Evaluating Theories Guiding Human-Robot Interaction with Implications for Ethics

Jessica K. Barfield and Jiangen He

Abstract. This chapter reviews the role of theory in guiding research on human-robot interaction (HRI) with a discussion of how the theories relate to the design and use of robots from the perspective of ethics. As discussed in the chapter, studies on HRI are often based on a theoretical framework derived from the disciplines of psychology, communication, information science, robotics, and marketing. Each of these disciplines have either directly developed theory to guide research on HRI or in a different context have developed theory to study human behavior which was then adopted for use in robotics by the HRI community. After a discussion of the benefits of using theory to guide research on HRI a conceptual framework is proposed to categorize theories which have been previously used to explore different aspects of HRI. In addition, summary tables are provided which presents key aspects of HRI theories and with implications for ethics. The chapter concludes with a critical review of how to review theories and a discussion of future directions in the use of theory to guide HRI research.

Keywords: Theory, Human-Robot Interaction, Anthropomorphism, Uncanny Valley Effect, AI, Design

1. Introduction

This chapter summarizes theories which have been used to evaluate human interaction with robots experienced in different social contexts. Throughout the chapter an attempt is made to relate the theories to issues of ethics for the design and use of robots. As the chapter discussion and summary tables show many theories guiding research on human-robot interaction (HRI) were developed primarily within the academic disciplines of communication, psychology, robotics, and marketing **(**Asemi, Ko, and Nowkarizi, 2021; Manthiou et al. 2021; Prewett et al. 2010; Ullrich and Diefenbach, 2017; Walther, 1992). Robotics researchers either used prior theory to guide research on HRI or they developed new theory to describe broad principles of human behavior with applications to HRI. In either case, numerous theoretical approaches have been used to guide HRI research offering the robotics community different theoretical assumptions and predictions for HRI. Concerning the scope of the chapter it presents a review of HRI theories by summarizing two main theoretical approaches for investigating HRI and includes a discussion of how selected theories relate to issues of ethics in the design and use of robots when experienced in social contexts. The chapter concludes with a critical evaluation of HRI theories and comments on future directions in the use of theory when investigating issues of ethics for human interaction with robots.

The term, "theory-based approach" (Auttin et al. 2023; Hmelo, Gotterer, and Bransford, 1997) is a common term used among researchers and for this chapter has applications to ethics offering several distinct advantages to the robotics community (Hoorn, 2020a,b). More specifically, by using the term theory-based approach, we refer to HRI studies that that were guided by the assumptions and predictions of a theory (thus, the approach is theory-based). For example, the use of theory allows researchers to propose hypothesis which can be tested within a particular theoretical framework thus providing a context in which to make predictions for human interaction with robots (Barfield, 2023a; Trafton, Raymond, and Khemlani, 2021). Further, a theory-based approach to guide HRI research is useful for connecting researchers from different academic disciplines interested in the same or a similar topic within HRI thus providing the opportunity for interdisciplinary collaboration between researchers. Additionally, there are other benefits associated with the use of theory to guide ethics considerations in HRI research. For example, Teo et al. (2017) describing the value of theories for human interaction with robots commented that theories can help organize the body of knowledge in a particular field, provide direction for HRI research, and importantly, contribute to additional theory building. And from an applied perspective, a theory-based approach for HRI

may help researchers focus on identifying the key performance indicators which may guide ethical interactions between humans and robots and in fostering the development of robots that interact with people based on the ethics principles of transparency, repeatability, reproducibility, and trust (Marvel, et al. 2020). In addition, discussing the benefits of using theory to guide research in HRI, Teh and colleagues (2021) concluded that theory could be used to provide first-cut insights into large-scale HRI scenarios that otherwise could be too costly or challenging to perform with computer simulations or with observing robots in the wild.

Some additional background on what constitutes a theory and its elements is useful to set the stage for the chapter. However, we should note early in the chapter that there is no one definition of what a theory is within or between academic disciplines, and this has historically been a topic of discussion within the philosophy of science literature (Cover, Curd, and Pincock, 2012; Popper, 2014). For example, Karl Popper offered unique perspectives on the nature of scientific inquiry and advocated for the idea of falsificationism in which scientific theories cannot be proven true, but they can be falsified through empirical testing (Popper, 2014). The current chapter adopts Popper's approach for the conceptualization of a theory and as a useful guide for selecting the theories in the summary tables presented below. From Popper, the theories presented in this chapter should meet the criteria that they can be subjected to rigorous testing and scrutiny, and they should be open to revision or rejection based on empirical evidence. Popper's emphasis on falsifiability is particularly relevant for HRI research and robot ethical design where hypotheses can be posed and directly tested in the laboratory or in a field-study evaluating robots in the wild. Thus, a theory usually includes some predictive capacity and attempts to explain the causal mechanisms of implementing the theory in a particular context (Hantula, 2011; Kwon et al, 2020; Shalley, 2012). To provide more perspective on what constitutes a theory according to Nilsen (2015) a theory consists of "analytical principles or statements which structure our observations, understanding, and explanation of the world" (pg. 2).

