Anthropomorphism in human–robot interactions: a multidimensional conceptualization

Kühne, R.; Peter, J.

DOI
10.1093/ct/qtac020

Publication date
2023

Document Version
Final published version

Published in
Communication Theory

Citation for published version (APA):
Kühne, R., & Peter, J. (2023). Anthropomorphism in human–robot interactions: a multidimensional conceptualization. *Communication Theory*, *33*(1), 42-52. https://doi.org/10.1093/ct/qtac020

Communication Theory, 2023, 42–52
https://doi.org/10.1093/ct/qtac020
Advance access publication 18 October 2022
Original Article

International Communication Association
OXFORD

Anthropomorphism in human–robot interactions: a multidimensional conceptualization

Rinaldo Kühne * and Jochen Peter

Amsterdam School of Communication Research, ASCoR, University of Amsterdam, Amsterdam, The Netherlands

*Corresponding author: Rinaldo Kühne, Amsterdam School of Communication Research, ASCoR, University of Amsterdam, P.O. Box 15791, 1001 NG Amsterdam, The Netherlands. Email: r.j.kuhne@uva.nl

Abstract

With robots increasingly assuming social roles (e.g., assistants, companions), anthropomorphism (i.e., the cognition that an entity possesses human characteristics) plays a prominent role in human–robot interactions (HRI). However, current conceptualizations of anthropomorphism in HRI have not adequately distinguished between precursors, consequences, and dimensions of anthropomorphism. Building and elaborating on previous research, we conceptualize anthropomorphism as a form of human cognition, which centers upon the attribution of human mental capacities to a robot. Accordingly, perceptions related to a robot's shape and movement are potential precursors of anthropomorphism, while attributions of personality and moral value to a robot are potential consequences of anthropomorphism. Arguing that multidimensional conceptualizations best reflect the conceptual facets of anthropomorphism, we propose, based on Wellman's (1990) Theory-of-Mind (ToM) framework, that anthropomorphism in HRI consists of attributing thinking, feeling, perceiving, desiring, and choosing to a robot. We conclude by discussing applications of our conceptualization in HRI research.

Keywords: human–robot interaction (HRI), social robots, technology, Theory-of-Mind (ToM), digital agents

The cognition that an entity possesses human characteristics—called anthropomorphism—is a fundamental psychological response that is based on cognitive representations typically acquired during childhood (Bjorklund & Causey, 2018). While humans can basically anthropomorphize anything (Epley et al., 2007), perceptions of human-likeness play a crucial role in human–robot interactions (HRI) (Duffy, 2003; Fink, 2012; Kahn et al., 2007) because characteristics of humans are often used as guiding principles in robot design (Fong et al., 2003). Human communication, for example, has often been employed as a blueprint for the design of technology and has been regarded as key to improve robot usability (Guzman, 2018). Accordingly, research and development in robotics has focused on making robots communicate and behave more like humans to facilitate their use. This development may be particularly important for social robots (Breazeal, 2003). These robots are not primarily endowed with humanlike characteristics to increase their ease of use, but to create meaningful social interactions and relationships (Breazeal, 2003; Fong et al., 2003).

Despite the crucial role of anthropomorphic design and, accordingly, perceptions of human-likeness in HRI, there are two issues with the concept of anthropomorphism that currently hinder progress in the field. First, anthropomorphism has not been consistently distinguished from related psychological phenomena. Notably, it is not evident whether certain cognitions and behaviors present precursors, consequences, or constitutive elements of anthropomorphism. For instance, it is not clear whether perceptions of a robot's shape, movement, and behavior lead to the anthropomorphization of the robot or whether they are an aspect of anthropomorphization. Theoretical frameworks commonly used in HRI are also inconclusive in this respect. For example, the computers-are-social-actors (CASA) framework, a seminal framework in HRI research, posits that people treat computers as if they were real people. Research has confirmed this hypothesis by showing, for instance, that people are polite to computers or apply reciprocity norms to computers (Nass & Moon, 2000). While the CASA framework originally suggested that social responses to computers are typically mindless and automatic (Nass & Moon, 2000), recent research suggests that social responses are facilitated by more anthropomorphic computer representations (Gong, 2008). Accordingly, the role of anthropomorphism in the CASA framework seems to deserve further attention (Lee-Won et al., 2020).

Second, no consensus exists about the dimensional structure of anthropomorphism in HRI. Scholars disagree on whether anthropomorphism is best conceptualized as uni- or multidimensional, that is, whether anthropomorphism is a monolithic concept (i.e., unidimensional) or whether it has multiple facets, or dimensions, that constitute the overall concept (i.e., multidimensional). Proponents of a multidimensional view, in turn, disagree on which dimensions such a conceptualization should include (Ruijten, 2018; Zhang et al., 2008). Epley et al. (2007), for instance, propose that attributions of conscious thought and intentions are crucial dimensions of anthropomorphism. However, the authors indicate themselves that additional dimensions, such as the attribution of emotions or behavioral characteristics, may also be important for defining anthropomorphism. Accordingly, HRI researchers have employed different sets of criteria, as well as dimensions, when conceptualizing and operationalizing anthropomorphism (Bartneck et al., 2009; Złotowski et al., 2014).

These issues call for an overview of, and theoretical reflection on, the concept of anthropomorphism and suggestions

Received: 21 December 2021. **Revised:** 14 June 2022. **Accepted:** 25 August 2022

for how anthropomorphism can be clearly and stringently conceptualized in HRI. Without such a conceptualization, research findings about the role of anthropomorphism in HRI will remain difficult to compare across studies and cumulative evidence and progress in the field is hindered. At a more general level, our article thus also responds to scholars' call for a more systematic and coherent conceptual and operational definition of key concepts in the field (e.g., de Jong et al., 2019).

In addressing the aforementioned issues, we build heavily on previous work on anthropomorphism by Epley and colleagues (Epley et al., 2007; Epley & Waytz, 2010) and on Wellman's (1990) Theory-of-Mind (ToM) framework. We argue, first, that determinants, dimensions, and consequences of anthropomorphism should be clearly and consistently differentiated. Specifically, we maintain that attributions of human mental capacities are at the core of the anthropomorphism of robots, whereas perceptions related to shapes, movements, and behaviors are precursors of anthropomorphism. Moreover, attributions of personality and moral value to a robot should be regarded as potential consequences of anthropomorphism. We do not aim at a full-fledged theory of anthropomorphism that discusses all its predictors and consequences and their exact meaning. Instead, the goal is to clarify which frequently mentioned definitional facets of anthropomorphism should be regarded as predictors and consequences rather than as constitutive criteria. Furthermore, our conceptualization is agnostic regarding the question of robot ethics. Our conceptualization does not suggest that a robot's ontology ought (or ought not) to be assessed before rights can be assigned to a robot. Our conceptualization exclusively focuses on psychological mechanisms that unfold in HRI, not on how robot rights should be evaluated (for a detailed discussion of robot ethics see Gunkel, 2012, 2022).

Second, we posit that a multidimensional conceptualization is better suited to accurately reflect the conceptual complexity of anthropomorphism than a unidimensional conceptualization. Building on Wellman's (1990) ToM framework, we propose a multidimensional conceptualization of robot anthropomorphism, which comprises thinking, feeling, perceiving, desiring, and choosing. Thus, we propose a conceptualization which is more focused, as anthropomorphism is separated from precursors and consequences, as well as more detailed, as five facets of anthropomorphism are distinguished. We outline that this conceptualization has the advantage of being applicable across a broad range of HRIs and of facilitating a detailed analysis of anthropomorphism's relationships with other concepts. Finally, we discuss the limitations of this new conceptualization and its implications for future research.

