2000). To describe how these two ways of processing information work, scholars talk about dual-process models. The two ways of processing are sometimes labeled *system 1* versus *system 2* (Kahneman, 2011), *associative* versus *rule-based* (Sloman, 1996), and *automatic* versus *controlled* processing (Shiffrin and Schneider, 1984). Whatever their name, the dual-process model proposes that the primary way in which people process input and construct a response (whether that is an affective reaction, a decision, or a behavior) is automatic, with the possibility of tweaking this initial response through more deliberate and conscious processing.

As the name implies, automatic processing may occur outside of an individual's conscious awareness, based on the activation of cognitive and affective responses (Evans, 2008). Many such associations are formed through previous experience (McLaren et al., 2014). For example, if you have watched a large number of sci-fi movies that have portrayed robots as threatening villains, like the Terminator, you will most likely associate a robot you encounter for the first time with something rather negative. If, on the other hand, your initial experiences with robots are as friendly members of the family, like Doraemon or Astro Boy, then your first reaction to a robot might be positive.

This automatic processing forms the initial impression and sets the tone for what our intuitive expectations of a robot are. In contrast, deliberate processing is more conscious and intentional. According to some dual-process models, the deliberate system builds on the results from automatic processing, resulting in sequential processing (Evans and Stanovich, 2013). Others have proposed that the two modes of processing work in parallel and that the outcome is constructed from the output of both (Smith and DeCoster, 2000). Either way, it is important to realize that although the deliberate form of processing is conscious, this does not imply that when we use deliberate processing, we are perfectly rational or objective. We are just making a conscious effort at a task, whether that is figuring out the answer to an exam question or forming an opinion on how trustworthy a robot is. Deliberate processing takes effort and mental capacity, and therefore it only happens when we have the motivation and capability to do it (Evans, 2008; Złotowski et al., 2018).

Thus, when you run into a new robot, like in the example at the start of this chapter, you may form an instantaneous, automatic impression. If you have the motivation and mental capacity, you may also engage in more deliberate processing of how you feel and what you think about it. At times, these two impression-related processes may result in differing implicit and explicit attitudes; for example, de Graaf and colleagues (2016) have shown that people may actually be more negative about robots in implicit measures than in the impressions that they consciously and explicitly express.

In Section 4.2 (Chapter 4), we saw that in addition to a like/dislike distinction, impression formation can also entail people attributing essentially human features and characteristics to other entities (including robots). These characteristics include intentions, emotions, and dimensions of mind perception (e.g., agency and experience) (Gray et al., 2007), to name but a few. This attribution of traits and characteristics is called *anthropomorphization* (Epley et al., 2007, 2008; Eyssel, 2017). It has been proposed that this process can also be conceptualized in terms of a dual-process model (Złotowski et al., 2018; Urquiza-Haas and Kotrschal, 2015).

8.2 Anthropomorphism

In this section, we will discuss several theoretical frameworks that have been proposed to explain anthropomorphism (i.e., perceiving and judging a humanlike form), as well as the process of anthropomorphization (i.e., the attribution of humanlike characteristics to nonhuman entities).

8.2.1 Psychological anthropomorphism

In the early years of human–robot interaction (HRI) research, the conceptualization of what anthropomorphism is and entails was fairly limited, with anthropomorphism—at that time—being most often equated to humanlikeness in appearance, in line with the engineering approach to the concept. Thus, early work on anthropomorphism mainly focused on assessing the perceived appearance of the robot.

Going beyond the classical engineering perspective, recent theorizing in psychology has provided a complementary perspective on the nature of the phenomenon. The theoretical framework proposed by Nicholas Epley and colleagues (2007) has been influential in both psychology and robotics and serves to broaden our understanding of the notion of anthropomorphism, its causes, and its consequences. Whereas *anthropomorphism* until then had mainly referred to humanlike form, Epley and colleagues suggested that the phenomenon extends beyond the observable and includes cognitive and motivational processes—hence creating the notion of psychological anthropomorphism. Specifically, they suggested three core factors that determine anthropomorphic inferences about nonhuman entities: effectance motivation, sociality motivation, and elicited agent knowledge. Let us introduce these concepts briefly.