More on focus for this chapter, "ethical theory" refers to the study of morality, exploring what actions are considered right or wrong, and provides frameworks for making moral decisions. In addition, building ethical robots involves the process of design, and Jones and Gregor (2007) discussed components of design theories which included: (1) the purpose and scope of the theory, (2) its constructs, (3) principles of form and function defining the theory, (4) using testable propositions, that is, using a theory-based approach for design, and (5) considering principles of implementation. Additionally, in the field of design, Friedman (2003) commented that the process of design involved solving problems noting that theories need to be developed which focus on exploring how designs work using a theory-based explanation of the design process. In another discipline, Love (2000) discussing the philosophy of design commented that a critical analysis of design should clarify the relationships between individual design concepts including the underlying assumptions comprising the design. In another discipline the philosophy of design explores the fundamental questions surrounding design's purpose, ethics, and implications for society. Guided by the philosophy of design, the process of design is examined critically, often in conversation with aesthetics, ethics, and knowledge. Thus, the philosophy of design investigates the foundational concepts, assumptions, and implications of design itself. It examines design in relation to fundamental philosophical questions like ethics (how design impacts society), aesthetics (what makes a design good or beautiful), and epistemology (how we gain knowledge through design). Additionally, Weng and Hirata (2022) indicated that designers of socially intelligent machines should consider Value-Sensitive Design which is a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner. And in the field of engineering, design theory adopts the engineering design process which includes a series of steps that involve problem definition, brainstorming, prototyping, and testing with the application of fundamental engineering principles. Further, engineering design incorporates design theories in order to create safe, efficient, and reliable solutions to real-world problems.

For research on HRI, it is often based on the assumption that rules and theories that apply to interpersonal interactions between people should also apply to human interaction with social robots. This observation is particularly relevant for the role of ethics in HRI. Essentially, that people interact with robots

as they do people is the Computers as Social Actors (CASA) paradigm proposed by Nass and Brave (2005) and Nass, Steuer, and Tauber (1994). Supporting this notion, Fox and Gambino (2021) commented that social robots should mimic humans in form and behavior such that human interaction with robots is more likely to resemble human-human interaction. They discussed several theories which could be used to investigate HRI which included resource theory (Martini, Buzzell, and Wiese, 2015), interdependence theory (Wagner and Arkin, 2008), equity theory (Fox and Gambino, 2021), and social penetration theory (Fox and Gambino, 2021). Critiquing the theories, Fox and Gambino considered whether they were viable frameworks for studying HRI given their theoretical assumptions and claims. With relevance for ethics, they concluded that the above listed theories on interpersonal behavior might be unsuitable for examining current human-robot interactions given their view that current robots often fail to achieve social actor status and thus may not be considered co-actors in a social exchange (Fox and Gambino, 2021). However, most researchers within the HRI community have concluded otherwise. For example, numerous researchers have successfully used communication, management science, and behavioral science theories to guide research on HRI and to investigate issues on ethics related to HRI (Barfield, 2023b). As an example, Teh, Hu, and Soh (2021) used a theory-based approach to explore how humans and robots might adapt to one another over time. To investigate this topic, they used the Theoretical Human-Robot Scenarios (THuS) framework to evaluate the interactions between groups of humans with robots that were programmed with the ability to learn. More specifically, Teh and colleagues successfully used ThuS as a tool to quantitatively compare HRI scenarios that involved different agent types.

Not surprisingly, the field of psychology has provided the robotics community a rich set of theories which have been used to guide research with relevance to ethics and HRI (Ullrich and Diefenbach, 2017). For example, research on moral psychology has investigated HRI in moral contexts, and has asked how these results may impact debate in ethical theory. Interestingly, "HCI ethics" typically involves applying moral standards to the design, use, and management of technology, focusing on human welfare, privacy, ownership, bias, usability, trust, autonomy, and accountability. An example of this approach is seen through the use of the Semiotic Inspection Method as applied to HCI and HRI (de Souza, et al. 2006) which is based broadly on Semiotic Engineering theory (which uses concepts from semiotics and computer science to investigate the relationship between user and designer) (de Souza, et al. 2001). Bento and colleagues (2009) argued that the Semiotic Inspection Method is an example of an approach that has been successfully used in HCI research and adopted for the design of human-robot interfaces. Additionally, Gonsior et al. (2012) discussed the transfer of psychological knowledge, particularly on prosocial behavior from social psychology to the domain of HRI. More specifically, they investigated human helpfulness behavior towards a robot and found that situations evoking empathy and similarity between human and robot led to increased helpfulness towards the robot. Additionally, theories used to explain human-computer interaction (HCI) have often been adopted by the robotics community and widely used to guide research on HRI (Allan, et al., 2022 a, b; Higgins, 1997, 2005). Given the numerous benefits of using theory to guide research on HRI this chapter extends the discussion on the use of theory for evaluating HRI by presenting the results of a literature search broadly relating HRI theories to ethics and by presenting a summary table listing representative theories which have been used previously by scholars in different disciplines to guide research on HRI with implications for ethics.

1.1 Searching for HRI Theories

The search for HRI theories with implications for ethics was guided by the PRISMA statement and its extension (PRISMA, 2024; see also Hirvonen et al. 2024) which describe a systematic approach and reporting procedure for literature reviews. The approach consists of: (i) identifying candidate articles to review, (ii) a screening process applied to the articles found, and (iii) the use of criteria to select the articles used to summarize the search. Based on this procedure, and following Hirvonen et al. (2024), the following process was used to search for theories guiding HRI research.

1.1.1 Identifying Articles for Review

The first phase of the literature review consisted of identifying theories relating to the topic of HRI which could then be analyzed from the perspective of ethics and HRI. The process was based primarily on the use of the Web of Science, IEEE/ACME, and Google Scholar databases and a search for peer reviewed articles or papers from conferences proceedings focusing on the topic of HRI. Further, an effort was made to locate IEEE and ACM conference papers relating to HRI and to identify and search relevant journals on HRI.

