Computers in Human Behavior 153 (2024) 108098

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

It feels, therefore it is: Associations between mind perception and mind ascription for social robots

Kevin Koban [a,*], Jaime Banks [b]

[a] Department of Communication, University of Vienna, Austria
[b] School of Information Studies, Syracuse University, United States

ARTICLE INFO

Handling Editor: Matthieu Guitton

Keywords:
Social robots
Mind attribution
Affective capacities
Social cognition

ABSTRACT

As social robots are increasingly designed with sophisticated simulations of human skills, it becomes essential to understand boundary conditions of people's engagement of them as mindful actors. The present paper reports a comprehensive secondary analysis of six studies (total $N = 967$) on the relationship between mind *perception* (evaluation of mental capacities) and mind *ascription* (explicit assignment of a mind-having status) when people consider a humanoid robot in different scenarios. Results indicate there is a context-independent, moderate link between perceptions of affective mental capacity (i.e., ability to feel emotion) and explicit mind ascription. We further found hints for a weak relationship between perceptions of reality-interaction capacity (i.e., sensory and agentic abilities) and decisions to ascribe mind that may need larger samples in order to be validated. Perceptions of social-moral capacities (i.e., evaluations of people and goals) were not a significant predictor of mind ascription. Overall, these findings highlight the pivotal role played by robots' display of affective engagement with the world for the acceptance of robots as mindful "beings" in human spheres.

Although recent technological advances in robotics have yet to keep pace with predictions made by entrepreneurs or futurists, let alone approaching science fiction writers' envisioned future, social robots have become increasingly sophisticated over recent years. A social robot is an embodied entity that simulates social processes and shows behaviors in a manner aligned with the norms of its social environment (see Duffy, 2003). Humanoid robots like *Nadine* (Ramanathan, Mishra, & Thalmann, 2019) may be regarded as prime examples of the current state of social robotics (see Phillips, Zhao, Ullman, & Malle, 2018). From the outside, they *display* human-anatomical features with natural-looking artificial skin and hair. Beyond mere looks, they also have a mental architecture that enables them to *approximate* several human-communicative characteristics (e.g., speech synthesis or affect-expression capabilities).

With these advances, it is increasingly important to examine further how people attribute a social robot qualities typically reserved for human (or other animate) beings that may serve as criteria used to assign it the status of a *being* (i.e., one of an ontological class characterized by subjective awareness of and engagement with the world) rather than an inert *artifact* (Gunkel, 2012; see also Heidegger, 1927/1962). In other words, in understanding the implications of human-machine relations through simulated sociality, it is critical to understand boundary conditions for the engagement of robots as legitimate social actors rather than as mere tools or inert objects. Among many such conditions (see Gunkel, 2018; Guzman, 2020 for overviews), attributions of mindfulness are central to how humans understand others (Dennett, 1996; Gray, Gray, & Wegner, 2007; Thellman, de Graaf, & Ziemke, 2022). Attributions of mindfulness are a form of status assignment that is paramount for people's engagement with them, as mind ascription often comes with respect and moral status (Wegner & Gray, 2016), yet sometimes also with unease and fears (Müller, Gao, Nijssen, & Damen, 2021). Accordingly, mindful status may determine whether individuals (or society, broadly) show benevolent or hostile behaviors toward robot technology (Keijsers & Bartneck, 2018; Keijsers, Bartneck, & Eyssel, 2022).

This paper connects theoretical and empirical work about the social-psychological processes of *mind perception* (i.e., inferencing another agent's potential or actual mental capacities; Malle, 2019) and *mind ascription* (i.e., explicit assignment of a minded status to an agent; Gunkel, 2012) to examine how people's perception of abilities in a social robot are associated with their decision about it having a mind. In contrast to other perspectives (e.g., Thellman et al., 2022), we do not see

* Corresponding author. Währinger Straße 29 (R. 7.36), Vienna, 1090, Austria.
E-mail address: kevin.koban@univie.ac.at (K. Koban).

https://doi.org/10.1016/j.chb.2023.108098
Received 4 January 2023; Received in revised form 8 December 2023; Accepted 12 December 2023
Available online 15 December 2023

mind perception and mind ascription as referring to the same broad phenomenon (that is, mental state attribution) but as conceptually and operationally different: Mind perception is primarily a perceptual process attending to forms of mental capacity, whereas mind ascription is the result of decision- and meaning-making by which a being-status is assigned to the target. Better understanding how mind perception influences explicit mind ascription will illuminate the dynamics by which subtle evaluations may coalesce to support the recognition or rejection of robots as "beings" (Van Der Goot & Etzrodt, 2023; Wykowska, 2021). These dynamics hold implications for whether and how people may assign legal and moral rights to robots as persons (to some degree) and for how artificial agents need to be designed to be accepted in social contexts (Gunkel, 2012; Kühne & Peter, 2023).

1. Ontological categorization of social robots

Humans innately and purposively sort things in the world into categories as a means for efficiently understanding them, often through heuristics (see Bowker & Star, 2000). One kind of sorting is ontological categorization, or the identification of an agent as a *kind* of thing, where humans and machines constitute decidedly different agent categories (Etzrodt, 2021; Guzman, 2020). As summarized by Gunkel (2012), the standard Western position on the ontological categorization of robots (or, earlier on, machines) relies on the presence or lack of functional equivalents of different mental properties (e.g., self-consciousness, rationality, imagination, intentionality, autonomy, interactivity, adaptability, capacity for suffering/enjoyment), where the having or not-having of those functions determines whether an agent may be categorized as a "being." In line with this reasoning, a group of roboticists has argued that current and/or future robotic technologies must have the ontological status of an artifact because they lack (many of) those qualities (see Boden et al., 2017). A central challenge to this objectivist argument, however, is known as the "problem of other minds": Since we cannot directly observe another agent's inner states, any inference about the presence or absence of mind (or qualifying properties thereof) is drawn from observing its behavior; this makes such inferencing vulnerable to mere displays and, thus, epistemologically invalid (Seibt, 2017). Accordingly, the ontological categorization of robots must be based on an individual's evaluation of whether or not they are to be counted as beings and not on any sort of definitive proof of their ontological status (Putman, 1964). In other words, being-status is a matter of psychological and social construction.

In contrast to that standard philosophical position, Coeckelbergh (2011) advocates for determining a robot's ontological status not on the *presence* of any qualifying properties but on the *appearance* of these qualities (for a similar argument, see Danaher, 2020). In other words, once a robot is subjectively experienced as having certain properties, this robot may also be granted the ontological status of a being by the experiencing subject. Following Dennett (1996), perceived mindfulness is at the center of this process. People understand other agents by adopting one of three "stances" toward them: The assumption of mindlessness (i.e., "physical stance"), of mindful design (i.e., "design stance"), or of an assumed mindful status (i.e., "intentional stance"). People predict an agent's behavior in the physical stance according to the laws of physics and in the design stance according to the intentions of a designer, but in the intentional stance according to the assumption that it is an intentional system that acts relative to its own mind. Notably, previous research has shown that humans indeed adopt the intentional stance with non-trivial frequency (e.g., Etzrodt & Engesser, 2021; Marchesi et al., 2019, 2021, 2022) and similarly to when faced with fellow human beings (Thellman, Silvervarg, & Ziemke, 2017).

Relatedly, Wegner and Gray (2016) emphasize that "minds are perceived into existence" (p. 6), meaning that people assume another agent's mind not based on objective facts but on perceptions. By relying on subjective perceptions of mental capacities to determine a mindful status (i.e., by perception superseding manifestation), this position also recognizes that people often do not think of social robots as mindless artifacts. Instead, people often refer to robots as mindful social companions and experience feelings of intimacy (e.g., de Graaf & Allouch, 2017; Kertész & Turunen, 2019; Klamer, Ben Allouch, & Heylen, 2011). However, people typically do not experience robots as fully animate beings either. When asked to classify humans, animals, and robots into two distinct groups based on similarity, people typically group humans and animals, leaving robots outside of the realm of naturalness and animacy (Edwards, 2018). Children and adolescents across different age groups tend to understand social robots as "in-between" beings: partly animate, partly inanimate (Kahn & Shen, 2017). Adults have similar orientations toward robots that display intentionality (Levin, Killingsworth, Saylor, Gordon, & Kawamura, 2013) and after cooperating with voice-based agents (Etzrodt, 2021; Etzrodt & Engesser, 2021; Guzman, 2019). Others have shown that people alternate between viewing a robot as a mindless artifact or a mindful being, depending on the situation: When engaged in meaningful social interaction, they tended to think of it as a companion; when engaged in instrumental use, they tended to consider it a tool (Pradhan, Findlater, & Lazar, 2019). Rather than being a distinct third *kind* somewhere between mindless objects and mindful beings (Kahn & Shen, 2017), this suggests that people's ontological categorizations may be fluid.