Precursors and consequences of anthropomorphism

The first ambiguity in the conceptualization of anthropomorphism in HRI refers to the boundaries of the concept and, more specifically, the distinction of anthropomorphism from its precursors and consequences in HRI. Scholars concur that anthropomorphism constitutes a form of human cognition that centers on the assignment of human characteristics to a nonhuman entity and that can be elicited when humans observe or interact with a robot (Bartneck et al., 2009; Duffy, 2003; Epley et al., 2007; Fink, 2012). While anthropomorphism can be elicited in response to a broad range of robots, ranging from machinelike to humanlike robots (Tan et al., 2018), imbuing robots with humanlike characteristics is likely to facilitate anthropomorphism (Fink, 2012; Tan et al., 2018). Anthropomorphism thus needs to be distinguished from the anthropomorphic design of robots and, more specifically, from the properties of a robot that may (or may not) elicit anthropomorphism (Fink, 2012; Ruijten, 2018; Yogeeswaran et al., 2016). This distinction is reflected in HRI research that investigates how robot features that are expected to make a robot more humanlike influence the extent to which users anthropomorphize a robot (Kiesler et al., 2008; Salem et al., 2013).

However, authors disagree where the cognition of anthropomorphism begins and where it ends. Specifically, it is unclear which cognitions are precursors of anthropomorphism; which cognitions constitute anthropomorphism; and which cognitions are consequences of anthropomorphism. Some authors have argued that anthropomorphism includes the perception that a nonhuman entity has a *human form* (Bartneck et al., 2009; Liarokapis et al., 2013; Waytz, Epley, et al., 2010; Zhang et al., 2008; Złotowski et al., 2015). This includes the perception that a nonhuman entity has a human size (Zhang et al., 2008), a humanlike morphology or shape (Liarokapis et al., 2013; Zhang et al., 2008), or a humanlike face (Bartneck et al., 2009). The perception that a nonhuman entity *moves* or behaves like a human has also been regarded as an aspect of anthropomorphism (Bartneck et al., 2009; Liarokapis et al., 2013; Zhang et al., 2008).

Frequently, the attribution of *human mental capacities* is emphasized in definitions of anthropomorphism (Duffy, 2003; Epley et al., 2007; Eyssel, 2017; Haslam & Loughnan, 2014). In this view, anthropomorphism refers, broadly speaking, to the belief that a nonhuman entity can think and feel like a human. Other conceptualizations emphasize that attributing a *personality* (i.e., regularities and consistencies in orientations and responses; Higgins, 1990) to an entity indicates anthropomorphism (Kiesler & Goetz, 2002; Kim & Sundar, 2012; Salem et al., 2013). Finally, Kahn et al. (2007) suggest that the attribution of intrinsic *moral value* to a robot constitutes a psychological benchmark of whether the robot is perceived as humanlike. This implies that assigning moral value to a robot is an additional aspect of anthropomorphism. The definitional issue is further complicated because several conceptualizations mention multiple of the above aspects as constituents of anthropomorphism (e.g., Bartneck et al., 2009; Kahn et al., 2007).

Against the backdrop of diverse conceptualizations of anthropomorphism and the resulting complexity and definitional problems, we believe that a more focused definition of anthropomorphism—which zooms in on core aspects of the concept and excludes other aspects that were part of earlier definitions—is most useful for HRI research. We argue that the *attribution of human mental capacities to robots* presents the core of robot anthropomorphism. This conceptual facet has been included in most (Broadbent, 2017; Duffy, 2003; Eyssel, 2017; Fink, 2012), albeit not all (e.g., Zhang et al., 2008), discussions of anthropomorphism in robotics, as well as in psychological conceptualizations (Epley et al., 2007), including research on dehumanization (Haslam & Loughnan, 2014), and in philosophical reflections on personhood (Dennett, 1988). This more focused definition thus has the advantage that it probably reflects a scholarly consensus on the meaning of anthropomorphism.

Limiting the concept of robot anthropomorphism to the attribution of human mental capacities to robots raises the question of what role the excluded conceptual facets (i.e., perceptions of a robot's humanlike form and movement, attributions of personality and moral value to a robot) play. In general terms, we suggest that these facets constitute psychological responses that are related to, but are yet distinct from, anthropomorphism. More specifically, within the nomological network of anthropomorphism in HRI, perceptions of a robot's humanlike form and movement should be regarded as precursors of anthropomorphism, whereas attributions of personality and moral value to a robot should be regarded as consequences of anthropomorphism.

As Epley et al. (2007) point out, anthropomorphism is an "inference about *unobservable characteristics* (emphasis added) of a nonhuman agent, rather than descriptive reports of a nonhuman agent's observable or imagined behavior" (p. 865). Perceptions of robot shape and movement may thus be distinguished from anthropomorphism because they refer to *observable* robot characteristics. This view resonates with the position of Eyssel (2017), who argued, referring to Epley et al. (2007), that anthropomorphism in HRI goes "beyond simple judgments of an entity with regard to its anthropomorphic form or lifelikeness" (p. 366).

Research indicates that perceptions of observable robot characteristics do influence judgments regarding robots' mental capacities. For instance, Barco et al. (2020) and Hegel et al. (2008) demonstrated that robots with a humanlike form are more anthropomorphized than functional and zoomorphic robots. Moreover, the perception that a nonhuman target moves at a speed similar to the speed of human movement has been shown to increase mind attribution (Morewedge et al., 2007). As Urquiza-Haas and Kotrschal (2015) discuss in their review, neurological research corroborates that the human brain's "social information processing center (...) [is triggered by] (...) biological motion, human body motion, hand and mouth movement and facial expressions" (p. 169). Similarly, communicative behaviors and, in particular, the use of speech have been shown to increase the likelihood that an entity is judged as human (Schroeder & Epley, 2016). In sum, it is plausible to regard perceptions related to observable robot characteristics, such as shape and movement, as "upstream" cognitions—or precursors—that may result in robot anthropomorphism.

A clearer conceptual demarcation is also required on the "downstream" side—from the consequences—of anthropomorphism. Notably, anthropomorphism has to be distinguished from the attribution of personality and moral value to a robot, as the latter two concepts have previously been included in conceptualizations of anthropomorphism. Both philosophical (Dennett, 1987, 1988) and psychological (Epley & Waytz, 2010) accounts of anthropomorphism imply that attributions of personality and moral value to a robot should be regarded as consequences rather than dimensions of robot anthropomorphism.

Dennett (1987, 1988) proposed that humans can use three different strategies to analyze and predict their environment. Next to the physical stance (i.e., analyzing one's environment in terms of physical properties and laws) and the design stance (i.e., analyzing entities in terms of their intended functions or purpose), Dennett proposes what he calls the intentional stance, which is the most central to this article. This stance basically corresponds to anthropomorphism (i.e., the attribution of human mental capacities) as it implies that people can treat other entities as "rational agents" (Dennett, 1987, p. 17) to predict their behavior. This stance is the most useful to understand and predict the behavior of other humans, but it is also applied to nonhuman entities, such as animals and natural phenomena. As Dennett (1987) points out, taking the intentional stance is a gateway for subsequent "social" cognitions:

> First you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. (p. 17).

A similar view stems from Epley and Waytz (2010), who argue that anthropomorphism is a precursor of cognitions about other people's mental states and behaviors. Likewise, Margolin (2021) proposes that people seek information about the "choice-making tendencies of agents" (p. 718) when trying to understand their behavior. In line with these views, neurological research shows a "neurological distinction in the processing of the physical and social aspects of the world" (Urquiza-Haas & Kotrschal, 2015, p. 169) and that perceptions that an entity is human can stimulate "social" brain circuits (Hortensius & Cross, 2018). Accordingly, it makes sense to regard the anthropomorphism of robots a precursor of attributing a specific personality to a robot: Humans first decide whether a robot can be regarded as humanlike, and if this is the case, they can assess the robot along person-related criteria, including personality traits.