First, effectance motivation concerns our desire to explain and understand the behavior of others as social actors. This motivation might be activated when people are confronted with an unfamiliar interaction partner that they are unsure about how to deal with. Most people are still relatively unfamiliar with robots as social interaction partners, so it is easy to imagine how approaching the robot as if it had humanlike characteristics would function as a default option. People might therefore attribute humanlike characteristics to robots to psychologically regain control over the novel situation they find themselves in. In this case, anthropomorphization can reduce the stress and anxiety associated with HRI. Effectance motivation might explain the intriguing finding that robot movement, whether or not the robot has an explicitly social role, is commonly interpreted by people as a social cue (Erel et al., 2019).

Second, anthropomorphization of robots could also be caused by a sociality motivation, particularly in people who lack social connections. In this case, people may turn to nonhuman entities as social interaction partners to address their feelings of situational or chronic loneliness. Supporting this idea, previous research has shown that people who have been made to feel lonely in an experimental situation or who are chronically lonely anthropomorphize robots to a greater extent than people who are sufficiently socially connected (Eyssel and Reich, 2013).

Lastly, *elicited agent knowledge* refers to the way in which people use their commonsense understanding of social interactions and actors to understand robots. For example, Powers et al. (2005) showed that people who considered women to be more knowledgeable about dating norms behaved as if male and female robots also had differing competencies regarding dating. For instance, they used more time and words to explain dating norms to a male robot. This factor in particular can be used to guide the design and technical implementation of social robots for various tasks.

These three determinants shed light on the psychological mechanisms of people humanizing nonhuman entities. This includes the attribution of emotions, intentions, typical human traits, or other essentially human characteristics to any type of nonhuman entity, real or imagined (Epley et al., 2007). The basic assumption is that people use self-related or anthropocentric knowledge structures to make sense of the nonhuman things—or in our case, robots—around them. Human resemblance in appearance and behavior triggers anthropomorphic judgments, and people may thus attribute traits and emotions to a technical system despite the fact that the system, indeed, is merely a piece of technology (see Figure 8.1). This, in turn, not only affects the social perception of robots but also the actual behavior displayed toward them during an interaction.

8.2.2 The process of anthropomorphization

Early on in the history of human–agent research, the prominent media equation hypothesis was formulated by Reeves and Nass (1996), who demonstrated in an array of HCI studies that people readily ascribe humanlike traits to

This material has been published by Cambridge University Press as Human-Robot Interaction by
Christoph Bartneck, Tony Belpaeime, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Sabanovic.
ISBN: 9781009424233 (https://www.cambridge.org/9781009424233).

Figure 8.1 The Telenoid telepresence robot's (2010–present) design uses abstracted humanlike features to inspire anthropomorphization while also aiming to let the unique identity of the person interacting through the robot to be perceived by the person interacting with it. (Source: Photo by Selma Šabanović)

machines. Back then, their research merely involved personal computers because social robots were not yet developed enough to serve as research platforms in such interactive setups. However, later on, the ideas from the so-called "computers as social actors" (CASA) approach were translated to the domain of social robots and have been validated in extensive empirical research ever since. Research on the CASA approach touches on the notion of automaticity of social judgments about technologies. Likewise, the model by Złotowski et al. (2018) differentiates automatic and controlled components related to forming anthropomorphic inferences about robots.

As mentioned earlier, we can distinguish two processes, system 1 and system 2, that supposedly are involved in the anthropomorphization of robots. According to this, people engage in fast, initial snap judgments of a given target—"Is the target humanlike or not?" Following that, more deliberate, controlled processes can alter the initial judgment from system 1. Złotowski et al. (2018) have coined the notions of *implicit* versus *explicit* anthropomorphism to refer to these two distinct outcomes of system 1 and system 2.