Because people cognitively default to human ways of being (Dacey, 2017), the notion of assigning being-qualifying properties can be understood as a matter of *anthropomorphism*. Anthropomorphism is the phenomenon of human beings perceiving and attributing (prototypically) human properties to nonhuman entities (Epley, Waytz, & Cacioppo, 2007). Importantly, anthropomorphism can occur in different ways ranging from very subtle (and often intuitive and cue-driven) inferences about a nonhuman agent's potential or actual mental capacities (i.e., mind perception) to explicit decisions on whether this agent can be assigned some kind of minded status, which typically comes with associated attributions of personhood (i.e., mind ascription; see Złotowski et al., 2018). However, the relationship between mind perception and mind ascription remains unclear thus far.

2. Mind perception and mind ascription

Extant literature has demonstrated that humans *subjectively* make inferences about other minds based on various immediate and mentally stored information (see Achim, Guitton, Jackson, Boutin, & Monetta, 2013). More specifically, people evaluate both human and nonhuman agents for presumed mental capacities from the moment they step into a social situation with them (Epley & Waytz, 2010). These perceptions extend beyond humans, also including reactions to a wide range of animate (e.g., human beings, animals, plants), inanimate (e.g., computers, robots, brands, institutions), and imaginary agents (fictional characters, deities; e.g., Malle, 2019).

2.1. Mind perception

Current scholarship considers mind perception (i.e., inferencing of mental capacities) to be multidimensional (but see, e.g., Tzelios, Williams, Omerod, & Bliss-Moreau, 2022). Most prominently, Gray et al. (2007) established mind perception as a two-dimensional construct consisting of perceived *agency* (i.e., capacities for intentionality and action) and perceived *experience* (i.e., capacities for sensation, feeling, and individuality). Theoretical work from other research areas (e.g., dehumanization, Haslam, Kashima, Loughnan, Shi, & Suitner, 2008; social cognitions, Fiske, Cuddy, & Glick, 2007) highlighted similar differentiations that distinguish between perceptions of agentic and experiential mental capacities. More recent approaches expanded beyond this two-dimensional structure. Weisman et al. (2017) revised Gray et al.'s original work by focusing on similarities among mental capacities (instead of people's perceptions of how strongly certain agents may display them), an approach that resulted in three distinct dimensions:

physiological sensations (i.e., the “body”), social-cognitive and self-regulatory abilities (i.e., the “heart”), and perceptual-cognitive abilities (i.e., the “mind”).

Building upon both Gray et al.’s (2007) original conceptualization and Weisman et al.’s (2017, 2021) re-conceptualization, another three-dimensional model focusing on perceptions of robots was advanced by Malle (2019). Malle argued for a three-dimensional conceptualization that splits the relatively heterogeneous agency dimension into a perceptual-cognitive component (*reality-interaction capacity:* basic agentic capacities such as perception, memory, knowledge, communication) and a social-cognitive component (*social-moral capacity:* higher-order agentic capacities, such as social reasoning, moral cognition, cognitive control); these stand alongside perceived *affective capacity* similar to the original experience dimension, as the ability to sense and feel. When confronted with a robot agent within a social situation, people will make sense of it by determining its perceptual- and social-cognitive agentic and experiential capacities. Throughout the social situation, these estimates are continuously adjusted (Epley, 2015).

2.2. Mind ascription

Mind ascription, in comparison, is the appraisal of an agent’s perceived mindfulness, as people overtly decide whether or not they believe the robot has a (structurally or functionally equivalent) mind and will respond to it as such (see Seibt, 2017). That is, if mind perception is about “seeing mind,” mind ascription addresses whether people are “assigning mind” to another agent. Prior research has postulated that people uniformly deny stating that social machines may be mindful, even though they themselves had shown humanlike social behaviors towards them immediately before (Nass & Moon, 2000). People have been found to be well aware of what separates humans from robots. Guzman (2020) revealed that people can choose from various criteria to differentiate between human and artificial agents. Some of these criteria may be (theoretically or experientially) approachable for social robots in the foreseeable future (e.g., intelligence, autonomy, morality, or emotions); other criteria, however, may more likely be used to refuse them a being-status (e.g., origin, naturalness, or soul). Accordingly, even if robots were to display human-level mental capacities—and even if they may perceive them to have mental capacities—people may still deny mindful status to robots based on these criteria. That is, people can base decisions for or against ascribing mind on a perceived presence (or absence) of certain criteria without factoring in the perceived presence (or absence) of others (Edwards, 2018). On the other hand, evidence also suggests people may be open to machine mind ascription. Some researchers have argued that children may not engage in pretense when they assign mindfulness to social robots during playful interactions (e.g., Kahn et al., 2011; Severson & Carlson, 2010). Among adults, Guzman (2019) demonstrated that many understand artificial entities to be independent social agents, just as Marchesi et al. (2019, 2021, 2022) repeatedly revealed that many people opt for explicitly mentalistic descriptors (and against accommodating or explicitly mechanistic descriptors) when evaluating ambiguous actions of robots. In another domain, anecdotal evidence on soldiers’ relationships with helper robots where machines were awarded full military honors and funerals support the assumption that people readily assign a mind—and even associated status benefits—to a robot that displays very limited mental capacities (Carpenter, 2016). It is therefore conceivable that people’s categorical denial of mindfulness—a position that is long-established in the literature (see Gambino, Fox, & Ratan, 2020 for an overview)—might be due to people’s unwillingness to report mind-perception experiences given that they realized that they had violated a common norm by socially interacting with a very rudimentary computer agent (see Banks, 2019 or, more generally, Nosek, 2007). Rapid technological advances in artificial intelligence and robotic technologies—in tandem with changing social norms about human-robot interactions—might promote more contemporary moderation of this unwillingness (Gambino et al., 2020).

Because empirical evidence is scarce, it is unclear whether people’s explicit ascription of a mind to a robot corresponds with particular perceptions of that robot’s mental capacities. It may be that people make sense of social robots as a function of perceiving certain capacities or that an ascribed mindful status may prompt attention to particular perceived capacities. Alternatively, these capacity perceptions might not necessarily be reflected in their willingness to explicitly ascribe them mindfulness. As evidenced by Nosek (2007), correlations between people’s mental operations and their *subjective experience* of these operations can deviate depending on the phenomenon in question, as people might be unable to specify their mental processing accurately. However, following Nosek’s argumentation, this partial disjoint does not necessarily mean that people are entirely unaware of their mental processing either but rather that they can become aware (and reflect upon it) given an opportunity. Since the relationship between both capacity and status assessments of robots’ mindfulness is not yet well understood, we ask.

RQ1. (How) do perceptions of a robot’s mental capacities (i.e., reality-interaction capacity, affective capacity, and social-moral capacity) correspond with explicit mind ascription?

3. Method

To address the posed research question, we drew from six studies of human evaluations of a particular social robot to conduct secondary analyses examining the relation between mind perception and mind ascription. Those six studies all (a) used the same stimulus robot and (b) adopted identical measures of mind perception and ascription. The present research question was not asked, and the present analyses were not conducted in those source studies. The datasets can be drawn from the studies’ respective OSF sites.[1]

In each study, people viewed or participated in different activities with a robot before answering items measuring mind perceptions and mind ascription. In Studies 1 and 2, participants participated in or observed a series of canonical Theory of Mind tests with a stimulus robot (Banks, 2021). In Study 3, participants engaged in different activities varied according to norms and goals (those associated with social interaction, tasks, and play), in which a social robot partner demonstrated agentic and experiential capacities (Banks, Koban, & Chauveau, 2021). In Study 4, participants considered videos depicting robots’ behaving (im)morally, and in Studies 5 and 6, participants viewed and evaluated morally ambiguous robot behaviors (Banks & Koban, 2021).