Several authors also suggest that anthropomorphism precedes the attribution of a moral value to an entity. Haslam and Loughnan (2014) and Waytz, Gray, et al. (2010), for example, suggest that moral value and moral responsibility are typically attributed to entities that are capable of agency and experiences—two essential mental capacities of humans. Agency—including intentionality and autonomous decision making—is crucial for the assignment of moral responsibility, whereas the ability to experience happiness and pain is essential for the assignment of moral value and rights. Accordingly, anthropomorphizing the robot influences whether a robot is regarded as a moral agent (Sullins, 2006). Based on the literature, it is thus plausible to distinguish anthropomorphism—the attribution of human mental capacities to a nonhuman entity—from upstream cognitions, which predict anthropomorphism (i.e., perceptions of shape and movement as humanlike), as well as from downstream cognitions, which follow from anthropomorphism (i.e., attributions of personality and moral value).

Compared to other conceptualizations of anthropomorphism, our conceptualization is more focused, which we believe is advantageous for at least two reasons. First, a more focused conceptualization of anthropomorphism is more "adequate," which increases its heuristic value and facilitates future research in HRI. An adequate definition is neither too narrow nor too broad in its meaning (Westermann, 2000). A broad conceptualization of anthropomorphism runs the risk of subsuming different types of cognitions, which potentially influence each other and form a causal mechanism, under one concept. This may hinder research on the interplay of concepts that deserve attention in their own right. In contrast, distinguishing anthropomorphism from potential precursors and consequences enables a theoretically sound causal analysis of the relationships of anthropomorphism with related concepts—not only with the concepts mentioned above, but with a broad range of potential precursors (e.g., elicited agent knowledge, effectance motivation, and sociality motivation; see Epley et al., 2007) and consequences (e.g., the acceptance of robots; see Duffy, 2003) that are discussed in HRI research. Our more focused conceptualization not only facilitates "vertical" distinctions along the causal chain, but also "horizontal" distinctions between anthropomorphism and similar concepts such as animacy (Scholl & Tremoulet, 2000) and social presence (Biocca et al., 2003). However, an in-depth discussion of distinctions between anthropomorphism and similar concepts is beyond the scope of this article because it requires an elaborate conceptual analysis of each similar concept.

Second, a more focused definition of anthropomorphism supports a "stable" conceptualization of anthropomorphism, replacing context-dependent conceptualizations. A definition is stable or durable if it can be used under a variety of conditions (McLeod & Pan, 2005). Anthropomorphism can occur in response to a broad range of entities (e.g., objects, animals, artifacts, natural phenomena). In HRI, more specifically, anthropomorphism can emerge in response to a large variety of robots, including anthropomorphic, zoomorphic, caricatured, and functional robots as well as embodied and non-embodied robots (Fong et al., 2003). Thus, a broad definition of anthropomorphism runs the risk of stipulating criteria of anthropomorphism that are only relevant for a subset of robots, with the result that anthropomorphic responses cannot be compared across robot models. For instance, if humanlike shape is a criterion of anthropomorphism (which it would be in a broad definition), embodied and non-embodied robots can hardly be compared in their degree of anthropomorphism. This does not mean that concepts such as robot characteristics and perceptions of such characteristics should be discarded because they are only relevant in a subset of HRI contexts. Rather, we suggest that these robot characteristics and corresponding perceptions are context-dependent predictors of anthropomorphism. Anthropomorphism as we define it here, by contrast, is a more general psychological phenomenon that may arise across a broad range of HRI contexts. A more focused definition thus accounts for the generality of the phenomenon of anthropomorphism, while still allowing for the meaningful study of context-dependent predictors that elicit anthropomorphism.

Uni- versus multidimensional conceptualizations of anthropomorphism

As mentioned before, HRI scholars do not fully agree on, first, whether anthropomorphism is best conceptualized as uni- or multidimensional, and, second, which dimensions a multidimensional conceptualization should include (Ruijten, 2018; Zhang et al., 2008). Viewing a concept as unidimensional implies that the concept's facets largely overlap in their meaning and, accordingly, that its indicators, statistically speaking, load on one factor. A multidimensional view implies that the concept's facets have a (partially) distinct meaning and, accordingly, that its indicators load on multiple factors.

Unidimensional conceptualizations of anthropomorphism are frequently employed (Barco et al., 2020; Bartneck et al., 2009; Salem et al., 2013), the most prominent being the conceptualization by Bartneck and colleagues (2009). Unidimensional conceptualizations are not per se problematic if they are rooted in an accurate theoretical definition. However, we believe that a multidimensional conceptualization of anthropomorphism is preferable for two reasons. First, humans possess a variety of mental capacities (e.g., feeling and thinking) and they are likely to refer to, and differentiate between, multiple mental capacities when evaluating whether a robot is humanlike. It is thus plausible to assume that the anthropomorphism of robots includes multiple, distinct facets. Second, a unidimensional conceptualization, which does not allow for multiple, distinct facets, runs the risk of being overly abstract and imprecise. Accordingly, unidimensional conceptualizations have elicited discrepant operational definitions. For instance, Bartneck et al. (2009) used indicators such as "humanlike," "lifelike," and "natural" to form a unidimensional measure of anthropomorphism. Waytz, Morewedge, et al. (2010) employed a series of different unidimensional operationalizations, including the single-item measure that the computer appears to "behave as if it has its own beliefs and desires" (p. 414) and a unidimensional measure consisting of the three indicators that a gadget has "a mind of its own," "intentions, free will, consciousness," and that it appears "to experience emotions" (p. 415). A multidimensional conceptualization, in contrast, may increase conceptual precision by specifying more concrete subdimensions and, thus, supports a more consistent and comparable operational definition of anthropomorphism.

One frequently employed multidimensional conceptualization is rooted in Haslam's dual model of dehumanization, which addresses why people deny other persons or social groups humanness (Haslam, 2006; Haslam et al., 2009; Haslam & Loughnan, 2014). The model differentiates two categories of mental capacities that may or may not be attributed to others: uniquely human characteristics, which distinguish humans from animals and nonhuman entities (e.g., refinement), and human nature characteristics, which are essentially or typically human (e.g., interpersonal warmth). Dehumanization occurs when people perceive others as lacking human nature and/or uniquely human characteristics (Haslam & Loughnan, 2014).

Haslam's dehumanization model (Haslam, 2006; Haslam & Loughnan, 2014) is a potentially useful foundation for the conceptualization of anthropomorphism in HRI because it describes mental criteria that individuals may use when evaluating a robot's human-likeness. Accordingly, HRI researchers have posited that the anthropomorphism of robots includes a "uniquely human" and a "human nature" dimension (Eyssel et al., 2010; Złotowski et al., 2014) and several HRI studies refer to this distinction when measuring anthropomorphism (Eyssel et al., 2011; Ferrari et al., 2016; Salem et al., 2013). However, uniquely human and human-nature characteristics constitute themselves broad mental categories, which subsume a variety of psychological concepts. According to Haslam (2006), human uniqueness refers to civility, refinement, moral sensibility, rationality and logic, and maturity. Human nature refers to emotional responsiveness, interpersonal warmth, cognitive openness, agency and individuality, and depth. Because of this conceptual breadth, distinguishing uniquely human and human-nature characteristics as constitutive dimensions of anthropomorphism does not sufficiently clarify the meaning of anthropomorphism.