Other models of anthropomorphism have shed light on the time scale of the process of attribution, differentiating different phases of anthropomorphization—namely, the pre-initial stage, the initialization stage, familiarization, and finally, stabilization (Lemaignan et al., 2014a). According to this model, individuals form an a priori impression of a given entity before

This material has been published by Cambridge University Press as Human-Robot Interaction by
Christoph Bartneck, Tony Belpaeime, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Sabanovic.
ISBN: 9781009424233 (https://www.cambridge.org/9781009424233).

the first encounter, and they might revise and extend these judgments in the subsequent initialization phase. Once a person hits the familiarization stage, a more realistic impression of the agent can be formed due to exposure to it and experience with it. As a consequence, anthropomorphic inferences likely decrease in this stage. Finally, people come to a comprehensive judgment of the agent of interest in the stabilization phase. Such a conceptualization thus integrates initial snap judgments with more deliberate considerations about the humanlike nature of a given entity.

This model was further complemented by the original authors when they introduced a three-stage model to reflect the cognitive processes involved in anthropomorphization (Lemaignan et al., 2014b). That is, phase I involves automatic evaluations without necessarily involving actual HRI. In phase II, people get to interact with the entity of interest, and based on this, they create a mental model of the robot that reflects its real or imagined functionalities or characteristics. This mental model is finally adapted as a function of actual "contextualized" interaction, that is, based on meaningful interactions with the robot, for example, in the user's home context (Lemaignan et al., 2014b).

Above and beyond the socio-cognitive perspective, the integrative framework of anthropomorphism (IFA) by Spatola et al. (2022) is a model that takes individual and cultural variables into consideration. For instance, an individual's tendency to endorse spiritualism, mentalization, and humanization might be affected by the cultural context. For example, Japanese culture features animism, the belief that things such as mountains, statues, or trees have a spiritual essence. This is also believed to spill over to robotics, with robots being given certain spiritual qualities.

8.3 Measuring anthropomorphization

8.3.1 Explicit measurements

Tightly related to the theorizing on what anthropomorphism entails is the question of operationalization: How does one measure anthropomorphization? In order to solve this issue, one needs to clearly define what anthropomorphism is and what it is not so that a measurement can be constructed that targets anthropomorphization and nothing else. In short, we need to know not just why and when people anthropomorphize but also how.

Psychological anthropomorphism has been measured under many names. Common terms include *mental state ascription/attribution*, *mind perception*, and *theory of mind*. Although all these terms have different connotations, they are referring to the same underlying phenomenon (Thellman et al., 2022).

Focusing on agents in general rather than robots specifically, Gray et al. (2007) proposed two dimensions of mind perception: agency and experience. *Agency* refers to the ability to, for example, plan, think, and exert self-control, whereas *experience* entails the ability to, for example, have hopes and dreams, feel emotions, and have a personality. These measures of mind perceptions have been adapted to research on social robots by Eyssel and

This material has been published by Cambridge University Press as Human-Robot Interaction by
Christoph Bartneck, Tony Belpaeime, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Sabanovic.
ISBN: 9781009424233 (https://www.cambridge.org/9781009424233).

Loughnan (2013), who combined it with a measure of racism. White American participants were asked to evaluate a robot that had been given either a White or a Black skin color. An interesting pattern emerged in which participants' level of racism did not lower the overall level of mind attribution but lowered perceived agency and heightened experience.

These two scales of mind attribution bear some semblance to the warmth and competence scales that appear to be the key dimensions of social judgments in human cognition. Accordingly, Cuddy et al. (2008) have posited that people initially judge a person's or group's perceived warmth (e.g., tolerant, warm, good-natured, sincere) and then determine the target's competence (e.g., competent, confident, independent, competitive, intelligent) (Fiske et al., 202, 2007; Wojciszke, 2005). Recently, the primacy-of-warmth assumption has been challenged in replication research (Nauts et al., 2014), but the basic tenets of warmth and competence (or agency and communion) as core dimensions of social evaluation still hold (Abele et al., 2016). Not surprisingly, HRI researchers also inquire about the warmth and competence of social robots (Eyssel and Hegel, 2012; Carpinella et al., 2017; Christoforakos et al., 2021; Mieczkowski et al., 2019).