In the first phase of the literature review the search parameters used to locate theories which guided HRI research were initially broad and returned thousands of records (see Table 1). For example, at the time of the search, using high-level search terms on HRI shows that 15,276 records were produced using the Web of Science database and 123,783 records were returned using Google Scholar. Thus, the search was refined such that it was more focused on articles which used a theoretical approach to investigate HRI and other criteria discussed below. This was done primarily by narrowing the terms used in the search query. For example, using the Web of Science database and using "robot anthropomorphism" as a search topic returned 996 articles. However, when the search was narrowed even further using the following search terms "robot anthropomorphism AND experiment" 205 records were returned. Further, refinement of the search queries produced less returned articles in a search, but more on point for the topic of the chapter. For example, "robot anthropomorphism AND experiment AND theory" produced 2 records and both were found to be relevant for the chapter topic. Additionally, using the Web of Science and Google Scholar databases and `using the above refinement procedure terms such as "gender", "race", and "ethnicity", given they may relate to issues of ethics in the design and use of robots, were paired to the terms "robot" or "human-robot interaction" and used to guide the search. Another search method was to search using the names of prominent researchers in HRI which were identified by looking at paper citations for relevant articles and chaining respective articles' references in key areas of HRI research.

1.1.2 Screening Process

After the above search procedure which initially resulted in several thousand returned records, in the second phase of the search process for HRI theories the articles were screened for relevance based on inclusion and exclusion criterion. Additionally, for searches that produced several hundred articles their abstracts were first reviewed to check whether they used a theory-based approach to investigate HRI and if so, the article was downloaded for further screening. For all articles and paper abstracts, screening occurred by the first author reviewing the records found. For inclusion, the theory had to relate to human interaction with robots, could represent any application area, and the theory stated assumptions that could be tested empirically through the use of hypothesis testing (see Popper, 2014); in either case a quantitative or qualitative study was deemed appropriate for the literature review.

Based on the above approach for the review, several criteria were used to select the articles which were placed in summary tables (presented below) organized around two conceptual themes that have prominently guided HRI research. As the main criterion the article had to include a human and robot engaged in a social context. Further, the article had to identify a theory and make a theoretical contribution to HRI and have implications for human-robot interface design and ethics. Table 1 provides examples of keywords and databases used to guide the search and Table 2 presents some key researchers doing theory-guided HRI research.

Table 1. Example of literature search queries, prompts, and search parameters used across online databases with key articles found.

Search Method or Database	**Example of Keywords Used**	**Key Articles Found**
University Online Databases: Web of Science	"robot AND ethnicity" and "robot ethnicity" 49 records each "human-robot interaction AND ethnicity" 12 records "culture AND robot" 1,059 records "human-robot interaction AND theory" 869 records "uncanny valley" 714 records "robots AND uncanny valley" 351 records "humanistic theory AND robot" 6 records "communication theory AND robot" 1,979 records "social role AND robot" 1,886 records "robot anthropomorphism" 996 records "robot anthropomorphism and communication theory" 16 records "robot anthropomorphism and experiment" 205 records	Fox, J., Gambino, A., Relationship Development with Humanoid Social Robots: Applying Interpersonal Theories to Human/Robot Interaction*, Cyberpsychology Behavior and Social Networking,* Vol. 24(5), 294-299, 2021. Bernotat, J., Eyssel, F., and Sachse, J., Shape It - The Influence of Robot Body Shape on Gender Perception in Robots, *9th International Conference on Social Robotics (ICSR)*, 75-84, 2017. Otterbacher, J and Talias, M., S/he's too Warm/Agentic! The Influence of Gender on Uncanny Reactions to Robots, *12th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 214-223, 2017. Eyssel, F., and Hegel, F., (S)he's Got the Look: Gender Stereotyping of Robots, *Journal of Applied Social Psychology*, Vol. 42(9), 2213-2230, 2012. Sparrow, R., Do Robots have Race? Race, Social Construction, and HRI, *IEEE Robotics and Automation Magazine*, Vol. 27(3), 144-150, 2020. For anthropomorphism these search terms provide an example of the refinement search process used to produce a more focused search]
Google Scholar	"robot anthropomorphism" 39,300 records "robot anthropomorphism and communication theory and experiment" 3,160 records "social construction theory AND robot" 23 records	Tajfel, H., Turner, J. C., Austin, W. G., and Worchel, S., An Integrative Theory of Intergroup Conflict, *Organizational identity: A Reader*, 56-65, 1979. Duffy, B. R., Anthropomorphism and the Social Robot, *Robotics and Autonomous Systems*, Vol. 42 (3-4), 177-190, 2003.

Table 2. Key HRI researchers identified.

Key researchers identified from conferences participation and literature review	Robert Sparrow; Fredirike Eyssel; Christoph Bartneck; Kerston Haring	Eyssel, F., and Loughnan, S., It Don't Matter if You're Black or White? Effects of Robot Appearance and User Prejudice on Evaluations of a Newly Developed Robot Companion, 422-433, In: Herrmann G., Pearson M. J., Lenz A., Bremner, P., Spiers A., & Leonards U. (eds) Social Robotics, ICSR 2013, *Lecture Notes in Computer Science*, Vol. 8239. Springer, Cham, 2013. Sparrow, R., Robotics Has a Race Problem, *Science Technology & Human Values, Vol. 45* (3), 538-560, 2020. Bartneck, C., Nomura, T., Kanda, T., Suzuki, T., and Kennsuke, K., Cultural Differences in Attitudes Towards Robots. *Proceedings of the AISB Symposium on Robot Companions: Hard Problems And Open Challenges In Human-Robot Interaction*, Hatfield, 1-4, 2005. Bartneck, C., Yogeeswaran, K., Ser, Q-M., Woodward, G., Sparrow, R., Wang, S., Eyssel. F., Robots and Racism, *Proceedings of ACM/IEEE International Conference on Human Robot Interaction (HRI '18)*, 1-9, 2018. Eyssel, F., and Kuchenbrandt, D., Social Categorization of Social Robot: Anthropomorphism as a Function of Robot Group Membership, *British Journal of Social Psychology*, Vol. 51 (4), 724-731, 2012a.