A second multidimensional conceptualization of anthropomorphism is rooted in Gray, Gray, and Wegner's (2007) work on mind perception. They propose that people consider two mental faculties when assessing whether other entities have minds: whether the entity can act and decide autonomously (agency) and whether it is sentient (experience). Accordingly, HRI scholars have argued that the anthropomorphism of robots is a two-dimensional concept consisting of agency and experience (Hortensius & Cross, 2018; Złotowski et al., 2015). Scholars have recognized a conceptual overlap between Haslam's and Gray et al.'s frameworks, with uniquely human characteristics broadly mapping onto agency, and human-nature characteristics onto experience (Hortensius & Cross, 2018; Złotowski et al., 2015). However, like the distinction between human uniqueness and human nature (Haslam, 2006), the agency-experience distinction results in subdimensions of anthropomorphism that are conceptually somewhat fuzzy. Moreover, while the agency-experience distinction has influenced operationalizations of anthropomorphism and mind perception in HRI (e.g., Ferrari et al., 2016; Fraune, 2020; Fraune et al., 2020; Trovato & Eyssel, 2017), it has not yet become an established theoretical standard in HRI.

Using the ToM framework to specify dimensions of anthropomorphism

Both uni- and multidimensional conceptualizations of anthropomorphism in HRI are currently ambiguous: Unidimensional conceptualizations do not distinguish relevant subdimensions of anthropomorphism and multidimensional conceptualizations propose subdimensions that are themselves not entirely clear. Consequently, anthropomorphism has been inconsistently operationalized in HRI. Against this backdrop, we propose that a set of core dimensions of anthropomorphism, which account for and complement the main dimensions in extant frameworks (Haslam & Loughnan, 2014; Gray et al., 2007), can be based on the ToM framework. The main assumption of the ToM framework is that humans possess a naïve theory of mind, which proposes a set of mental faculties that people possess (e.g., feeling and thinking) and that can be used to predict people's behavior (Bjorklund & Causey, 2018).

The ToM framework has been mainly employed in developmental research that investigates at which age children acquire specific aspects of a ToM (Bjorklund & Causey, 2018). The framework has also been mentioned in the context of HRI. In design-oriented studies, scholars have posited that robots should be equipped with a ToM to facilitate proper social interactions between humans and robots (Benninghoff et al., 2013; Scassellati, 2002). In user-centered research, some scholars have noticed a conceptual relationship between ToM and anthropomorphism. Stafford et al. (2014) posited that "the theory of mind perception is related to anthropomorphism, in that people attribute capacities of mind to non-human characters" (p. 20). A similar view was expressed by Airenti (2018, p. 10): "In anthropomorphic attribution, children use the same theory of mind abilities that they use in interactions with humans." With respect to HRI, Banks (2020) has recently argued that robots' anthropomorphic design (e.g., visual and behavioral social cues) may facilitate the application of a ToM to robots. She found that human-likeness indeed increased the use of ToM with robots. Beyond HRI research, Epley and Waytz (2010) have pointed out the that mind perception is an inferential process that "is usually referred to as a *theory of mind*" (p. 505).

The above accounts indicate that employing ToM in human–human interactions (HHI) and anthropomorphizing robots in HRI involve related processes, except that, in HHI, the other is a human and, in HRI, the other is a robot. However, there is an additional difference between HHI and HRI, which is particularly relevant in the context of anthropomorphism in HRI: Anthropomorphism *precedes*[1] the application of ToM, that is, the explanation of robot behavior on the basis of mental capacities. Applying ToM to understand a robot's behavior requires that the robot is first (at least partly) anthropomorphized or, in the terminology of Dennett (1987, 1988), that the intentional stance toward the robot is taken. As Dennett (1987) points out, one has to first decide to treat an object as a rational agent, before beliefs and desires can be attributed to the object. Accordingly, Epley (2018) argues that "instead of treating other people like objects, we instead attribute a mind to another person, complete with concepts like intentions, desires, attitudes, and beliefs that can be used to explain his or her behavior" (pp. 592–593). While initial assessments of others' mental faculties are relevant in human interactions (for instance, in interactions with children or people with mental disabilities), they are essential in HRI because (social) robots widely differ in their capacities and in their similarity to humans (Fong et al., 2003).

The fact that anthropomorphism precedes the application of ToM elucidates why anthropomorphism and ToM entail corresponding subdimensions: The application of ToM to a robot logically presupposes that the robot possesses *a specific set of human mental faculties*: Only if the robot possesses these mental faculties can explanations of the robot's behavior be based on them. Anthropomorphism thus constitutes a cognition in which an individual evaluates whether and to which degree the robot possesses the *respective human mental faculties*. Each assessment of a distinct mental faculty, in turn, constitutes a dimension of anthropomorphism. Accordingly, an inspection of the main elements of people's ToM should reveal the main dimensions of anthropomorphism.

Wellman's (1990; see also Wellman & Bartsch, 1988) model of ToM is particularly suited to derive the main dimensions of anthropomorphism because it explicitly discusses the mental faculties that people's theory of mind comprises. At the core of the model is the notion of belief-desire reasoning: People's actions are assumed to be rooted in their beliefs and desires. *Beliefs* constitute an individual's knowledge about the state of the world. *Desires* are what the individual wants, wishes, or hopes for. Beliefs and desires are assumed to predict people's *intentions*, which reflect what people aim or plan to do; and intentions, in turn, are regarded as the main predictor of *actions*: "Intentions are plans to actualize certain desires; intentions harness various beliefs onto actions" (Wellman, 1990, p. 111). Individuals' beliefs and desires are themselves influenced by perceptions and basic emotions and physiology: *Perceptions* (e.g., seeing, hearing) are assumed to inform other individuals about the external world and to form the basis of beliefs, whereas *basic emotions and physiology* (e.g., hunger, fear) are assumed to function as an internal source of information, which predicts desires.

Wellman's (1990) model of ToM comprises two additional concepts, cognitive emotions and sensations. However, the two concepts largely overlap with concepts discussed above: Cognitive emotions are an "intermediate concept between emotion and thinking" (Wellman, 1990, p. 114), and sensations may be located between perception and physiology (the role of sensations within the model is not explicated in detail). Thus, the two concepts do not seem to constitute distinct mental faculties that may inform dimensions of anthropomorphism.

Based on Wellman's (1990) ToM model, we propose that the anthropomorphism of robots includes five dimensions that reflect the main components of people's ToM. Individuals attribute these capacities to robots in a non-binary way: They evaluate to which degree a robot possesses the capacities. The first dimension is the attribution of *thinking* to a robot. We use the term thinking instead of beliefs because, as Wellman (1990) points out, beliefs belong to the broader concept of thinking and are produced by thinking processes, such as reasoning, learning, and remembering. Accordingly, we define thinking as all higher-level cognitive processes (Blanchette & Richards, 2010) or all "conscious and effortful" (Smith & DeCoster, 2000, p. 108) types of information processing which involve "the intentional retrieval of explicit, symbolically represented rules (...) and their use to guide processing" (Smith & DeCoster, 2000, p. 108). When one attributes thinking to a robot, one believes that the robot's hard- and software enable some forms of higher cognitive process. Thinking is part of previous conceptualizations of anthropomorphism and mind perception: It reflects aspects of uniquely human characteristics in dehumanization theory (Haslam, 2006; Haslam et al., 2009) and of agency in Gray et al.'s (2007) theory of mind perception. The attribution of thinking (or higher-level cognitive processes) has previously also been identified as an aspect of anthropomorphism in HRI (Duffy, 2003; Fussell et al., 2008; Trovato & Eyssel, 2017).