3.1. Participants

Descriptive information for each study is displayed in Table 1. Sampling strategies varied across studies: (1) a mixed approach combining convenient samples of university students and participants from crowdsourcing platforms; (2, 3, 5) convenience samples of university students; and (4, 6) U.S.-representative samples (in terms of age, gender, education, and political affiliation) garnered from panel services. Notably, the respective experimental manipulations in Studies 1, 2, and 4 included a condition in which participants were not watching or interacting with a social robot but a human. Participants interacting with a human (rather than a robot) were excluded from the present analyses.

[1] Study 1 materials: https://osf.io/hcgkm/, Study 2 materials: https://osf.io/8yb67/, Study 3 materials: https://osf.io/n87bg/, Study 4/5/6 materials: https://osf.io/6kqbn/.

Table 1
Demographic information for all included studies.

	Age in years *M (SD)*	Sex *n*	Ethnicity/Race identification *n*	Highest educational degree *n*	Political affiliation *n*
Study 1 (*n* = 125)	33.25 (11.22)	64 female 60 male 1 NA	96 Caucasian 9 Asian 6 African 4 Hispanic 5 Mixed/other		
Study 2 (*n* = 74)	20.30 (1.78)	39 female 35 male	50 Caucasian 7 Asian 7 Middle Eastern 4 African 3 Hispanic 3 Mixed		
Study 3 (*n* = 106)	20.42 (3.55)	66 female 40 male	66 Caucasian 18 Hispanic 8 African 5 Asian 1 Middle Eastern 1 Native American 7 Mixed		
Study 4 (*n* = 176)	48.39 (17.97)	90 female 86 male		38 High School/GED 37 Technical School/ Associate 35 Bachelor 33 Graduate 31 Less than High School/ GED	72 Moderate 56 Conservative 48 Liberal
Study 5 (*n* = 76)	20.83 (3.87)	51 female 25 male	38 Caucasian 16 Hispanic 11 African 7 Mixed 2 Asian 1 Middle Eastern 1 Indian		
Study 6 (*n* = 410)	45.90 (18.43)	213 female 197 male		121 Technical School/ Associate 113 High School/GED 79 Bachelor 54 Less than High School/ GED 43 Graduate	139 Moderate 139 Conservative 132 Liberal

3.2. Measures

Across the studies, mind perception was measured with a 20-item scale proposed by Malle (2019). Participants indicated how much they agree or disagree with the robot being capable of several mental capacities on a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree). The scale consists of three dimensions: reality-interaction capacity (four items, e.g., "moving on their own"), affective capacity (eight items, e.g., "feeling happy"), and social-moral capacity (eight items, e.g., "telling right from wrong"). Mind ascription was operationalized with a dichotomous item asking to ascribe ("yes") or deny ("no") mindful status, i.e., by stating whether or not they think the robot has "a mind" based on what they saw and heard from it (see Table 2). Each study measured several other variables relevant to the original analyses but irrelevant to this paper's research question. The extracted data and scripts specific to the present analyses are available online: https://osf.io/eab7u.

Table 2
Descriptive information and Cronbach's alpha scores for all included measures for each included study.

	Reality-Interaction Capacity α, *M (SD)*	Affective Capacity α, *M (SD)*	Social-Moral Capacity α, *M (SD)*	Mind Ascription	
				'No' *n* (%)	'Yes' *n* (%)
Study 1 (*N* = 125)	.78, 5.53 (1.24)	.95, 2.68 (1.64)	.92, 3.81 (1.52)	37 (29.60%)	88 (70.40%)
Study 2 (*N* = 74)	.74, 5.50 (1.16)	.89, 3.73 (1.42)	.85, 4.07 (1.20)	35 (47.30%)	39 (52.70%)
Study 3 (*n* = 106)	.78, 4.80 (1.24)	.89, 2.55 (1.26)	.89, 3.73 (1.51)	56 (52.83%)	50 (47.17%)
Study 4 (*N* = 176)	.78, 4.59 (1.37)	.96, 3.01 (1.78)	.92, 4.27 (1.88)	109 (61.93%)	67 (38.07%)
Study 5 (*N* = 76)	.63, 5.10 (0.99)	.92, 3.36 (1.43)	.88, 4.49 (1.35)	34 (44.74%)	42 (55.26%)
Study 6 (*N* = 410)	.78, 4.87 (1.18)	.93, 3.50 (1.61)	.90, 4.62 (1.54)	180 (43.90%)	230 (56.10%)

3.3. Stimulus robot

In all cases, participants watched/interacted with a *Robothespian 4* (Engineered Arts, U.K) that was equipped with lighted white body shells and the Socibot head, using a female American-English voice (default *Heather* voice; see individual study materials for detailed information and videos). In Studies 3, 4, 5, and 6, the robot exhibited a humanlike face (default *Pris* face); in Studies 1 and 2, the robot exhibited either that humanlike face or a machine-like face (Robothespian's default *Robot 1* face) for a group of participants. The robot was consistently named "Ray" and addressed with feminine pronouns (minimizing idiosyncratic gendering by participants; see Seaborn & Frank, 2022). All displayed behaviors were executed by a hidden human controller via Wizard of Oz procedure, using a pre-determined set of responses that were tailored in each study according to its respective requirements.

3.4. Analysis procedure

We analyzed the datasets using a hybrid approach—individual-study analyses followed by a "mini meta-analysis" (see Goh, Hall, & Rosenthal, 2016).

The individual studies' analyses were conducted discretely using the same procedure, addressing RQ1 via logistic regressions conducted individually for each study. In those regressions, participants' perceptions of reality interaction, affective, and social-moral capacity were continuous predictors, whereas mind ascription was a dichotomous criterion variable. To estimate overall fit, regression models were tested via Chi-Squared test and compared regarding their relative model fit (lower scores indicate better fit) using the Akaike information criterion (AIC) against a baseline model that only included an intercept. Additionally, we conducted sensitivity analysis to estimate the models' ability to correctly identify "true positives" (i.e., how many people who answered positively were correctly predicted as such by the model) as well as specificity analysis to estimate the models' ability to accurately identify "true negatives" (i.e., how many people who answered negatively were correctly predicted as such by the model). Both the models' sensitivity and specificity were plotted against each other in a ROC curve to evaluate how well each model discriminates between people with positive and negative answers. Here, the area under the curve (AUC) is estimated as an indicator with scores between .5 (no discrimination above chance) and 1 (perfect discrimination), where scores above 0.7 are typically considered as acceptable, above 0.8 as excellent, and above

0.9 as outstanding (Hosmer & Lemeshow, 2000). Lastly, individual predictors' unstandardized Beta coefficients were tested for significance, and odds ratios were reported to facilitate interpretability. An odds ratio >1 indicates that positive answers become more likely, while an odds ratio <1 indicates that positive answers become less likely with increasing predictor scores.

Because potential covariates (e.g., attitudes towards robots/technology, robot experience) did not show any meaningful impact on the analysis, they were not included in the present analyses (see online supplements for full results). Following the single-study analyses, a meta-analytic analysis was conducted to discern any aggregate associations.

4. Results by individual study

4.1. Study 1: Perception/ascription in the context of mentalizing tests (mediated)

The regression model predicted mind ascription significantly better than the null model (including only the intercept; see Table 3). Sensitivity analysis showed that observed positives (i.e., 'yes' answers) were predicted by the model as positive (i.e., true positives) with a probability of 88.6%; specificity analysis demonstrated a 67.6% probability that observed negatives (i.e., 'no' answers) were predicted as negative (i.e., true negatives). When plotted against each other, the empirical ROC curve revealed an AUC of 0.891 (indicating excellent discrimination). Concerning the individual predictors, only perceptions of affective capacity were significantly associated with mind ascription (Table 3). These results indicate that people who perceive the robot as capable of feeling are more likely to ascribe a mind to that robot.

4.2. Study 2: Perception/ascription in the context of mentalizing tests (co-present)

This regression model predicted mind ascription significantly better than the null model (see Table 4). Sensitivity and specificity analysis showed that true positives were predicted with a 65.7% probability, while true negatives were predicted with a 71.8% probability. The empirical ROC curve had an AUC of 0.784 (suggesting an acceptable discrimination). Analysis of the individual predictors was in line with the results of Study 1. Only participants' perception of affective capacity significantly predicted participants' decision to ascribe mind to the robot (Table 4).