1.2 Organization of the Theories

In order to provide a conceptual framework to organize the theories discovered through the search procedures, Table 3 presents a description of two broad conceptual areas which were used to categorize and group individual theories which have been used previously to investigate HRI. Based on the above search procedure, the framework emerged from a comprehensive review of the HRI literature and prior HRI research by the Barfield (2021 a,b, 2023a,b,c). While there is some overlap between the two conceptual areas, the categories reflect a broad framework around which to organize HRI theories and are useful in identifying the general characteristics of theories evaluating human performance and ethical interactions with robots. To start, the process of anthropomorphism for robots is a well-replicated finding within HRI research and several theories have been proposed to describe the processes leading to robot anthropomorphism (Jones, Niichel, and Armstrong, 2018; Kühne and Peter, 2023). Given the important finding that people evaluate robots in a similar manner as they do humans it is expected that more research and theory development will be done to explore the factors which influence how robots are anthropomorphized based on different characteristics of the robot. In addition, the Uncanny Valley effect first discussed by Mori (1970) which describes the negative or eerie reaction among humans for robots approaching, but not quite reaching, human levels of appearance has been broadly used to study HRI and has served as framework to develop theories to guide HRI research. The Uncanny Valley effect suggests a non-linear relationship between a robot's anthropomorphism and reaction to the robot and proposes that by increasing the human likeness of robot appearance such that it is no longer perceived to be within the dip of the Uncanny Valley, there will be an increased affinity towards the robot (Seymour et al. 2019).

Table 3. Two conceptual categories for conceptualizing HRI theories.

.

Theory Categorization	Description	Relevance for Ethics and HRI	Example
Theories based on the process of robot anthropomorphism	Theories describing HRI may be based on the perceptual and cognitive processes associated with interpreting the features and behavior of a robot in human terms. Further, anthropomorphizing robots has been shown to lead to an emotional response from users; therefore, among others, theories dealing with human affect have been used to evaluate the process of robot anthropomorphism.	The process of anthropomorphism allows relationships to be formed between humans and robots. Humanoid robots especially have the power to convince their user that they are able to provide them with genuine, reciprocal affection and real social relations. Anthropomorphism can distort moral judgments about AI, such as those concerning its moral character and status, as well as judgments of responsibility and trust.	Fraune (2020) investigated whether robot anthropomorphism decreased differences between humans and robots on in-group favoritism. Generally, participants favored the in-group over the out-group robots in various scenarios.
Theories based on the Uncanny Valley effect	The Uncanny Valley, the eerie feeling elicited when humans evaluate robots that approach but not reaching humans in likeness, has been replicated in numerous studies; such studies often vary the physical appearance and behavior of the robot and its voice characteristics.	The Uncanny Valley effect refers to the unsettling feeling evoked by humanoid robots that appear almost, but not quite, human, and its ethical implications in robot design and interaction are significant, impacting moral expectations, decision-making, and potential misuse. In moral psychology there are several outstanding questions on how robot appearance and other perceived properties of the robots influences the way their decisions are evaluated.	Thepsoonthorn, Ogawa, and Miyake (2021) examined the influence of the robot's nonverbal behavior on the Uncanny Valley effect. They observed a relationship between the participants' ratings on human-likeness of the robot's nonverbal behavior and affinity toward the robot's nonverbal behavior.

1.3 Comments on Theories for HRI

To summarize the above discussion, Figure 1 suggests that Robot Anthropomorphism and the Uncanny Valley effect can serve as broad conceptual frameworks to guide HRI research and specifically to frame issues of ethics, and as such, they have both led to the use, development, and adoption of other theories from different disciplines which have been used to study HRI. For example, from Figure 1 this includes the use of theories accounting for social roles adopted by robots in social interactions, theories related to how people categorize robots, humanistic theories which focus on how robots are experienced by users in social contexts, and theories from psychology, communications, and information science which focus on the communication process between human and robot. In addition, the other topics shown in Figure 1 have also played an important role in guiding HRI research; for example, people may place robots in social categories, or react to robots as if they had a race, gender, or ethnicity which of course implicates issues of ethic in HRI. In these areas most often HRI researchers have adopted theories from other fields to guide their research. There is also a relationship between robot anthropomorphism and the Uncanny Valley effect in

that robots perceived to be in the dip of the Uncanny Valley are attributed fewer human characteristics than robots perceived to be outside of the Uncanny Valley. On this point, Kühne and Peter (2023) commented that anthropomorphism plays a prominent role in HRI stating that robot shape, which is a factor in determining whether a robot is perceived to be in the Uncanny Valley, is a potential precursor of anthropomorphism.