The concepts of physiology, emotion, and sensation in Wellman's (1990) model are summarized in the second dimension of anthropomorphism: the attribution of *feeling* to a robot. Feelings are an individual's experiences of bodily changes and reactions (LeDoux, 1999), such as hunger, sexual arousal, and pain. An important aspect of feeling is the experience of emotions such as anger, fear, and happiness. We follow Leventhal and Scherer (1987) in proposing that feelings can be elicited by sensory motor, schematic, and conceptual processes. Feelings can be produced by processes ranging from automatic and reflex-like, which maps onto Wellman's (1990) concept of sensations, to volitional and reflective, which maps onto Wellman's concept of cognitive emotions. Attributing feelings to a robot does not presume that the robot possesses a human physiology. Similar to the attribution of thinking to a robot, it rather reflects the belief that a robot's hard- and software enable subjective experiences. The feeling dimension of anthropomorphism reflects an aspect of human nature in dehumanization theory (i.e., emotionality; Haslam, 2006; Haslam et al., 2009). It closely maps onto the experience dimension in Gray et al.'s (2007) conception of mind perception. Moreover, several HRI scholars have mentioned the attribution of emotions as an aspect of anthropomorphism (Duffy, 2003; Fussell et al., 2008; Lemaignan et al., 2014; Wang & Krumhuber, 2018).

The third dimension of anthropomorphism is the attribution of *perception* to a robot. Perception means that the robot is able to sense its environment (notably, vision, hearing, touch, taste, and smell) and interpret this input. Again, attributing perception to a robot does not presume a human physiology because sensors can emulate human senses. We subsume sensation under feeling rather than perception because Wellman's definition refers to internal experiences such as dizziness, nausea, and pain. Accordingly, we posit that feelings refer to the experience of internal states, whereas perceptions refer to the experience of external stimuli. Perception is typically not identified as a distinct dimension of anthropomorphism by HRI scholars. However, the dimension is relevant because perceptions are an essential aspect of the ToM framework and the application of ToM to a robot presumes that the robot has at least minimal access to its environment. Moreover, scholars have proposed that artificial agents, such as robots, should be able to perceive their environment to function as, and be considered, social entities. Hubbard (2011) notes that "in order to be a 'living' entity of any sort, one must have the ability to interact meaningfully with the environment by receiving and decoding inputs from, and sending intelligible data to, its environment" (p. 419). Research on social robots—i.e., robots that emulate socialness for the purpose of enabling social relationships with humans—indicates that robots can only fulfil social functions if they can perceive their (social) environment (Breazeal, 2003; Fong et al., 2003).

The fourth dimension of anthropomorphism is the attribution of *desiring* or wanting to a robot. This faculty has also been referred to as motivation (Baumeister, 2016; Ryan & Deci, 2000). According to Baumeister (2016), "it is a condition of an organism that includes a subjective sense (not necessarily conscious) of desiring some change in self and/or environment" (p. 1). Attributing desiring to a robot corresponds to believing that the robot possesses needs and preferences that it wishes to gratify. Desiring (or wanting) is closely related to intention, as indicated above and as elaborated by Wellman (1990, p. 110): "One can have a desire (a wish, a hope, a fancy) to do something but in spite of this never actually intend (plan, aim, decide) to do it."

Individuals may possess a variety of needs, but they can choose which ones they try to satisfy, and through which means. Accordingly, individuals' behaviors are not fully determined by their desires (and beliefs) because they can choose a course of action. Different terms have been used to describe this idea within the context of artificial agents, including intentionality (Bigman et al., 2019), free will (Krausová & Hazan, 2013), agency (Himma, 2009), and autonomy (Beer et al., 2014). We use the term *choice* to describe the ability to freely choose between multiple courses of action, which is the fifth and final dimension of the anthropomorphism of robots.

Desiring and choice both constitute human nature characteristics in dehumanization theory (Haslam, 2006; Haslam et al., 2009). Moreover, desiring reflects an aspect of the experience dimension and choice an aspect of the agency dimension in Gray et al.'s (2007) theory of mind perception. However, like perception, desiring and choice are typically not identified as distinct dimensions of anthropomorphism in HRI.

In summary, building and elaborating on previous research and relying notably on Wellman's (1990) ToM framework, we propose that the anthropomorphization of robots consists in attributing five human mental capacities—thinking, feeling, perceiving, desiring, and choosing—to a robot. These mental capacities constitute robot anthropomorphism because they are integral to the explanation and prediction of human behavior. Accordingly, if any one of the dimensions were not to be included in a definition of robot anthropomorphism, a crucial benchmark of a mind's human-likeness would be missing. As our conceptualization explicates the main dimensions of anthropomorphism, it represents a specification of Dennett's (1987) intentional stance and existing conceptualizations of anthropomorphism (e.g., Epley et al., 2007). By specifying the main dimensions of anthropomorphism, our conceptualization improves the definitional precision and decreases ambiguity (McLeod & Pan, 2005). Our conceptualization of anthropomorphism is depicted in Figure 1. In addition, Table 1 explicates the advantages of our conceptualization over some previous conceptualizations. The table lists three main quality criteria of definitions—stability, adequacy, and precision—and specifies: (a) problems of existing conceptualizations related to these criteria; (b) how our conceptualization addresses these problems; (c) and advantages of our conceptualizations for future research.

Future research and limitations

Our conceptualization of anthropomorphism may advance research on anthropomorphism and provides multiple avenues for future research. First, clearly demarcating the conceptual boundaries of anthropomorphism facilitates research on the precursors and consequences of anthropomorphism. This type of research is not limited to the study of concepts that were previously regarded as elements of anthropomorphism. Rather, our conceptualization allows for the study of diverse potential predictors (e.g., robot, interaction, and user characteristics) and consequences (e.g., robot acceptance and relationship formation with robots). Moreover, because our conceptualization is multidimensional, it enables research on whether specific dimensions of anthropomorphism respond differently to precursors, whether they have distinct consequences, and how the dimensions are related to each other.

Second, our conceptualization facilitates clear distinctions of anthropomorphism from similar concepts that are frequently studied in HRI, such as animacy (Scholl & Tremoulet, 2000) and social presence (Biocca et al., 2003) and stimulates investigations into the differential roles of these concepts in HRI. Such distinctions make it possible to study, for example, the differential effects of robot morphology on anthropomorphism, animacy, and social presence (e.g., Barco et al., 2020).

Third, our conceptualization of anthropomorphism excludes context-dependent upstream cognitions (notably, perceptions of robots' shapes, movements, and behaviors). Consequently, it can be applied to many HRI contexts, ranging from different robot types (e.g., anthropomorphic, zoomorphic, caricatured, and functional, Fong et al., 2003) to robots with both a highly anthropomorphic design (i.e., androids) and a minimally anthropomorphic design (i.e., industrial robots) to diverse interaction scenarios (e.g., playing or collaborating with a robot). At the same time, our conceptualization can be linked to context-dependent upstream cognitions to investigate idiosyncratic psychological processes. For instance, when studying humanoid robots, one can investigate whether the perception that the robot has a humanlike shape influences anthropomorphism.