HRI researchers have also applied the principles of dehumanization and infrahumanization theory to robots. Dehumanization is the process in which humans perceive others as somehow being "less" human by reducing the ascription of human traits (Haslam, 2006; Haslam and Loughnan, 2014; Loughnan and Haslam, 2007). The theory differentiates between uniquely human and human-nature traits (Haslam, 2006), with the first relating to capabilities that supposedly set humans apart from other animals (e.g., rationality, civilization, and refinement) and the latter being qualities that, although shared with other animals, still are considered fundamental to being human (e.g., curiosity, emotionality, and warmth) (Haslam et al., 2008). In intergroup research, these traits have been used to assess dehumanization of other humans as animal-like (denial of uniquely human traits) or machinelike (through denial of human-nature traits). In turn, in the context of nonhuman entities, these traits have been applied to measure the anthropomorphism of social robots (Eyssel et al., 2011; Spatola et al., 2021).

Infrahumanization (Leyens et al., 2000; Leyens, 2009) is a more subtle form of dehumanization. Rather than blatantly reducing someone's ascribed ability to experience emotion or engage in rational thought, perceived humanness is compromised through a lower ascription of secondary emotions, which are considered as more exclusive to humans (e.g., compassion and regret) compared to primary emotions like anger, fear, or joy (Vaes et al., 2003). Numerous studies have shown that although people attribute primary emotions to in-groups and out-groups alike, they tend to deny secondary emotions to others who belong to an out-group. In trying to adapt these ideas from dehumanization research to the study of the humanization of nonhuman entities, research by Eyssel et al. (2010) has shown that measuring the attribution of primary and secondary emotions can be used as a measure of anthropomorphism in robots. More recent work has used measured reaction time to reflect the automatic perception of robots as having primary and secondary emotions (Spatola and Wudarczyk, 2021).

A measure for anthropomorphism that was specifically developed for HRI is the Godspeed questionnaire. It has been widely used in the field and has been translated into several languages (Bartneck et al., 2009). More recently, researchers have started developing additional related scales, such as the Robotic Social Attributes Scale (RoSAS) (Carpinella et al., 2017) and the revised Godspeed questionnaire (Ho and MacDorman, 2010) or the Human–Robot Interaction Evaluation Scale (HRIES) (Spatola et al., 2021), a questionnaire measure that integrates ideas underlying dehumanization research and items from the RoSAS (Carpinella et al., 2017).

8.3.2 Implicit measurements

Although many of these measures rest on self-reports and questionnaires, other, more subtle behavioral indicators (e.g., language use, application of social norms that are used in human–human interaction, such as in proxemics) may also be used to investigate the consequences of implementing humanlike form and function in social robots (see Figure 8.2). Enriching the repertoire of measurements from direct to more indirect approaches that are based on reaction times (Spatola and Wudarczyk, 2021; Akdim et al., 2021; Li et al., 2022), for example, will be beneficial not only for the current research in the field of social robotics but likewise as a form of external validation of theorizing in psychology. Wykowska (2021) outlines a variety of HRI experiments that included neurophysiological measurements to shed light on the processes involved. This is certainly useful in order to avoid relying predominantly on self-report measures.