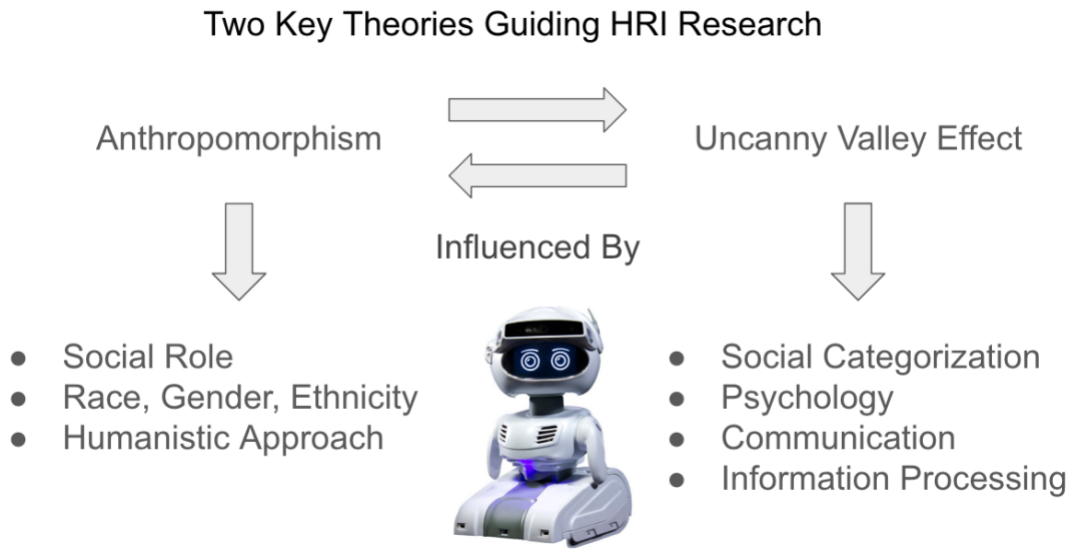

Figure 1. Robot Anthropomorphism and the Uncanny Valley effect can be used as conceptual frameworks to guide HRI research. Further, their broad conceptualizations have influenced different theoretical approaches for HRI derived from theories on social role, robot categorization, humanistic, psychology, communication, and information processing approaches.

2. Summary of Theories for Ethical HRI Design

Based on the above two classifications schemes which can be used to categorize theories that have guided HRI research, and on a comprehensive search of the HRI literature over a period of months, Tables 3 and 4 list different theories that were identified meeting the above search criteria and which have previously been used to guide HRI research. From the tables it is interesting to note that most theories identified were adopted from the fields of psychology, information and communication sciences, and to some extent, marketing which is not surprising given a main goal within these disciplines is to understand human behavior in a particular context. To reiterate, what follows next are tables which organizes theories which have previously been used to guide research on the topic of HRI based on the Uncanny Valley Effect and robot anthropomorphism. The tables provide a brief description or explanation of the theory, its relation to HRI and to ethics, and a standard reference in which the theory was used to explore HRI. Finally, we should note that there are other theories which could be included in the following summary tables; for brevity, what are listed are representative theories for each conceptual framework.

Table 3. The Uncanny Valley Effect used to guide HRI research.

Mail Theory Uncanny Valley Effect	**Description or Explanation**	**Relation to HRI and Ethics**	**Reference**
Uncanny Valley Effect	The uncanny valley reflects the relation between a robot's degree of resemblance to a human being and the emotional response to the robot. The concept suggests that humanoid robots that imperfectly resemble actual human beings provoke feelings of eeriness and revulsion in observers	Robots that appear human but not quite so have been found to result in negative evaluations of the robot. This implicates moral and ethical issues in robot design.	Mori (1970)
Terror Management Theory (TMT)	TMT proposes that humans may feel threatened by humanoid appearing robots	Humanoid robots may violate norms of human appearance and movement thus influencing human moral and ethical behavior towards robots.	Gray and Wegner (2012)
Appraisal Theory	Appraisal theory is derived from psychology research which states that emotions are extracted from our evaluations of events that cause specific reactions in different people. Our appraisal of a situation can cause an emotional or affective response	For robots varying levels of human-likeness of attributes (i.e., visual, vocal and verbal) influence consumption outcomes (e.g., service encounter evaluation, revisit intentions and positive word of mouth intentions). By appraising a situation an emotional response may be generated and can lead to the ethical or unethical treatment of robots.	Lu, Zhang, and Zhang (2021) Demutti et al (2022)
Theory of Mind Perception	Mind perception entails ascribing mental capacities to other entities	Mind perception is linked to individual differences and to design features of a robot which may affect the evaluation of its mental capacities and may help to explain the Uncanny Valley effect. The mental attributes ascribed to robots may influence its treatment as ethical or unethical.	Schein and Gray (2015) Rabe et al. (2022)

The next summary table emphasizes that anthropomorphism, ascribing human characteristics to robots has been a lead to numerous theories to guide HRI research and has implications for ethnics in robot design and use.

Table 4. The construct of anthropomorphism used to guide HRI research and theory development.

Anthropomorphism	**Description or Explanation**	**Relation to HRI and Ethics**	**Reference**
Anthropomorphism Theory	Anthropomorphism is the extent to which people attribute human characteristics to a non-human entity	The human-like appearance of a robot has a direct effect on the robot's evaluation and may trigger moral, ethical, or unethical treatment of robots.	Epley, Waytz, and Johnn (2007)
Humanized-AI Social Interactivity Framework	This framework extends previous studies in the consumer behavior literature by offering an increased understanding of how individuals choose to interact with robots in domestic environments based on humanness and social interaction	Different levels of robot interaction can affect an individual's liking of robots and the moral and ethical treatment of robots.	Letheren, Jetten, Roberts, and Donovan (2021)
Emotion Theory .	Generally, theories of emotion present broad theoretical perspectives representing all major schools of thought on the study of the nature of emotion	There are different roles that emotion plays in decision-making, learning, communication, and social interactions; thus robots that collaborate with people should consider human affect so that robots are treated morally and ethically.	Breazeal (2004)
SEEK Theory of Anthropomorphism	SEEK states that anthropomorphism occurs subconsciously within moments of interacting with a non-human agent, here an individual's mind shifts, recognizing for example, that a service is not being delivered by a something but rather by someone	Anthropomorphism can lead to the impression that the robot will operate similarly to a human and if so, issues of ethics and morals are implicated.	Epley, Waytz, and John (2007)
Theory-Context-Characteristics-Methodology (TCCM)	A TCCM approach uses qualitative input from researchers who are active in HRI and helps identify where opportunities for further development and growth lie.	TCCM assists in the design of service robots by comparing service robots with humans, thus representing a major design oriented method to create ethical robots.	de Keyser and Kunz (2022)
Theory of Affective Bonding	In psychology an affectional bond is a type of attachment behavior one individual has for another, e.g., a caregiver for her or his child in which the two partners tend to remain in proximity to one another	People may experience a mediated or simulated interaction with a social robot such that the more emotionally aroused, the more likely to develop an affective relationship with a robot and engage in more ethical treatment of the robot.	Hoorn (2020 a,b)