There are at least three limitations of our conceptualization of anthropomorphism. First, our conceptualization presumes that the anthropomorphism of robots comprises explicit cognitions that individuals can consciously access; it does thus not cover implicit (Urquiza-Haas & Kotrschal, 2015) or mindless (Kim & Sundar, 2012; Nass & Moon) anthropomorphism. However, our conceptualization is compatible with the notion that automatic and unconscious as well as reflective and conscious processes can elicit anthropomorphism (Echterhoff et al., 2006; Kim & Sundar, 2012). Accordingly, our conceptualization can be used to complement the CASA framework, which posits that social responses to machines result from the mindless activation of social scripts (Nass & Moon, 2000). In line with recent broader conceptualizations of the CASA framework (Lombard & Xu, 2021), our conceptualization adds to the framework by specifying an additional route—the mindful or explicit attribution of human mental capacities—through which robots can elicit the use of social rules and heuristics. This explicit route may become increasingly important in the future as robots successively acquire more humanlike characteristics. In fact, we suggest that future research addresses how mindful and mindless processes produce anthropomorphism. This is particularly relevant as mindful and mindless processes may foster different results about the human-likeness of a robot (Banks, 2020; Echterhoff et al., 2006).

Second, our conceptualization of anthropomorphism has not yet been translated into a measurement instrument. Depending on the context of application, different measurement models may be better suited. A reflective model implies that variation in anthropomorphism creates variation in its indicators, which are assumed to be highly correlated (Coltman et al., 2008). Such a model may be adequate in situations in which individuals make quick, intuitive judgments of a robot's human-likeness (e.g., when evaluating robots that are either very human-like or not human-like at all). In contrast, a formative model implies that variation in the

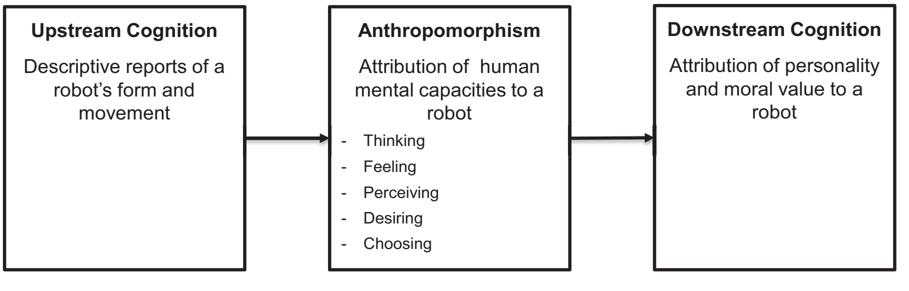

Figure 1. Conceptualization of Anthropomorphism and Distinction from Up- and Downstream Cognitions

Table 1. Summary of Advantages of a More Focused Conceptualization of Anthropomorphism

Adjustment	Affected Quality Criterion	Problem (P), Solution (S), and Advantage (A) for Human–Robot Research
Distinction of anthropomorphism from its precursors	Stability	P: If a definition of anthropomorphism encompasses the perception that a robot moves and behaves like a human, it can only be reasonably applied to robots whose movements and behaviors resemble those of humans, such as NAO (Softbank), but not to robots that move like machines, such as Cozmo (Digital Dream Labs), or animals, such as the robot seal Paro (Intelligent System). For instance, Zhang et al. (2008) propose that anthropomorphism includes "perceivable mannerisms and behaviors," which is a criterion that can be hardly applied to a zoomorphic robot, such as Paro. S: Our conceptualization of anthropomorphism focuses on mental capacities and excludes the shape or movement of robots. Instead, it explicitly defines such concepts as precursors of anthropomorphism. A: Our conceptualization of anthropomorphism can be applied to different types of robots.
Distinction of anthropomorphism from its consequences	Adequacy	P: If an anthropomorphic robot such as Nao (Softbank) is studied and a definition of anthropomorphism subsumes both the attribution of human mental capacities and the attribution of moral value to a robot (e.g., Kahn et al., 2007), causal relationships between the two concepts may be overlooked. In that case, researchers may not consider the possibility that the attribution of human mental capacities may cause, rather than be part of, the attribution of moral value to the robot. S: Our conceptualization of anthropomorphism focuses on mental capacities and does not refer to the attribution of personality and moral value to robots. It explicitly defines such concepts as consequences of anthropomorphism. A: Our conceptualization stimulates a more nuanced and focused thinking about the relationships between anthropomorphism and its consequences.
Distinction of anthropomorphism from similar concepts	Adequacy	P: If a definition of anthropomorphism is not clearly demarcated from similar concepts, such as social presence and animacy, important differences in psychological responses to robots may be overlooked. Similarly, differential effects of anthropomorphism and similar concepts may be overlooked. For instance, Ho and MacDorman's (2010) concept of perceived humanness includes the distinction between inanimate and living, which creates a conceptual overlap with animacy and hinders analyses of the distinct effects of social robots on anthropomorphism and animacy. S: By explicating the conceptual core of anthropomorphism (and by specifying the main dimensions of anthropomorphism), the distinction of anthropomorphism from similar concepts is facilitated. A: Differential effects on anthropomorphism and similar concepts can be studied as well as differential effects of anthropomorphism and similar concepts on other outcomes.
Specification of concrete dimensions of anthropomorphism	Precision	P: Extant definitions of anthropomorphism typically provide an abstract discussion of anthropomorphism that often only cursorily discusses subdimensions and/or only lists examples of subdimensions. For instance, Bartneck et al. (2009) argue that "anthropomorphism includes the attribution of human form, human characteristics, and human behavior to robots" (p. 74). However, without specifying which forms, characteristics, or behaviors should be considered, and without providing a theoretical justification of each aspect of anthropomorphism, the concrete conceptualization and operationalization are left to the discretion of researchers and, thus, crucial dimensions may be overlooked. S: Our conceptualization defines five main dimensions of anthropomorphism, which reduces the room for interpretation. Moreover, we provide a theoretical justification for the main dimensions of robot anthropomorphism. By using ToM as a foundation, we establish that our conceptualization of anthropomorphism is comprehensive and reflects those five mental capacities that have been demonstrated to be essential in the interpretation and prediction of human agents' behavior—i.e., thinking, feeling, perceiving, desiring, and choosing. A: A more consistent and comprehensive conceptualization and operationalization of anthropomorphism is facilitated.

indicators, which are not assumed to be highly correlated, creates variation in anthropomorphism (Coltman et al., 2008). Such a model may be adequate when individuals gradually analyze the capacities of a robot (e.g., when evaluating borderline cases).

Third, a measure of explicit anthropomorphism may need to be complemented with measures of implicit anthropomorphism in HRI to arrive at a comprehensive picture of the concept. Main candidates to assess implicit anthropomorphism are indirect self-report measures (Kim & Sundar, 2012) and neurological measures (e.g., Urquiza-Haas & Kotrschal, 2015), given progress in understanding the neurological foundations of HRI in general and anthropomorphism in particular (Hortensius & Cross, 2018; Urquiza-Haas & Kotrschal, 2015). We believe that the combination of indirect self-report and neurological measures will help us to grasp anthropomorphism better.

In sum, this study has distinguished robot anthropomorphism from its precursors and consequences and highlighted core dimensions of anthropomorphism. Against the backdrop of important developments in social robotics, it is our hope that our conceptualization helps to clarify what anthropomorphism is—and what it is not. Ultimately, we believe that our more focused conceptualization will sharpen and systematize studies in HRI to contribute to truly cumulative knowledge on this important subject.

Funding

Work on this article was made possible by funding from the European Research Council (ERC) under the European Union s Horizon 2020 research and innovation programme, under grant agreement no. 682733 to the second author.

Notes

1 We regard our conceptualization of anthropomorphism as a sufficient but not a necessary condition for being able to use ToM on a robot. Anthropomorphizing—that is, the explicit attribution of human mental capacities to a robot—allows for the application of ToM to a robot. However, the application of ToM to a robot does not presume that human mental capacities have been explicitly attributed to the robot because implicit processes may elicit ToM (see Future research and limitations).