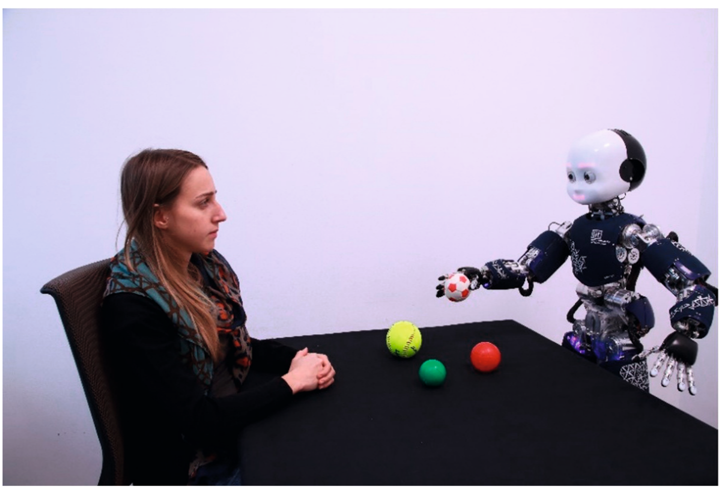


Figure 8.2 An interaction between an iCub robot and a person. Photos like these are used to study whether people believe the robot to have mental states (Marchesi et al. 2019). (Source: Serena Marchesi/IIT)

This material has been published by Cambridge University Press as Human-Robot Interaction by
Christoph Bartneck, Tony Belpaeime, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Sabanovic.
ISBN: 9781009424233 (https://www.cambridge.org/9781009424233).

8.4 Consequences of anthropomorphism

Clearly, it is important to empirically investigate the impact of physical (i.e., appearance-focused) versus psychological anthropomorphism. Perceiving an entity such as a social robot as more or less humanlike comes with a wide array of consequences. For instance, the perceived human-likeness of the robot's appearance or behavior might trigger expectations regarding the entity's functions and capabilities. Often, these expectations far exceed the actual skill set of the respective robot. For example, a robot that features a humanlike face, arms, and legs might be expected to be able to engage in meaningful interaction, display gesture and gazing behaviors, and navigate the social space on two feet. However, most often, these expectations are disappointed in light of the actual capabilities of contemporary robots. That is, specific affordances (see Chapter 4) result in specific perceptions.

Take, for example, the Geminoid robot developed by Ishiguro and Dalla Libera (2018) and Sakamoto et al. (2007) (see Figure 4.7). An android might raise high expectations in end users, given the nearly perfectly humanlike appearance. At the same time, the actual reality of the tele-operated digital twins appears to result in disappointment on the part of the users.

Anthropomorphism, however, can have more consequences than just disappointment. For example, mind attribution to robots affects the perceived suitability of robots for certain jobs and thus might be crucial regarding ultimate deployment and uptake (Wiese et al., 2022).

In addition, psychological anthropomorphism has been related to perceived threat, that is, people feeling threatened in their sense of humanness (Ferrari et al., 2016; Złotowski et al., 2017). This idea is also reflected in qualitative data regarding the perception of autonomous robots (Stapels and Eyssel, 2022). Here, potential end users report fear of being replaced, outperformed, or monitored by robots, which might breach their privacy and misuse their data. Once conflicting evaluations of the same attitude object exist, we experience ambivalence and inner conflict (Stapels and Eyssel, 2021). On the positive side, humanlike perceptions of technology might also increase trust in artificial intelligence (AI) in general (Troshani et al., 2021; Li and Suh, 2021; Kaplan et al., 2021), in intelligent personal assistants (Chen and Park, 2021; Seeger and Heinzl, 2018), in autonomous vehicles (Waytz et al., 2014; Large et al., 2019; Ruijten et al., 2018), and in HRI (Kulms and Kopp, 2019; Christoforakos et al., 2021). Therefore, let us briefly address the notion of trust in social robots and HRI.

8.4.1 Trust in technologies

Many definitions of trust are available, originating from psychology, sociology, economics, and philosophy. These definitions have in common that trust is defined to include having confidence in a person or a system to conduct the appropriate action (Li and Betts, 2003; Biros et al., 2004; Barney and Hansen, 1994). Sabel's definition from 1993, however, focuses on the interaction between each partner's vulnerabilities, defining trust as follows: "Trust is the mutual confidence that no party to an exchange will exploit another's vulnerabilities" (Sabel, 1993, p. 1133). Being confident that an interaction partner will not exploit another partner's vulnerability implies trust in an interaction partner's positive attitudes, benevolence, integrity, trustworthiness, and performance (Lee and See, 2004; Muir, 1994).