3. Analysis of Theories for Robot Design: Issues of Ethics

In addition to the above summary of theories presented in Tables 3 and 4, we conducted an in-depth analysis of the articles citing the ten seminal papers on the Uncanny Valley effect and the Theory of

Anthropomorphism. This was done to assess the influence of these theories specifically on robot design- a central focus of this chapter. To perform this analysis we harvested 6,411 articles that cited the twelve articles listed in Tables 3 and 4 using the Dimensions Analytics API, which enables full-text search, data retrieval, and supports data analyses and visualizations. Given the large volume of articles resulting from the analysis, we employed a two-stage filtering process to pinpoint those articles with a primary focus on robot design. In the first stage of our process, we leveraged the GPT-4.1 language model via the OpenAI API. For each article, we concatenated its title and abstract and supplied them to the model, along with a detailed system prompt outlining clear inclusion and exclusion criteria for robot design studies. The system prompt was as follows:

> *You are an expert in robotics research. Your task is to analyze research articles and determine if they are specifically studies about "robot design".*
>
> *Robot design studies include but are not limited to:*
> *- Physical robot design*
> *- Robot morphology and embodiment*
> *- Robot appearance and aesthetics*
> *- Human-robot interaction design*
> *- Robot prototyping and development*
>
> *Robot design studies DO NOT include:*
> *- Pure software/algorithm development*
> *- Robot control systems (unless specifically about design implications)*
> *- Theoretical robotics without design aspects*
> *- Studies that only use robots as tools but don't design them*
> *- General AI or machine learning without robot design focus*
>
> *Respond with a JSON object containing:*
> *- "is_robot_design": boolean (true if this is a robot design study)*
> *- "confidence": float between 0 and 1 (how confident you are in this classification. Below 0.5 means you recommend to manually check the article)*
>
> *Response format: {"is_robot_design": boolean, "confidence": float}*

The model returned, for each article, a result which contained a boolean flag (*is_robot_design*) indicating whether the study focused on robot design, and a corresponding confidence score between 0 and 1. The confidence score provided an estimate of the model's certainty in its classification for each article. In the second stage of our filtering process, we automatically accepted the classification if the model's confidence was at least 0.7. Articles with lower confidence scores were flagged for further manual review. Through this two-stage screening process, we ultimately identified 974 articles that met our criteria of focused studies on robot design. Additionally, to understand the research landscape of the robot design studies that applied to the theories summarized in Table 3 and 4, we conducted computational topic analysis using the BERTopic framework (Grootendorst, 2022). BERTopic is a topic modeling approach that leverages transformer-based language models to generate document embeddings, which are then clustered to identify coherent topics. For this analysis, we used SPECTER2 (Singh et al., 2022), a model specifically pretrained on scientific literature, to generate document embeddings from each article's title and abstract. Document embeddings are numerical representations of the articles in a multi-dimensional space. These representations capture the semantic meaning of the text (titles and abstracts), so that articles with similar topics or content are located closer together in this space. Based on training with scientific literature data, SPECTER2 can provide embeddings that enable more accurate and meaningful clustering when analyzing

research articles. The document embeddings of the 974 articles generated by SPECTER2, were used for the topic modeling (BERTopic). The BERTopic modeling pipeline included standard steps to extract and label meaningful research themes, such as dimensionality reduction, clustering, and keyword selection. To generate interpretable topic labels, we employed GPT-4.1 within the BERTopic pipeline to analyze the titles and keywords of articles within each identified topic (see https://maartengr.github.io/BERTopic/getting_started/representation/llm.html#openai). All topic labels were required to be more specific than generic "robot design" to ensure meaningful differentiation between research areas. The system prompt for labeling is as follows:

> *I have a research topic from robot design studies that includes the following research papers with these titles:*
> *[TITLES]. Keywords: [KEYWORDS]*
>
> *Please provide one short topic label (as short as possible). All the papers are about robot design, so the topic should be more specific than "robot design".*
>
> *Return the result in JSON format as follows:*
> *{*
> *"label": "<short robot design topic label>",*
> *}*

Figure 2 shows a visual representation of the major research topics identified from the 974 articles focusing on robot design and their engagement with the key theories of anthropomorphism and the Uncanny Valley effect. The visualization was created by using DataMappPlot (see https://datamapplot.readthedocs.io/en/latest/demo.html). In the figure, each dot represents a paper on "robot design" and each cluster represents a distinct topic area, labeled according to the most salient theme derived from the contents of the corresponding articles. The spatial relationships between clusters reflect semantic similarities, with related topics appearing closer together. This map illustrates how different aspects of anthropomorphism and design theory are distributed across contemporary robot design research.