References

Airenti, G. (2018). The development of anthropomorphism in interaction: Intersubjectivity, imagination, and theory of mind. *Frontiers in Psychology*, *9*, 2136. https://doi.org/10.3389/fpsyg.2018.02136

Banks, J. (2020). Theory of mind in social robots: Replication of five established human tests. *International Journal of Social Robotics*, *12*(2), 403–414. https://doi.org/10.1007/s12369-019-00588-x

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). https://doi.org/10.1145/3371382.3378348

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. https://doi.org/10.1007/s12369-008-0001-3

Baumeister, R. F. (2016). Toward a general theory of motivation: Problems, challenges, opportunities, and the big picture. *Motivation and Emotion*, *40*(1), 1–10. https://doi.org/10.1007/s11031-015-9521-y

Beer, J. M., Fisk, A. D., & Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of Human–Robot Interaction*, *3*(2), 74–99. https://doi.org/10.5898/JHRI.3.2.Beer

Benninghoff, B., Kulms, P., Hoffmann, L., & Krämer, N. (2013). Theory of mind in human–robot-communication: Appreciated or not? https://doi.org/10.17185/DUEPUBLICO/31357

Bigman, Y. E., Waytz, A., Alterovitz, R., & Gray, K. (2019). Holding robots responsible: The elements of machine morality. *Trends in Cognitive Sciences*, *23*(5), 365–368. https://doi.org/10.1016/j.tics.2019.02.008

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators and Virtual Environments*, *12*(5), 456–480. https://doi.org/10.1162/105474603322761270

Bjorklund, D. F., & Causey, K. B. (2018). *Children's thinking. Cognitive development and individual differences*. Sage.

Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgment, decision making and reasoning. *Cognition & Emotion*, *24*(4), 561–595. https://doi.org/10.1080/02699930903132496

Breazeal, C. (2003). Toward sociable robots. *Robotics and Autonomous Systems*, *42*(3–4), 167–175. https://doi.org/10.1016/S0921-8890(02)00373-1

Broadbent, E. (2017). Interactions with robots: The truths we reveal about ourselves. *Annual Review of Psychology*, *68*(1), 627–652. https://doi.org/10.1146/annurev-psych-010416-043958

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, *61*(12), 1250–1262. https://doi.org/10.1016/j.jbusres.2008.01.013

de Jong, C., Peter, J., Kühne, R., & Barco, A. (2019). Children's acceptance of social robots: A narrative review of the research 2000–2017. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, *20*(3), 393–425. https://doi.org/10.1075/is.18071.jon

Dennett, D. C. (1987). *The intentional stance*. MIT Press.

Dennett, D. C. (1988). Conditions of personhood. In M. F. Goodman (Ed.), *What is a person*? (pp. 145–167). Humana Press. https://doi.org/10.1007/978-1-4612-3950-5_7

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3-4), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

Echterhoff, G., Bohner, G., & Siebler, F. (2006). "Social Robotics" und Mensch-Maschine-Interaktion ["Social Robotics" and human-machine-interaction]. *Zeitschrift für Sozialpsychologie*, *37*(4), 219–231. https://doi.org/10.1024/0044-3514.37.4.219

Epley, N. (2018). A mind like mine: The exceptionally ordinary underpinnings of anthropomorphism. *Journal of the Association for Consumer Research*, *3*(4), 591–598. https://doi.org/10.1086/699516

Epley, N., & Waytz, A. (2010). Mind perception. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *The handbook of social psychology* (5th ed., pp. 498–541). Wiley.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Eyssel, F. (2017). An experimental psychological perspective on social robotics. *Robotics and Autonomous Systems*, *87*, 363–371. https://doi.org/10.1016/j.robot.2016.08.029

Eyssel, F., Hegel, F., Horstmann, G., & Wagner, C. (2010). Anthropomorphic inferences from emotional nonverbal cues: A case study. In *19th International Symposium in Robot and Human Interactive Communication* (pp. 646–651). https://doi.org/10.1109/ROMAN.2010.5598687

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 - HRI '11* (p. 61). https://doi.org/10.1145/1957656.1957673

Ferrari, F., Paladino, M. P., & Jetten, J. (2016). Blurring human–machine distinctions: Anthropomorphic appearance in social robots as a threat to human distinctiveness. *International Journal of Social Robotics*, *8*(2), 287–302. https://doi.org/10.1007/s12369-016-0338-y

Fink, J. (2012). Anthropomorphism and human likeness in the design of robots and human–robot interaction. In S. S. Ge, O. Khatib, J.-J. Cabibihan, R. Simmons, & M.-A. Williams (Eds.), *Social robotics. 4th International Conference, ICSR 2012, Chengdu, China, October 29-31, 2012, Proceedings* (pp. 199–208). Springer. https://doi.org/10.1007/978-3-642-34103-8_20

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, *42*(3–4), 143–166. https://doi.org/10.1016/S0921-8890(02)00372-X

Fraune, M. R. (2020). Our robots, our team: Robot anthropomorphism moderates group effects in human–robot teams. *Frontiers in Psychology*, *11*, 1275. https://doi.org/10.3389/fpsyg.2020.01275

Fraune, M. R., Oisted, B. C., Sembrowski, C. E., Gates, K. A., Krupp, M. M., & Šabanović, S. (2020). Effects of robot–human versus robot–robot behavior and entitativity on anthropomorphism and willingness to interact. *Computers in Human Behavior*, *105*, 106220. https://doi.org/10.1016/j.chb.2019.106220

Fussell, S. R., Kiesler, S., Setlock, L. D., & Yew, V. (2008). How people anthropomorphize robots. In *Proceedings of the 3rd International Conference on Human Robot Interaction - HRI '08* (p. 145). https://doi.org/10.1145/1349822.1349842

Gong, L. (2008). How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, *24*(4), 1494–1509. https://doi.org/10.1016/j.chb.2007.05.007

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science (New York, N.Y.)*, *315*(5812), 619–619. https://doi.org/10.1126/science.1134475

Gunkel, D. J. (2012). *The machine question. Critical perspectives on AI, robots, and ethics*. MIT Press.

Gunkel, D. J. (2022). The symptom of ethics: Rethinking ethics in the face of the machine. *Human-Machine Communication*, *4*, 67–83. https://doi.org/10.30658/hmc.4.4

Guzman, A. L. (2018). What is human–machine communication, anyway?. In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology, and ourselves* (pp. 1–28). Peter Lang.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *10*(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology*, *65*(1), 399–423. https://doi.org/10.1146/annurev-psych-010213-115045

Haslam, N., Loughnan, S., Kashima, Y., & Bain, P. (2009). Attributing and denying humanness to others. *European Review of Social Psychology*, *19*(1), 55–85. https://doi.org/10.1080/10463280801981645

Hegel, F., Krach, S., Kircher, T., Wrede, B., & Sagerer, G. (2008). Understanding social robots: A user study on anthropomorphism. In *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 574–579). https://doi.org/10.1109/ROMAN.2008.4600728

Higgins, E. T. (1990). Personality, social psychology, and person-situation relations: Standards and knowledge activation as a common language. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research*. (pp. 301–338). The Guilford Press.