4.3. Study 3: Perception/ascription in the context of socializing, tasks, and play (co-present)

The regression model turned out to be significantly better than the null model (see Table 5). True positives were predicted with a probability of 80.4%; true negatives with a probability of 68.0%. The empirical ROC curve demonstrated an AUC of 0.770 (again, indicating acceptable discrimination). Similar to the previous analysis, participants' perception of affective capacity was the sole significant predictor of mind ascription (Table 5). According to these data, higher perceptions of affective capacity are linked with a greater tendency to ascribe a mind to a social robot.

4.4. Study 4: Perception/ascription in the context of moral and immoral behavior (mediated)

The regression model predicting mind ascription demonstrated a better fit than the null model (Table 6). True positives were predicted

Table 3
Results of the logistic regression analysis in Study 1.

	B (SE)	z	p	Odd's ratio
Model 0:				
Residual deviance (df = 124) = 151.86				
AIC: 153.86				
Intercept	−0.87 (0.20)	−4.42	<.001	
Model 1:				
Residual deviance (df = 121) = 96.50				
χ^2 (3) = 55.36, p < .001				
AIC: 104.50				
McFadden's R^2 = .36				
Intercept	−7.56 (1.83)	−4.14	<.001	
Reality Interaction Capacity	0.44 (0.31)	1.41	.158	1.55
Affective Capacity	**0.77 (0.23)**	**3.41**	**.001**	**2.16**
Social-Moral Capacity	0.43 (0.31)	1.40	.161	1.54

Note: Rows in bold indicate significant univariate effects.

Table 4
Results of the logistic regression analysis in Study 2.

	B (SE)	*z*	*p*	Odd's ratio
Model 0:				
Residual deviance (df = 73) = 102.37				
AIC: 104.37				
Intercept	0.11 (0.23)	0.47	.642	
Model 1:				
Residual deviance (df = 70) = 81.51				
χ^2 (3) = 20.86, p < .001				
AIC: 89.51				
McFadden's R^2 = .20				
Intercept	−3.68 (1.52)	−2.42	.016	
Reality Interaction Capacity	−0.13 (0.31)	−0.43	.666	0.88
Affective Capacity	**0.61 (0.26)**	**2.35**	**.019**	**1.84**
Social-Moral Capacity	0.56 (0.36)	1.55	.120	1.75

Note: Rows in bold indicate significant univariate effects.

Table 5
Results of the logistic regression analyses in Study 3.

	B (SE)	z	p	Odd's ratio
Mind ascription				
Model 0:				
Residual deviance (df = 105) = 146.61				
AIC: 148.61				
Intercept	−0.11 (0.19)	−0.58	.560	
Model 1:				
Residual deviance (df = 102) = 118.57				
χ^2 (3) = 28.03, p < .001				
AIC: 126.57				
McFadden's R^2 = .19				
Intercept	−4.04 (1.08)	−3.75	<.001	
Reality Interaction Capacity	0.49 (0.27)	1.79	.074	1.63
Affective Capacity	**0.71 (0.24)**	**2.91**	**.004**	**2.03**
Social-Moral Capacity	−0.06 (0.22)	−0.28	.778	0.94

Note: Rows in bold indicate significant univariate effects.

Table 6
Results of the logistic regression analyses in Study 4.

	B (SE)	z	p	Odd's ratio
Mind ascription				
Model 0:				
Residual deviance (df = 175) = 233.87				
AIC: 235.87				
Intercept	−0.49 (0.16)	−3.14	.002	
Model 1:				
Residual deviance (df = 172) = 193.45				
χ^2 (3) = 40.42, p < .001				
AIC: 201.45				
McFadden's R^2 = .17				
Intercept	−2.86 (0.71)	−4.02	<.001	
Reality Interaction Capacity	0.35 (0.21)	1.67	.096	1.42
Affective Capacity	**0.61 (0.15)**	**4.03**	**< .001**	**1.84**
Social-Moral Capacity	−0.27 (0.16)	−1.72	.085	0.76

Note: Rows in bold indicate significant univariate effects.

with a probability of 85.3%; true negatives with a probability of 50.7%. The empirical ROC curve had an AUC of 0.780 (suggesting an acceptable discrimination). Participants' perception of affective capacity was the only significant predictor of mind ascription (Table 6).

4.5. Study 5: Perception/ascription in the context of moral ambiguity (mediated)

The regression model for mind ascription had no better fit than the null model (Table 7). True positives were predicted with a 50.0% probability and true negatives with a 73.8% probability. The empirical ROC curve had an AUC of 0.647 (indicating poor discrimination). None of the mental-capacity dimensions were significant predictors of mind ascription (Table 7).

4.6. Study 6: Perception/ascription in the context of moral ambiguity (co-present)

The regression model for mind ascription had a better fit than the null model (Table 8). True positives were predicted with a 65.0% probability; true negatives with an 80.9% probability. The empirical ROC curve had an AUC of 0.810 (indicating excellent discrimination). Perceptions of reality-interaction and affective capacities were significant predictors of mind ascription (Table 8).

4.7. Meta-analysis

Although meta-analyses typically comprise a large number of studies (see Anker et al., 2010), researchers have advocated for smaller-scaled ("mini") meta-analysis of a series of studies within a manuscript in order to obtain a more reliable estimate for the overall size of the effect in question (Goh et al., 2016). The necessary requirements for conducting such an analysis were met as all six studies applied similar methods (e.g., same stimulus robot, measures of mind perception, and mind ascription) across different scenarios (i.e., mediated vs. live interactions) using different sampling strategies (i.e., convenience, representative, and mixed sampling). We, therefore, performed three separate meta-analyses ($k = 6$, $N = 967$), one for each dimension of mind perception, using fixed-effects models in which each predictor's odds ratios were weighted by standard error and served as the mean effect size.

4.8. Reality-interaction capacity

Perceived reality-interaction capacity emerged as a significant predictor of mind ascription only in Study 6. Nevertheless, the overall effect turned out significant ($M_{OR} = 1.40$, z (5) = 3.44, $p = .001$) and homogeneous across all six studies ($I^2 = 0$, Q (5) = 3.86, $p = .570$; Fig. 1), indicating that perceptions of a robot's perceptual-cognitive abilities are possibly, albeit weakly, linked with a greater tendency to ascribe a mind to it.

Table 7
Results of the logistic regression analyses in Study 5.

	B (SE)	z	p	Odd's ratio
Mind ascription				
Model 0:				
Residual deviance (df = 75) = 104.51				
AIC: 106.51				
Intercept	0.21 (0.23)	0.92	.360	
Model 1:				
Residual deviance (df = 72) = 98.65				
χ^2 (3) = 5.86, $p = .119$				
AIC: 106.65				
McFadden's $R^2 = .06$				
Intercept	−1.54 (1.28)	−1.20	.231	
Reality Interaction Capacity	0.01 (0.33)	0.04	.972	1.01
Affective Capacity	0.34 (0.20)	1.67	.094	1.41
Social-Moral Capacity	0.13 (0.24)	0.52	.602	1.13

Table 8
Results of the logistic regression analyses in Study 6.

	B (SE)	z	p	Odd's ratio
Mind ascription				
Model 0:				
Residual deviance (df = 409) = 562.27				
AIC: 564.27				
Intercept	0.25 (0.10)	2.46	.014	
Model 1:				
Residual deviance (df = 406) = 433.17				
χ^2 (3) = 129.10, $p < .001$				
AIC: 441.17				
McFadden's $R^2 = .23$				
Intercept	−4.46 (0.61)	−7.36	<.001	
Reality Interaction Capacity	**0.42 (0.15)**	**2.80**	**.005**	**1.53**
Affective Capacity	**0.53 (0.10)**	**5.57**	**< .001**	**1.70**
Social-Moral Capacity	0.19 (0.12)	1.62	.105	1.21

Note: Rows in bold indicate significant univariate effects.

4.9. Affective capacity

Participants' perceptions of Ray as having the capacity to feel emotions were significant predictors of mind ascription in all studies, except for Study 5. Across all six studies, the overall effect was significant ($M_{OR} = 1.76$, z (5) = 8.50, $p < .001$) and homogeneous ($I^2 = 0$, Q (5) = 2.53, $p = .772$; Fig. 2). These results suggest a solid association between perceptions of affective capacity and mind ascription.