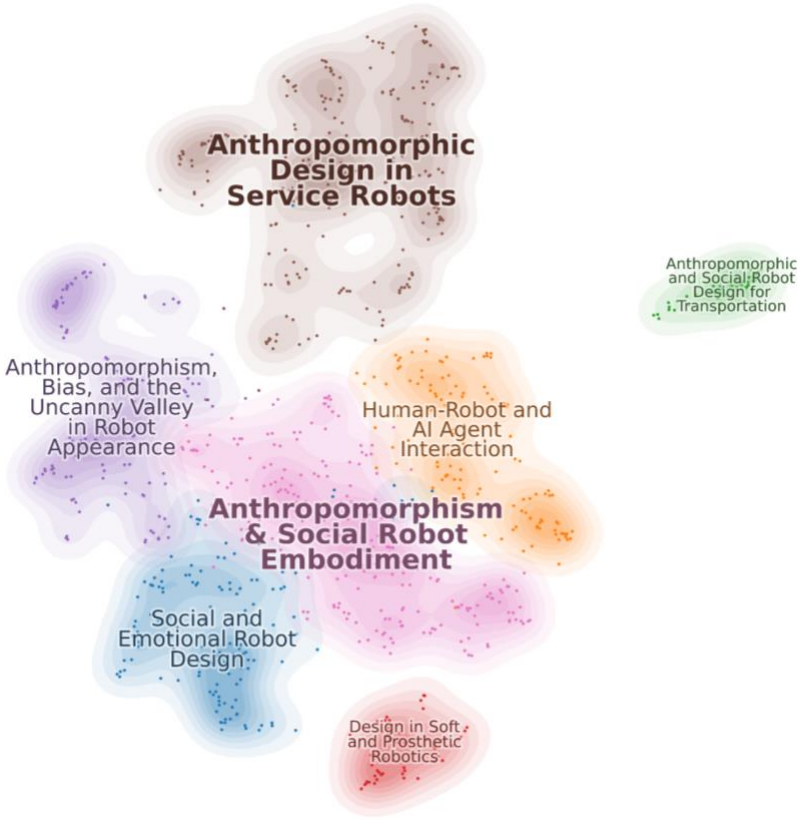

Figure 2. Topic map of robot design studies citing foundational theories on anthropomorphism and the Uncanny Valley effect. Each dot represents a study and colored cluster represents a research topic, with labels generated using GPT-4.1 for clarity and specificity. The distance between clusters indicates the semantic relatedness of the topics.

For each research topic in Figure 2, we randomly sampled 50 articles to identify ethical issues in robot design and develop ethical design guidelines to mitigate these issues. From this, we identified several common ethical concerns across different research topics in robot design and also analyzed the unique issues, guidelines, and recommendations specific to individual research topics or design contexts.

Concerning common ethical issues and recommendations for robot design a major recurring issue is the risk of reinforcing social stereotypes through robot appearance, voice, or behavior, especially regarding gender, race, and ability (Paterson, 2025; Perugia et al., 2023; Bernotat et al., 2021). Many studies warn that anthropomorphic cues can inadvertently perpetuate gender or racial biases, particularly when default designs for robots are white, able-bodied, or male-presenting (Perugia et al., 2022; Eyssel et al., 2012 a,b). From the literature it has been recommended that designers use gender-neutral or androgynous designs for robots experienced in social contexts and to involve diverse stakeholders in the design process to promote inclusivity and representation in HRI. From our analysis, privacy and data protection are also major concerns across design contexts, particularly when considering service and emotional robots. The collection and use of personal or physiological data (e.g., facial recognition, voice, EEG) can infringe on user privacy and autonomy, especially among vulnerable groups (Xie and Lei, 2022; Henkel et al., 2019; Wu et al., 2025). Ethical design should emphasize minimizing data collection to what is strictly necessary, ensuring informed user consent, and implementing robust data security and transparency measures (Li et al., 2022; Salvini et al., 2010).

Additionally, as noted in the chapter, the Uncanny Valley effect is another well-studied issue in HRI and plays a role in ethical interactions in HRI. Excessive or ambiguous anthropomorphism can reduce

trust, acceptance, and even influence the emotional well-being of users (Strait et al., 2017; Zhang et al., 2023). Many studies recommend moderate or context-appropriate anthropomorphism, congruence between appearance and behavior, and iterative user testing to avoid negative emotional responses (Akdim et al., 2023; Nakane et al., 2014). Further, transparency and prevention of deception are critical issues to consider, especially as robots become more emotionally expressive or autonomous. Overly human-like cues can mislead users about a robot's true capabilities, agency, or sentience (Balkenius and Johansson, 2022; Akbulut et al., 2024). Best practices call for clear communication about robot limitations and artificiality, avoiding cues that might suggest consciousness or deep emotions, and making reasoning and decision processes explicit (Datey et al., 2023; Papagni and Koeszegi, 2021). From our analysis we note the following points and guidelines.

- Cultural insensitivity and lack of contextual adaptation frequently undermine user acceptance and equity. Many emotional or social cues, colors, and gestures do not have universal meanings; robots designed for one context may cause confusion or exclusion in another (Liu et al., 2025). Participatory and cross-cultural design, as well as customizable features, are recommended to ensure local relevance and accessibility by existing studies (You et al., 2025; Wang et al., 2024).

- Ethical Issues in Specific Design Contexts**.** While many ethical issues, such as privacy, bias, and transparency, are present across various domains of robot design, unique challenges arise in specific contexts. Recognizing these distinct issues is essential for developing context-sensitive and ethically robust design guidelines. We identified numerous issues in specific design contexts shown in Figure 2 as follows:

    - Service Robots (Workplace, Hospitality, Retail). A unique ethical issue in the service robot context is the threat to job security and its impact on worker morale. The presence of service robots may evoke fears of displacement, social comparison, and declining job satisfaction among human employees (Tojib et al., 2023, Liu et al., 2025; Wang et al., 2023). Additionally, service robots are vulnerable to misconduct, vandalism, or abuse in public-facing and low-supervision environments, necessitating robust and resilient design as well as consideration of social status and setting (Wan and Chen, 2024). These concerns are less prominent in social or therapeutic robot research.