Himma, K. E. (2009). Artificial agency, consciousness, and the criteria for moral agency: What properties must an artificial agent have to be a moral agent? *Ethics and Information Technology*, *11*(1), 19–29. https://doi.org/10.1007/s10676-008-9167-5

Ho, C.-C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, *26*(6), 1508–1518. https://doi.org/10.1016/j.chb.2010.05.015

Hortensius, R., & Cross, E. S. (2018). From automata to animate beings: The scope and limits of attributing socialness to artificial agents. *Annals of the New York Academy of Sciences*, *1426*(1), 93–110. https://doi.org/10.1111/nyas.13727

Hubbard, F. P. (2011). "Do androids dream?": Personhood and intelligent artifacts. *Temple Law Review*, *83*(2), 405–474. https://ssrn.com/abstract=1725983

Kahn, P. H., Ishiguro, H., Friedman, B., Kanda, T., Freier, N. G., Severson, R. L., & Miller, J. (2007). What is a Human?: Toward psychological benchmarks in the field of human–robot. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, *8*(3), 363–390. https://doi.org/10.1075/is.8.3.04kah

Kiesler, S., & Goetz, J. (2002). Mental models of robotic assistants. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems - CHI '02* (pp. 576–577). https://doi.org/10.1145/506443.506491

Kiesler, S., Powers, A., Fussell, S. R., & Torrey, C. (2008). Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition*, *26*(2), 169–181. https://doi.org/10.1521/soco.2008.26.2.169

Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior*, *28*(1), 241–250. https://doi.org/10.1016/j.chb.2011.09.006

Krausová, A., & Hazan, H. (2013). Creating free will in artificial intelligence. In J. Romportl, P. Ircing, E. Zackova, M. Polak, & R. Schuster (Eds.), *Beyond AI: Artificial golem intelligence. Proceedings of the International Conference Beyond AI 2013* (pp. 96–109). University of West Bohemia.

LeDoux, J. E. (1999). *The emotional brain: The mysterious underpinnings of emotional life*. Phoenix.

Lee-Won, R. J., Joo, Y. K., & Park, S. G. (2020). Media equation. In J. Bulck (Ed.), *The International Encyclopedia of Media Psychology* (1st ed., pp. 1–10). Wiley. https://doi.org/10.1002/9781119011071.iemp0158

Lemaignan, S., Fink, J., Dillenbourg, P., & Braboszcz, C. (2014). The cognitive correlates of anthropomorphism. In *2014 Human–Robot Interaction Conference, Workshop "HRI: A bridge between Robotics and Neuroscience."*

Leventhal, H., & Scherer, K. (1987). The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition & Emotion*, *1*(1), 3–28. https://doi.org/10.1080/02699938708408361

Liarokapis, M. V., Artemiadis, P. K., & Kyriakopoulos, K. J. (2013). Quantifying anthropomorphism of robot hands. In *2013 IEEE International Conference on Robotics and Automation* (pp. 2041–2046). https://doi.org/10.1109/ICRA.2013.6630850

Lombard, M., & Xu, K. (2021). Social Responses to Media Technologies in the 21st Century: The Media are Social Actors Paradigm. *Human-Machine Communication*, *2*, 29–55. https://doi.org/10.30658/hmc.2.2

Margolin, D. B. (2021). The theory of informative fictions: A character-based approach to false news and other misinformation. *Communication Theory*, *31*(4), 714–736. https://doi.org/10.1093/ct/qtaa002

McLeod, J. M., & Pan, Z. (2005). Concept explication and theory construction. In S. Dunwoody, L. B. Becker, D. M. McLeod, & G. M. Kosicki (Eds.), *The evolution of key mass communication concepts: Honoring Jack M. McLeod*. (pp. 13–76). Hampton Press.

Morewedge, C. K., Preston, J., & Wegner, D. M. (2007). Timescale bias in the attribution of mind. *Journal of Personality and Social Psychology*, *93*(1), 1–11. https://doi.org/10.1037/0022-3514.93.1.1

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Ruijten, P. A. M. (2018). Perceptions of human-likeness in human–robot interaction research. Unpublished manuscript.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

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. https://doi.org/10.1007/s12369-013-0196-9

Scassellati, B. (2002). Theory of mind for a humanoid robot. *Autonomous Robots*, *12*(1), 13–24. https://doi.org/10.1023/A:1013298507114

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299–309. https://doi.org/10.1016/S1364-6613(00)01506-0

Schroeder, J., & Epley, N. (2016). Mistaking minds and machines: How speech affects dehumanization and anthropomorphism. *Journal of Experimental Psychology. General*, *145*(11), 1427–1437. https://doi.org/10.1037/xge0000214

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*(2), 108–131. https://doi.org/10.1207/S15327957PSPR0402_01

Stafford, R. Q., MacDonald, B. A., Jayawardena, C., Wegner, D. M., & Broadbent, E. (2014). Does the robot have a mind? Mind perception

and attitudes towards robots predict use of an eldercare robot. *International Journal of Social Robotics*, *6*(1), 17–32. https://doi.org/10.1007/s12369-013-0186-y
Sullins, J. P. (2006). When is a robot a moral agent? *International Review of Information Ethics*, *6*, 23–30. doi: https://doi.org/10.29173/irie136
Tan, H., Wang, D., & Sabanovic, S. (2018). Projecting life onto robots: The effects of cultural factors and design type on multi-Level evaluations of robot anthropomorphism. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 129–136). https://doi.org/10.1109/ROMAN.2018.8525584
Trovato, G., & Eyssel, F. (2017). Mind attribution to androids: A comparative study with Italian and Japanese adolescents. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 561–566). https://doi.org/10.1109/ROMAN.2017.8172358
Urquiza-Haas, E. G., & Kotrschal, K. (2015). The mind behind anthropomorphic thinking: Attribution of mental states to other species. *Animal Behaviour*, *109*, 167–176. https://doi.org/10.1016/j.anbehav.2015.08.011
Wang, X., & Krumhuber, E. G. (2018). Mind perception of robots varies with their economic versus social function. *Frontiers in Psychology*, *9*, 1230. https://doi.org/10.3389/fpsyg.2018.01230
Waytz, A., Epley, N., & Cacioppo, J. T. (2010). Social cognition unbound: Insights into anthropomorphism and dehumanization. *Current Directions in Psychological Science*, *19*(1), 58–62. https://doi.org/10.1177/0963721409359302
Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383–388. https://doi.org/10.1016/j.tics.2010.05.006
Waytz, A., Morewedge, C. K., Epley, N., Monteleone, G., Gao, J.-H., & Cacioppo, J. T. (2010). Making sense by making sentient: Effectance motivation increases anthropomorphism. *Journal of Personality and Social Psychology*, *99*(3), 410–435. https://doi.org/10.1037/a0020240
Wellman, H. M. (1990). *The child's theory of mind*. MIT Press.
Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition*, *30*(3), 239–277. https://doi.org/10.1016/0010-0277(88)90021-2
Westermann, R. (2000). *Wissenschaftstheorie und Experimentalmethodik: Ein Lehrbuch zur Psychologischen Methodenlehre [Philosophy of science and experimental methods: A textbook* on psychological methdology]. Hogrefe.
Yogeeswaran, K., Złotowski, J., Livingstone, M., Bartneck, C., Sumioka, H., & Ishiguro, H. (2016). The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research. *Journal of Human–Robot Interaction*, *5*(2), 29. https://doi.org/10.5898/JHRI.5.2.Yogeeswaran
Zhang, T., Zhu, B., Lee, L., & Kaber, D. (2008). Service robot anthropomorphism and interface design for emotion in human–robot interaction. In *2008 IEEE International Conference on Automation Science and Engineering* (pp. 674–679). https://doi.org/10.1109/COASE.2008.4626532
Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and challenges in human–robot interaction. *International Journal of Social Robotics*, *7*(3), 347–360. https://doi.org/10.1007/s12369-014-0267-6
Złotowski, J., Strasser, E., & Bartneck, C. (2014). Dimensions of anthropomorphism: From humanness to humanlikeness. In *Proceedings of the 2014 ACM/IEEE International Conference on Human–Robot Interaction - HRI '14* (pp. 66–73). https://doi.org/10.1145/2559636.2559679