4.10. Social-moral capacity

Perceptions of social-moral capacities did not emerge as a significant predictor in any of the six studies. Likewise, the overall effect was not significant ($M_{OR} = 1.08$, z (5) = 0.96, $p = .335$) with little heterogeneity across studies ($I^2 = 0.46$, Q (5) = 9.20, $p = .101$; Fig. 3), indicating that perceived ability to make social and moral inferences may not be linked to overt ascription of minded status.

5. Discussion

In conducting secondary analyses of six studies, the present inquiry explored whether the perception of various mental capacities is associated with mind ascription to a humanoid social robot. Our analyses indicate that people's perception of a robot's affective capacity is the primary indicator of their likelihood of elevating that robot to the status of a mindful agent. A small-scale meta-analysis indicates these associations were stable across all six studies, irrespective of different sampling strategies or whether they accessed the robot through live exposure or video presentation.

5.1. Affective capacity as the primary correlate of mental status

Altogether, perceptions of humanoid robots' affective capacity are associated with mind ascription. Perceptions of reality-interaction capacity were only weakly (and rarely significantly) linked, and social-moral capacities had no meaningful relationship to participants' tendency to ascribe a mind to that social robot. Although our data analysis does not allow for reasonable inferencing as to the temporal relationship between mind perception and mind ascription (as both constructs were measured via self-reports after participants were exposed to the robot, thus making them vulnerable to intentional or unintentional bias; see Nosek, 2007), these correlative findings align with prior research highlighting perceptions of affective capacity as an important milestone for people's ontological categorizations. Our results extend previous work suggesting that emotions are a pivotal mental property considered

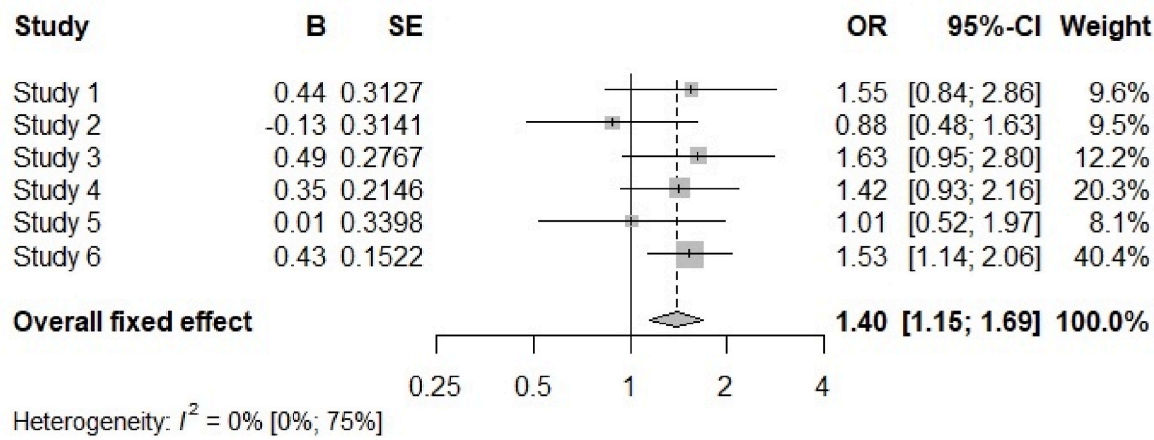

Fig. 1. Forest plot of the link between reality-interaction capacity and mind ascription.
Note: The Odds Ratio (OR) scale is logarithmic. The solid vertical line represents an OR of 1 (i.e., no effect). The dashed vertical line represents the mean effect size. The size of each square indicates the weight of the respective study; whiskers represent the 95% confidence interval of the effects.

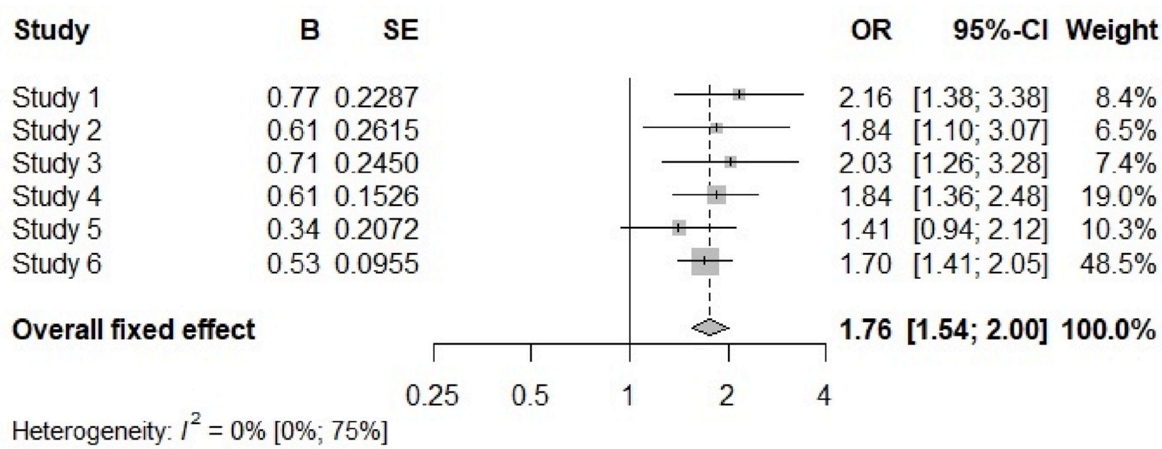

Fig. 2. Forest plot of the link between affective capacity and mind ascription.
Note: The Odds Ratio (OR) scale is logarithmic. The solid vertical line represents an OR of 1 (i.e., no effect). The dashed vertical line represents the mean effect size. The size of each square indicates the weight of the respective study; whiskers represent the 95% confidence interval of the effects.

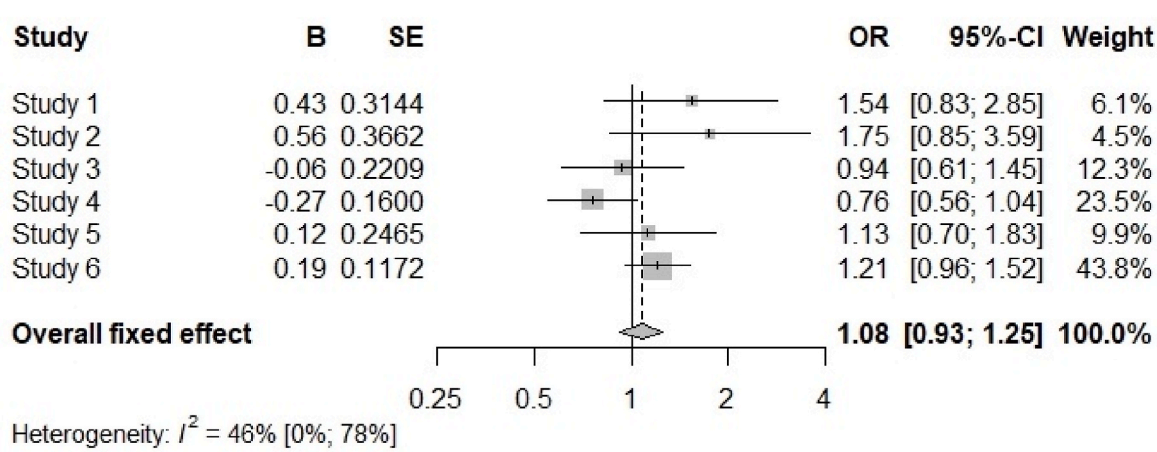

Fig. 3. Forest plot of the link between social-moral capacity and mind ascription.
Note: The Odds Ratio (OR) scale is logarithmic. The solid vertical line represents an OR of 1 (i.e., no effect). The dashed vertical line represents the mean effect size. The size of each square indicates the weight of the respective study; whiskers represent the 95% confidence interval of the effect.

when people are understanding agents as beings in general (Edwards, 2018; Guzman, 2020) and closely related to social cognitions toward robots (e.g., Spatola & Wudarczyk, 2021; Ward, Olsen, & Wegner, 2013). Similarly, research on dehumanization has not only indicated that people refer to a perceived lack of affective capacity when they assign another human agent with the status of a robot (i.e., mechanistic dehumanization; Haslam, 2006) but also that they contrast robots to humans primarily with notions of missing emotional states (Haslam et al., 2008). In other words, social robots are not expected to engage emotionally with their environments; if they are nonetheless perceived to be emotionally engaged and thus have affective capacity, this may manifest a (potentially positive) expectancy violation. If so, this perceived affective capacity may be more influential for people's being-status assignments than are perceived physical or logical engagements—those may be somewhat expected already and, therefore, less likely to influence being-status assignments.