According to Parasuraman and Riley (1997), automation is most simply defined as the process by which a machine carries out a function previously completed by a human. Works in the domain of human–automation trust have thus predominantly emphasized the performance of automated systems.

Existing works on trust in automation focus predominantly on improving human users' trust in automation by modifying the performance of the system based on human expectations or matching these with information about the system performance (Schaefer et al., 2016). Perceptions of trust in HRI have been modeled by Hancock et al. (2011, 2021) and Kessler et al. (2017) to consider robot, human, and environmental factors as determinants of trust. Most recent meta-analytic findings (Hancock et al., 2021) have emphasized the role of human-related factors in particular, which is in line with the general paradigm shift to more human-centered research. Despite the clear need for a construct-valid definition of trust, there seems to be no overarching consensus regarding a definition of trust yet. Nevertheless, various scales are available in the literature that appear to tap trust in automation or in social robots (see Krausman et al. (2022) for an overview).

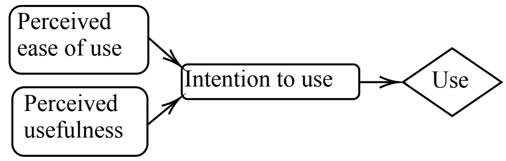
8.4.2 Accepting robots

For obvious reasons, it is important that a social robot is accepted by its human users. At a general level, existing research on social robot acceptance has mostly relied on the classic technology acceptance model (TAM; see Figure 8.3a) and extensions (Heerink et al., 2009). The basic TAM proposes that people's willingness to use a specific type of technology depends on the perceived usefulness and perceived ease of use (Mlekus et al., 2020). Thus, TAM takes the perspective of the robot as an object or a tool that has to be adopted. The TAM has been used to study production systems (Bröhl et al., 2016) and smart objects to investigate the interplay between anthropomorphic features and acceptance.

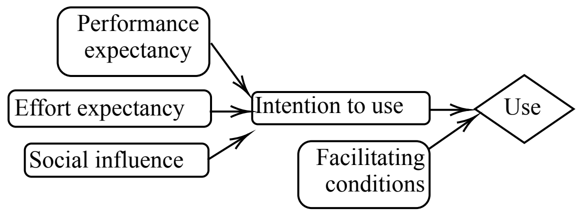
The classic TAM approach fails to consider the role of context factors (de Graaf et al., 2019). Other models therefore have expanded on the TAM by including context factors. For example, in the context of child–robot learning scenarios, the unified theory of acceptance and use of technology (UTAUT; see Figure 8.3b) has been applied (Conti et al., 2017). The UTAUT expands the component of "ease of use" to "effort expectancy" and the component of "perceived usefulness" to "performance expectancy"; it furthermore adds both a social component (e.g., seeing others interact with a robot) and an environmental component ("facilitating conditions").

This material has been published by Cambridge University Press as Human-Robot Interaction by
Christoph Bartneck, Tony Belpaeime, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Sabanovic.
ISBN: 9781009424233 (https://www.cambridge.org/9781009424233).

Figure 8.3 The TAM and UTAUT models.



a. Technology acceptance model (TAM)



b. Unified theory of acceptance and use of technology (UTAUT)

Both the TAM and the UTAUT have an emphasis on cognitive factors. The so-called Almere model (Heerink et al., 2010) builds on these models by adding affective factors such as *trust*, *perceived enjoyment*, and *attitude*. This framework has been developed to examine seniors' perceived acceptance of novel assistive technologies.

In a more general criticism of the TAM, de Graaf et al. (2019) have proposed to take into account hedonic factors, social normative beliefs, and control beliefs when predicting robot acceptance. This could be done by considering user experience (UX). UX is a concept related to TAM, but in addition to the practical attributes of functionality and usability, this framework also takes experiential attributes into account, for example, hedonic values such as stimulation (Hassenzahl, 2003). Moreover, whereas the TAM and the models derived from it consider the *perceived* usefulness and ease of use, the UX model proposes qualities of the technology that would influence these perceptions.