Due to the cross-sectional nature of the source studies that precludes causal claims, the present findings can be interpreted in both directions: On the one hand, it could be that once a robot is perceived as possessing affective capacity, people are more likely to categorize it as a mindful being (rather than as a mindless artifact). This interpretation would follow a default-interventionist dual-process logic in which capacity perceptions are intuited very quickly during an encounter and (partially) inform the slower-processed reflective assignments of mind (Epley, 2015). On the other hand, it could also be that people quickly decided on ascribing a mind to the robot during exposure (as an intuited social cognition; see Bohl & van den Bos, 2012), which then may have led them to subsequently rationalize their mind ascription in giving higher affective-capacity scores. Future research should engage dual-process methodologies (e.g., two-response procedures; see Thompson, Prowse Turner, & Pennycook, 2011) toward important insights into which order may be at work here.

We also note that both mind perception and mind ascription were measured through self-reports varying in directness (i.e., how clearly the measurement refers to the construct in question; see Corneille & Hütter, 2020). While the assessment of mind perception (via Likert-scales) might be regarded as moderately direct (suited for assessing people's perceptions of mental capacities; Takahashi, Ban, & Asada, 2016), a dichotomous measure that prompted participants to explicitly decide the robot's (un)minded status may be considered more direct—i.e., tapping more strongly into people's reflection of their subjective experience of mental operations than into the operations themselves (Nosek, 2007). Previous research provides convincing evidence that more reflective reasoning does not necessarily result in a more accurate assessment but often attempts to rationalize people's first impressions (e.g., Haidt, 2012). Future studies should implement a combination of indirect *in-situ* measures in tandem with direct post-hoc self-reports to uncover the temporality of robot mind perception and mind ascription.

5.2. Implications of affect-driven mindfulness on being-status

The reported link between perceived affective capacity and explicit mind ascription can also be read as supporting evidence of people's aversion to robots that exhibit signs of emotion. Specifically, these results align with previous findings that robots perceived to have affective capacities elicit aversive reactions from people, who typically experience them as eerie or uncanny (Stein & Ohler, 2017). Against this backdrop, our findings suggest that these aversive responses may result primarily from a perceived threat to (or unpleasant uncertainty about) human distinctiveness (Złotowski, Yogeeswaran, & Bartneck, 2017). Previous research has shown that robots that appear subjectively aware of their emotional states and affectively engage with the living world around them are neither expected nor preferred (Malle & Thapa Magar, 2017). They contradict people's existing mental models of what a social robot should be able to do (Banks, 2020) and may even remind them of dystopian science fiction tropes (Bruckenberger et al., 2013). The present findings thus point to affect-driven mind perception as a potential driver of this uncertainty, fear, and anxiety in its link with ascribed mind—only a minded robot may create strong feelings of abjection (see Banks, Koban, & Haggadone, 2023) or pose an existential threat (Müller et al., 2021).

Beyond experiential capacities, recent studies have demonstrated that a robot's display of agency can provoke aversion in terms of how robots could function in society (Appel, Izydorczyk, Weber, Mara, & Lischetzke, 2020). According to our findings, it may be likely that such aversive feelings are partially grounded in perceived reality-interaction capacity (but not in social-moral capacity). That is, this operational aversion may not be primarily caused by people's (fatalistic) picture of robots' potentially superior higher-order agentic capacities (see Grundke, Stein, & Appel, 2022) but by presumptions about how their basic agentic capacities might affect the job market or whether they might potentially pose a threat to their physical safety (Stein, Liebold, & Ohler, 2019). Future research should consider the potential interplays between mental capacity and mindful status as an explanatory mechanism to elaborate people's feelings about artificial agents' threats.

5.3. Practical implications

Since robots increasingly engage in interactions with the living world that often include a non-trivial risk for human lives (e.g., rescue robots used in natural or human-caused disasters), our findings have important practical implications. While criminal law is constantly adjusting to ongoing technological advances (Simmler & Markwalder, 2019), moral psychologists have shown that robots with sophisticated mental architectures are more likely to be held responsible for wrongdoings when they appear to be not only physically but also affectively engaging with the world, and they projected that this trend would accelerate with advanced programming (Bigman, Waytz, Alterovitz, & Gray, 2019). The observed link between perceived affective capacities and mind ascription may inform the manufacture of robotic technologies in terms of the mental architectures or simulative cues that should be prioritized in development and design—in particular for scenarios where machine-agent actions may get moralized either directly or indirectly (see Coeckelbergh, 2021). Specifically, our findings point to differential recommendations depending on desired human reactions: It may be important to avoid affective architectures for contexts in which public outcries (and costly lawsuits) may be following circumstances that could be perceived as moral transgression; conversely, it may be important to engage the affect-expressive potentials of humanoid robots in contexts where emotion can benefit the adoption and acceptance of such a robot, as with supportive companion machines or within healthcare contexts where expressed empathy can be meaningful.

Notably, it is unclear whether our findings and, thus, these implications can be generalized from social robots to other artificial agents, particularly disembodied ones like virtual assistants (e.g., Alexa or Siri) or AI-powered chatbots (e.g., Replika or ChatGPT). Although existing research suggests that disembodied machine agents are often understood as remarkably similar to embodied agents (e.g., Etzrodt & Engesser, 2021; Guzman, 2019), others have stressed that embodiment and corporeality have meaningful impact on people's perception, evaluation, and ascription (e.g., Hoffmann, Bock, Rosenthal, & Pütten, 2018; Roesler, Manzey, & Onnasch, 2023). Until agent-comparative research addresses this issue, we call for caution when attempting to draw implications for disembodied machine agents.

5.4. Limitations

Although the source studies, together, avoid some standard limitations of empirical research (e.g., exclusive focus on a single stimulus scenario/presentation mode, convenience sampling, lack of replicating evidence), our analyses are still subject to various shortcomings. Most notably, all included studies were originally designed to investigate research questions different from the one this paper addresses. Since we are aware (and supportive) of current efforts to increase the robustness of human-machine communication (and, more broadly, psychological) scholarship, we emphasize that our findings are exploratory and should be confirmed with an independent preregistered study. Apart from that, it must be mentioned that each of the six studies used the same stimulus robot. Although this stimulus consistency adds to the internal validity of our findings, it also reduces generalizability as the focal humanoid robot cannot be considered representative of available robots—especially with respect to common zoomorphic and mechanomorphic embodiments. Lastly, our analyses are based on quantitative self-report measures, which may oversimplify the complexities of people's mind perceptions and mind ascription, do not allow for definitive evidence on the temporality of both processes, and may be vulnerable to desirability bias. We, therefore, encourage additional research to replicate the present findings, combining self-reports with behavioral or psychophysiological indicators to overcome these common methodological shortcomings.

6. Conclusion

The present analysis indicates that people's perceptions of a robot as an emotional entity are linked with a greater probability that they also explicitly assign it the status of a mindful actor (rather than an inert artifact). As perceptions of mindfulness are central to more overt assignment of a "being" status (at least among Western individuals; see Spatola, Marchesi, & Wykowska, 2022), these findings suggest that perception of emotional capacities is likely an *essential* cornerstone of people's ontological categorization of social robots and other artificial agents as deserving a stronger agentic status in society.

CRediT authorship contribution statement

Kevin Koban: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing, Resources. **Jaime Banks:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing - review & editing.

Declaration of competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability

Data/code will be openly shared.

Acknowledgment

This paper is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-19-1-0006.

References

Achim, A. M., Guitton, M., Jackson, P. L., Boutin, A., & Monetta, L. (2013). On what ground do we mentalize? Characteristics of current tasks and sources of information that contribute to mentalizing judgments. *Psychological Assessment, 25*(1), 117–126. https://doi.org/10.1037/a0029137

Appel, M., Izydorczyk, D., Weber, S., Mara, M., & Lischetzke, T. (2020). The uncanny of mind in a machine: Humanoid robots as tools, agents, and experiencers. *Computers in Human Behavior, 102*, 274–286. https://doi.org/10.1016/j.chb.2019.07.031

Banks, J. (2019). A perceived moral agency scale: Development and validation of a metric for humans and social machines. *Computers in Human Behavior, 90*, 363–371. https://doi.org/10.1016/j.chb.2018.08.028

Banks, J. (2020). Optimus primed: Media cultivation of robot mental models and social judgments. *Frontiers in Robotics and AI, 7*, 62. https://doi.org/10.3389/frobt.2020.00062

Banks, J. (2021). Of like mind: The (mostly) similar mentalizing of robots and humans. *Technology, Mind, and Behavior, 1*(2). https://doi.org/10.1037/tmb0000025

Banks, J., & Koban, K. (2021). Framing effects on judgments of social robots' (Im)Moral behaviors. *Frontiers in Robotics and AI, 8*, Article 627233. https://doi.org/10.3389/frobt.2021.627233

Banks, J., Koban, K., & Chauveau, P. de V. (2021). Forms and frames: Mind, morality, and trust in robots across prototypical interactions. *Human-Machine Communication, 2*, 81–103. https://doi.org/10.30658/hmc.2.4

Banks, J., Koban, K., & Haggadone, B. (2023). Avoiding the abject and seeking the script: Perceived mind, morality, and trust in a persuasive social robot. *ACM Transactions on Human-Robot Interaction, 12*(3), 32. https://doi.org/10.1145/3572036

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

Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., et al. (2017). Principles of robotics: Regulating robots in the real world. *Connection Science, 29*(2), 124–129. https://doi.org/10.1080/09540091.2016.1271400

Bohl, V., & van den Bos, W. (2012). Toward an integrative account of social cognition: Marrying theory of mind and interactionism to study the interplay of type 1 and type 2 processes. *Frontiers in Human Neuroscience, 6*, 274.

Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. MIT Press.

Bruckenberger, U., Weiss, A., Mirnig, N., Strasser, E., Stadler, S., & Tscheligi, M. (2013). The good, the bad, the weird: Audience evaluation of a "real" robot in relation to science fiction and mass media. In G. Herrmann, M. J. Pearson, A. Lenz, P. Bremner, A. Spiers, & U. Leonards (Eds.), *Social robotics* (Vol. 8239, pp. 301–310). Springer International Publishing. https://doi.org/10.1007/978-3-319-02675-6_30.

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

Coeckelbergh, M. (2011). Humans, animals, and robots: A phenomenological approach to human-robot relations. *International Journal of Social Robotics, 3*(2), 197–204. https://doi.org/10.1007/s12369-010-0075-6

Coeckelbergh, M. (2021). Should we treat teddy bear 2.0 as a kantian dog? Four arguments for the indirect moral standing of personal social robots, with implications for thinking about animals and humans. *Minds and Machines, 31*(3), 337–360. https://doi.org/10.1007/s11023-020-09554-3

Corneille, O., & Hütter, M. (2020). Implicit? What do you mean? A comprehensive review of the delusive implicitness construct in attitude research. *Personality and Social Psychology Review, 24(3), 212–232*. https://doi.org/10.1177/1088868320911325

Dacey, M. (2017). Anthropomorphism as cognitive bias. *Philosophy of Science, 84*(5), 1152–1164.

Danaher, J. (2020). Welcoming robots into the moral circle: A defence of ethical behaviourism. *Science and Engineering Ethics, 26*(4), 2023–2049. https://doi.org/10.1007/s11948-019-00119-x

Dennett, D. C. (1996). Kinds of minds: Toward an understanding of consciousness. *BasicBooks*.

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

Edwards, A. P. (2018). Animals, humans, and machines: Interactive implications of ontological classification. In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology and ourselves* (pp. 29–49) (Peter Lang).

Epley, N. (2015). *Mindwise: How we understand what others think, believe, feel and want*. Penguin Books.

Epley, N., & Waytz, A. (2010). Mind perception. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 498–541). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470561119.socpsy001014.

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

Etzrodt, K. (2021). The ontological classification of conversational agents: An adaptation of piaget's equilibration theory. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, E. Luger, M. Goodwin, et al. (Eds.), *Chatbot research and design* (Vol. 12604, pp. 48–63). Springer International Publishing. https://doi.org/10.1007/978-3-030-68288-0_4.

Etzrodt, K., & Engesser, S. (2021). Voice-based agents as personified things: Assimilation and accommodation as equilibration of doubt. *Human-Machine Communication, 2*, 57–79. https://doi.org/10.30658/hmc.2.3

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Gambino, A., Fox, J., & Ratan, R. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication, 1*, 71–86. https://doi.org/10.30658/hmc.1.5

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass, 10*(10), 535–549. https://doi.org/10.1111/spc3.12267

de Graaf, M. M. A., & Allouch, S. B. (2017). The influence of prior expectations of a robot's lifelikeness on users' intentions to treat a zoomorphic robot as a companion. *International Journal of Social Robotics, 9*(1), 17–32. https://doi.org/10.1007/s12369-016-0340-4

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

Grundke, A., Stein, J.-P., & Appel, M. (2022). Mind-reading machines: Distinct user responses to thought-detecting and emotion-detecting robots. *Technology, Mind, and Behavior, 3*(1). https://doi.org/10.1037/tmb0000053