The relevance of UX for social robots and HRI has been recently recognized (Alenljung et al., 2019; Lindblom et al., 2020; Shourmasti et al., 2021; De Graaf and Allouch, 2013). Recent literature reviews, such as those by Shourmasti et al. (2021) and Jung et al. (2021), highlight the usefulness of UX in HRI, despite the clear challenges associated with it (Lindblom and Andreasson, 2016). Outside of the specific HRI context, merging of the TAM and UX models has been proposed to generate a more complete model of user acceptance (Mlekus et al., 2020).

8.4.3 *(Dis-)Liking robots*

Likability refers to the affective evaluation of to what extent a robot is seen to have pleasant or appealing qualities (Sandoval et al., 2021). In social interaction, likability is commonly associated with a willingness to collaborate

This material has been published by Cambridge University Press as Human-Robot Interaction by
Christoph Bartneck, Tony Belpaeime, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Sabanovic.
ISBN: 9781009424233 (https://www.cambridge.org/9781009424233).

(Pulles and Hartman, 2017), allowing yourself to be persuaded (Smith and De Houwer, 2014), and general prosocial behavior (Cillessen and Rose, 2005). At the same time, likability is not exclusively used in a social context; it can also be applied to objects (Niimi and Watanabe, 2012) or brands (Nguyen et al., 2013).

As early as the 1970s, Mori (1970) theorized about a relationship between human-likeness and likability in his theory of the uncanny valley (see Chapter 4). According to this theory, human-likeness would increase likability[1] up to a point; however, when an agent is almost but not quite human, likability would drop.

Recent research has suggested that although there indeed appears to be a drop in likability as agents approach perfect human-likeness, this may be the result of a mismatch in human-likeness between different features (e.g., extremely humanlike skin texture but facial muscle movements that are ever so slightly off; Kätsyri et al., 2015). This "mismatch effect" on uncanny feelings has been replicated for zoomorphic robots (Löffler et al., 2020) and for robots with "mixed" (incongruent) gender cues (Paetzel et al., 2016). At the same time, there appears to be a novelty factor involved as well because feelings of uncanniness tend to reduce after both short- and long-term interaction with a robot (Paetzel-Prüsmann, 2020).

More generally, various studies have found a relationship between robot likability and anthropomorphism (Roesler et al., 2021; Arora et al., 2021; Gonsior et al., 2011). For instance, emotional cues (Eyssel et al., 2010) and robot movement (especially if this movement is in sync with the user) were found to enhance likability (Lehmann et al. 2015; but see Henschel and Cross (2020), who did not find such an effect). Yamashita et al. (2016) extended the relationship between human-likeness and likability to touch and found a correlation between more natural robot "skin" and liking for a robot. Taken together, these findings show that, indeed, the perception of a robot and the actual makeup of a robot—that is, its appearance and functions—interact.

8.5 Conclusion

When we encounter someone, our social cognition kicks in to make a quick and, later, deliberated assessment of that individual. We learned, among other things, that individuals and groups may be judged as low or high in warmth and competence (Cuddy et al., 2008). We also learned that people are pretty good at forming such first impressions in a fast manner, pointing us to the differences between automatic versus controlled processes in social cognition. Humans are likewise good at forming impressions of social robots, and measures of warmth and competence have been prevalent to reflect the basic dimensions of social judgments in social cognition. Moreover, impressions

[1] It should be noted that the original work did not speak of *likability* but rather of a term that has proven to be impossible to translate into English fully and accurately but that touches on familiarity, affinity, and likability.

This material has been published by Cambridge University Press as Human-Robot Interaction by
Christoph Bartneck, Tony Belpaeime, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Sabanovic.
ISBN: 9781009424233 (https://www.cambridge.org/9781009424233).

about robots also extend to the attribution of traits, humanlike characteristics, and mind perception. Such anthropomorphization beyond the merely visible has stirred great interest in engineers and social scientists alike. Finally, we also addressed the consequences of attributing humanlike traits to nonhuman entities, including acceptance, likability, and trust.