Gunkel, D. J. (2012). *The machine question: Critical perspectives on AI, robots, and ethics*. MIT Press.
Gunkel, D. J. (2018). *Robot rights*. MIT Press.
Guzman, A. L. (2019). Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior, 90*, 343–350. https://doi.org/10.1016/j.chb.2018.08.009
Guzman, A. (2020). Ontological boundaries between humans and computers and the implications for human-machine communication. *Human-Machine Communication, 1*, 37–54. https://doi.org/10.30658/hmc.1.3
Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Pantheon Books.
Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10*(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4
Haslam, N., Kashima, Y., Loughnan, S., Shi, J., & Suitner, C. (2008). Subhuman, inhuman, and superhuman: Contrasting humans with nonhumans in three cultures. *Social Cognition, 26*(2), 248–258. https://doi.org/10.1521/soco.2008.26.2.248
Heidegger, M. (1927/1962). *Being and time*. S.C.M. Press.
Hoffmann, L., Bock, N., Rosenthal, v.d, & Pütten, A. M. (2018). The peculiarities of robot embodiment (EmCorp-Scale): Development, validation and initial test of the embodiment and corporeality of artificial agents scale. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction* (pp. 370–378). https://doi.org/10.1145/3171221.3171242
Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. Wiley.
Kahn, P. H., Reichert, A. L., Gary, H. E., Kanda, T., Ishiguro, H., Shen, S., et al. (2011). The new ontological category hypothesis in human-robot interaction. In *Proceedings of the 6th international conference on human-robot interaction* (p. 159). https://doi.org/10.1145/1957656.1957710
Kahn, P. H., & Shen, S. (2017). NOC NOC, who's there? A new ontological category (NOC) for social robots. In N. Budwig, E. Turiel, & P. D. Zelazo (Eds.), *New perspectives on human development* (pp. 106–122). Cambridge University Press. https://doi.org/10.1017/CBO9781316282755.008.
Keijsers, M., & Bartneck, C. (2018). Mindless robots get bullied. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction* (pp. 205–214). https://doi.org/10.1145/3171221.3171266
Keijsers, M., Bartneck, C., & Eyssel, F. (2022). Pay them No mind: The influence of implicit and explicit robot mind perception on the right to be protected. *International Journal of Social Robotics, 14, 499–514*. https://doi.org/10.1007/s12369-021-00799-1
Kertész, C., & Turunen, M. (2019). Exploratory analysis of Sony AIBO users. *AI & Society, 34*(3), 625–638. https://doi.org/10.1007/s00146-018-0818-8
Klamer, T., Ben Allouch, S., & Heylen, D. (2011). "Adventures of Harvey" – use, acceptance of and relationship building with a social robot in a domestic environment. In M. H. Lamers, & F. J. Verbeek (Eds.), *Human-robot personal relationships* (pp. 74–82). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-19385-9_10.
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
Levin, D. T., Killingsworth, S. S., Saylor, M. M., Gordon, S. M., & Kawamura, K. (2013). Tests of concepts about different kinds of minds: Predictions about the behavior of computers, robots, and people. *Human-Computer Interaction, 28*(2), 161–191. https://doi.org/10.1080/07370024.2012.697007
Malle, B. F. (2019). How many dimensions of mind perception really Are there? In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st annual meeting of the cognitive science society* (pp. 2268–2274). Cognitive Science Society.
Malle, B. F., & Thapa Magar, S. (2017). What kind of mind do I want in my robot? Developing a measure of desired mental capacities in social robots. In *Proceedings of the companion of the 2017 ACM/IEEE international conference on human-robot interaction* (pp. 195–196). https://doi.org/10.1145/3029798.3038378
Marchesi, S., De Tommaso, D., Perez-Osorio, J., & Wykowska, A. (2022). Belief in sharing the same phenomenological experience increases the likelihood of adopting the intentional stance toward a humanoid robot. *Technology, Mind, and Behavior, 3*(3), 11. https://doi.org/10.1037/tmb0000072
Marchesi, S., Ghiglino, D., Ciardo, F., Perez-Osorio, J., Baykara, E., & Wykowska, A. (2019). Do we adopt the intentional stance toward humanoid robots? *Frontiers in Psychology, 10*, 450. https://doi.org/10.3389/fpsyg.2019.00450
Marchesi, S., Spatola, N., Perez-Osorio, J., & Wykowska, A. (2021). Human vs humanoid. A behavioral investigation of the individual tendency to adopt the intentional stance. In *Proceedings of the 2021 ACM/IEEE international conference on human-robot interaction* (pp. 332–340). https://doi.org/10.1145/3434073.3444663
Müller, B. C. N., Gao, X., Nijssen, S. R. R., & Damen, T. G. E. (2021). I, robot: How human appearance and mind attribution relate to the perceived danger of robots. *International Journal of Social Robotics, 13*(4), 691–701. https://doi.org/10.1007/s12369-020-00663-8
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
Nosek, B. A. (2007). Implicit–explicit relations. *Current Directions in Psychological Science, 16*(2), 65–69. https://doi.org/10.1111/j.1467-8721.2007.00477.x
Phillips, E., Zhao, X., Ullman, D., & Malle, B. F. (2018). What is humanlike? Decomposing robots' humanlike appearance using the anthropomorphic roBOT (ABOT) database. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction* (pp. 105–113). https://doi.org/10.1145/3171221.3171268
Pradhan, A., Findlater, L., & Lazar, A. (2019). "Phantom friend" or "just a box with information": Personification and ontological categorization of smart speaker-based voice assistants by older adults. *Proceedings of the ACM on Human-Computer Interaction, 3*, 1–21. https://doi.org/10.1145/3359316
Putman, H. (1964). Robots: Machines or artificially created life? *The Journal of Philosophy, 61*(21), 668. https://doi.org/10.2307/2023045
Ramanathan, M., Mishra, N., & Thalmann, N. M. (2019). Nadine humanoid social robotics platform. In M. Gavrilova, J. Chang, N. M. Thalmann, E. Hitzer, & H. Ishikawa (Eds.), *Advances in computer graphics* (Vol. 11542, pp. 490–496). Springer International Publishing. https://doi.org/10.1007/978-3-030-22514-8_49.
Roesler, E., Manzey, D., & Onnasch, L. (2023). Embodiment matters in social HRI research: Effectiveness of anthropomorphism on subjective and objective outcomes. *ACM Transactions on Human-Robot Interaction, 12*(1), 1–9. https://doi.org/10.1145/3555812
Seaborn, K., & Frank, A. (2022). What pronouns for pepper? A critical review of gender/ing in research. In *CHI conference on human factors in computing systems* (pp. 1–15). https://doi.org/10.1145/3491102.3501996
Seibt, J. (2017). Towards an ontology of simulated social interaction: Varieties of the "as if" for robots and humans. In R. Hakli, & J. Seibt (Eds.), *Sociality and normativity for robots* (pp. 11–39). Springer International Publishing. https://doi.org/10.1007/978-3-319-53133-5_2.
Severson, R. L., & Carlson, S. M. (2010). Behaving as or behaving as if? Children's conceptions of personified robots and the emergence of a new ontological category. *Neural Networks, 23*(8–9), 1099–1103. https://doi.org/10.1016/j.neunet.2010.08.014
Simmler, M., & Markwalder, N. (2019). Guilty robots? Rethinking the nature of culpability and legal personhood in an age of artificial intelligence. *Criminal Law Forum, 30*(1), 1–31. https://doi.org/10.1007/s10609-018-9360-0
Spatola, N., Marchesi, S., & Wykowska, A. (2022). Different models of anthropomorphism across cultures and ontological limits in current frameworks the integrative framework of anthropomorphism. *Frontiers in Robotics and AI, 9*, Article 863319. https://doi.org/10.3389/frobt.2022.863319
Spatola, N., & Wudarczyk, O. A. (2021). Ascribing emotions to robots: Explicit and implicit attribution of emotions and perceived robot anthropomorphism. *Computers in Human Behavior, 124*, Article 106934. https://doi.org/10.1016/j.chb.2021.106934
Stein, J.-P., Liebold, B., & Ohler, P. (2019). Stay back, clever thing! Linking situational control and human uniqueness concerns to the aversion against autonomous technology. *Computers in Human Behavior, 95*, 73–82. https://doi.org/10.1016/j.chb.2019.01.021
Stein, J.-P., & Ohler, P. (2017). Venturing into the uncanny valley of mind—the influence of mind attribution on the acceptance of humanlike characters in a virtual reality setting. *Cognition, 160*, 43–50. https://doi.org/10.1016/j.cognition.2016.12.010
Takahashi, H., Ban, M., & Asada, M. (2016). Semantic differential scale method can reveal multidimensional aspects of mind perception. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.01717
Thellman, S., de Graaf, M., & Ziemke, T. (2022). Mental state attribution to robots: A systematic review of conceptions, methods, and findings. *ACM Transactions on Human-Robot Interaction, 11*(4), 1–51. https://doi.org/10.1145/3526112
Thellman, S., Silvervarg, A., & Ziemke, T. (2017). Folk-psychological interpretation of human vs. Humanoid robot behavior: Exploring the intentional stance toward robots. *Frontiers in Psychology, 8*, 1962. https://doi.org/10.3389/fpsyg.2017.01962
Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology, 63*(3), 107–140. https://doi.org/10.1016/j.cogpsych.2011.06.001
Tzelios, K., Williams, L. A., Omerod, J., & Bliss-Moreau, E. (2022). Evidence of the unidimensional structure of mind perception. *Scientific Reports, 12*, Article 18978. https://doi.org/10.1038/s41598-022-23047-6
Van Der Goot, M., & Etzrodt, K. (2023). Disentangling two fundamental paradigms in human-machine communication research: Media equation and media evocation. *Human-Machine Communication, 6*, 17–30. https://doi.org/10.30658/hmc.6.2
Ward, A. F., Olsen, A. S., & Wegner, D. M. (2013). The harm-made mind: Observing victimization augments attribution of minds to vegetative patients, robots, and the dead. *Psychological Science, 24*(8), 1437–1445. https://doi.org/10.1177/0956797612472343
Wegner, D. M., & Gray, K. (2016). *The mind club: Who thinks, what feels, and why it matters. Viking*.
Weisman, K., Dweck, C. S., & Markman, E. M. (2017). Rethinking people's conceptions of mental life. *Proceedings of the National Academy of Sciences, 114*(43), 11374–11379. https://doi.org/10.1073/pnas.1704347114
Weisman, K., Legare, C. H., Smith, R. E., Dzokoto, V. A., Aulino, F., Ng, E., et al. (2021). Similarities and differences in concepts of mental life among adults and children in five cultures. *Nature Human Behaviour, 5*(10), 1358–1368. https://doi.org/10.1038/s41562-021-01184-8
Wykowska, A. (2021). Robots as mirrors of the human mind. *Current Directions in Psychological Science, 30*(1), 34–40. https://doi.org/10.1177/0963721420978609
Złotowski, J., Sumioka, H., Eyssel, F., Nishio, S., Bartneck, C., & Ishiguro, H. (2018). Model of dual anthropomorphism: The relationship between the media equation effect and implicit anthropomorphism. *International Journal of Social Robotics, 10*(5), 701–714. https://doi.org/10.1007/s12369-018-0476-5
Złotowski, J., Yogeeswaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies, 100*, 48–54. https://doi.org/10.1016/j.ijhcs.2016.12.008