Questions for you to think about:

- Think back to the first time you interacted with a robot. Was there something that surprised you? What does that tell you about your automatic expectations?
- Imagine that you are trying to design the most hated robot ever. What behavior would you give it to make sure that people don't like it?
- Name and explain the cognitive determinants of anthropomorphism according to Epley et al. (2007).
- Explain the relationship between the dehumanization of humans and the anthropomorphization of robots.

8.6 Exercises

The answers to these questions are available in the Appendix.

** **Exercise 8.1 Dual processing** What does the *model of dual processing* refer to? Select one option from the following list:

1. That the evaluation of agents depends on *cognitive* and *affective* factors.
2. That mind is attributed along the lines of *uniquely human* and *human-nature* traits.
3. That the processing of the world around us can happen in an *automatic* or more *deliberate* way.
4. That mind is attributed along the lines of *warmth* and *competence*.

** **Exercise 8.2 Social judgements** What are the basic dimensions of social judgments in social cognition? Select one or more options from the following list:

1. Human nature
2. Human uniqueness
3. Agency
4. Warmth
5. Competence
6. Experience

*** **Exercise 8.3 Acceptance** Miciah is developing a social robot and wants to test the user acceptance of her current prototype. She has to decide between using the TAM or the UTAUT. What are some considerations she should take into account? Select one or more options from the following list:

1. The TAM is wrong; Miciah should use the UTAUT.
2. If Miciah wants to test only the interaction between robot and user (i.e., ignoring context), she should use the TAM.
3. The TAM is used for prototyping robots, whereas the UTAUT is used for evaluating robots once their design is complete. Miciah should use the TAM because she's running a prototype.
4. If the robot is designed for a social setting (e.g., to help out in a classroom), the UTAUT would be more appropriate.
5. Both models are valid to use; it depends on what aspects of user acceptance Miciah wants to evaluate.

Future reading:

- Epley, Nicholas, Waytz, Adam, and Cacioppo, John T. On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4):864–886, 2007. doi: 10.1037/0033-295X.114.4.864. URL https://doi.org/10.1037/0033-295X.114.4.864
- Lemaignan, Séverin, Fink, Julia, and Dillenbourg, Pierre. The dynamics of anthropomorphism in robotics. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 226–227. Institute of Electrical and Electronics Engineers, Piscataway, NJ, 2014a. doi: 10.1145/2559636.2559814. URL http://doi.org/10.1145/2559636.2559814
- Lemaignan, Séverin, Fink, Julia, Dillenbourg, Pierre, and Braboszcz, Claire. The cognitive correlates of anthropomorphism. In *2014 Human-Robot Interaction Conference, HRI: A Bridge between Robotics and Neuroscience Workshop*. Institute of Electrical and Electronics Engineers, Piscataway, NJ, 2014b. doi: 10.1007/s12369-014-0263-x. URL https://doi.org/10.1007/s12369-014-0263-x
- Spatola, Nicolas, Marchesi, Serena, and Wykowska, Agnieszka. Different models of anthropomorphism across cultures and ontological limits in current frameworks the integrative framework of anthropomorphism. *Frontiers in Robotics and AI*, 9:863319, 2022. doi: 10.3389/frobt.2022.863319. URL https://doi.org/10.3389/frobt.2022.863319
- Złotowski, Jakub, Sumioka, Hidenobu, Eyssel, Friederike, Nishio, Shuichi, Bartneck, Christoph, and Ishiguro, Hiroshi. Model of dual anthropomorphism: The relationship between the media equation effect and implicit anthropomorphism. *International Journal of Social Robotics*, 10(5):701–714, 2018. doi: 10.1007/s12369-018-0476-5. URL https://doi.org/10.1007/s12369-018-0476-5
- Shourmasti, Elaheh Shahmir, Colomo-Palacios, Ricardo, Holone, Harald, and Demi, Selina. User experience in social robots. *Sensors*, 21(15):5052, 2021. doi: 10.3390/s21155052. URL https://doi.org/10.3390/s21155052