    - Social Robot Embodiment (Elderly, Children, Therapy). In social robot embodiment, especially for long-term companionship or assistive contexts, a unique challenge is the potential for deception through emotional mimicry, which means robots simulating emotions, agency, or sentience may lead users (notably older adults or children) to over-attribute human-like qualities, which can affect attachment and well-being (Henkel et al., 2019). Misalignment between developer and user values is also highly prominent in this domain, as designers' priorities may not reflect the needs or sensitivities of vulnerable users (Bradwell et al., 2019). The risk of erosion of human agency and over-reliance is heightened in social robot embodiment, as users may become dependent on robots for social or emotional needs (Osawa et al., 2011).

    - Social and Emotional Robot Design (Healthcare, Education, Marketing). Within emotional robot design, a unique concern is the ethics of emotional feedback and manipulation, the use of facial expressions, vocal tone, or affective cues to influence user emotions or behaviors, which may be especially problematic in healthcare or educational settings (Yan et al., 2024; de Melo et al., 2023). Overreliance and dependency on emotional robots is also more acute in these contexts, as they may inadvertently substitute for human relationships and affect users' social development or mental health (Terada and Takeuchi, 2017). In addition, emotional

authenticity and transparency take on special importance: users may mistake artificial emotional displays for genuine human empathy, potentially impacting expectations and therapeutic outcomes (de Melo et al., 2023).

- Appearance and Interaction (Sex, Intimacy, and Identity). In domains involving sex robots or identity exploration, unique ethical issues include the objectification and sexualization of embodied robots, particularly female-presenting forms, which can reinforce harmful gender stereotypes and raise complex questions about consent, agency, and social impact (Borau et al., 2021). The potential for abuse and discrimination—both toward robots and through robots as instruments—emerges strongly in these contexts (Winkle and Mulvihill, 2024). Additionally, the risk of overgeneralization and lack of individuation—where users treat robots as undifferentiated members of a category—can undermine trust and appropriate human-robot teaming (Abubshait et al., 2023).

- Human-robot and AI Agent Interaction (Conversational Agents, AR/VR, Digital Companions). For anthropomorphic AI agents and virtual companions, a unique challenge is overload and distraction—overly visual or persistent agents can fatigue or distract users, especially in multitasking or augmented reality environments (Reinhardt et al., 2020). Contextual inappropriateness is also a distinctive concern, as agents designed for one domain (e.g., home) may not translate well to another (e.g., workplace or healthcare), potentially causing confusion or ethical breaches (Xie et al., 2023). The erosion of human rights and dignity—notably in care and religious roles—has particular salience in highly humanlike AI agent deployments (Miller, 2020).

4. Conclusions

The design of robots include the consideration of ethical challenges that arise from integrating robots into human social environments, and among others, involves issues such as fairness, dependency, autonomy, and the risk of potential harm to humans. As discussed within the chapter, research on the role of ethics in HRI is multidisciplinary involving researchers from various fields working to develop ethical frameworks and guidelines for the design and use of robots. One reason why ethics is important to consider during the design process for HRI is that robots are becoming equipped with sophisticated social skills and operating in a range of applications from retail and education to healthcare. From this observation, there are numerous opportunities for ethical issues to be considered when designing human-robot interfaces. Among others, this means that theories from psychology, robotics, communication, information science, and management have been applied to different human-robot scenarios and with varying degrees of success. To some extent the wide range of theories that apply to ethics and HRI is explainable by considering the complexity of human and robot behavior during social interactions. Simply put, the complexity of human behavior results in numerous ethical issues when humans interact with robots.

In terms of the use of theory to guide HRI research (and with implications for ethics), from the above literature review and Tables 2, 3, and 4 in our view there are two main theories which stand out- these are the theory related to the Uncanny Valley effect and the theory which addresses the anthropomorphism of robots. The reason is that both of these theories tie the appearance and behavior of a robot to its perception and evaluation which has been a topic of interest amongst HRI researchers since the early days of robot development. Additionally, there are other areas where theories have been successfully applied to HRI in the context of ethical considerations and among others, they have focused on robots serving a particular social role during social interactions with a person. For example, considering ethics, theories which apply to the social categorization of robots such as by their perceived race, ethnicity, or gender have guided HRI research and has shown that biases may be directed against robots. Additionally, other studies exploring ethics and HRI have adopted a humanistic perspective and have sought to determine how robots are experienced by users in social contexts.

Considering the two conceptual frameworks used to categorize the theories presented in Tables 3 and 4 we note that there is some redundancy among the theories presented which is to be expected. For example, the Uncanny Valley effect and the Terror Management Theory both address human reaction to robots approaching the dip of the uncanny valley. However, Terror Management Theory has the requirement that people may feel threatened by humanoid appearing robots (which raises numerous issues of ethics for HRI). It could be argued that the Uncanny Valley effect makes the same prediction. Thus, as a call for action there is: (i) a need to critically evaluate the theories guiding HRI research in terms of their usefulness for HRI, (ii) a need to determine whether there is redundancy among the theories, and (iii) a need to begin the discussion of deciding whether the theories can be combined into broad conceptual frameworks that not only reduce the redundancy and number of theories applied to HRI but also provide explanatory power across a range of applications. This recommendation is supported by the conclusion of other researchers such as Trafton, Raymond, and Khemlani (2021) who argued that to advance our understanding of HRI we should focus not only on the results of individual studies but also on developing a theoretical framework to guide research on HRI.

In addition, in many examples of robotics, the robot is designed to replace the role of a human, thus to some extent there is a need to replicate classic studies on human behavior and ethics which originally examined human-human interaction to determine whether the results of such studies can be extended to human-robot interaction. Of course, studies on HRI that are replications or extensions of previous studies have already been attempted and have provided a valuable knowledge base for HRI researchers (Leichtmann, Nitsch, and Mara, 2022). As an example see the work by Wullenkord and Eyssel (2020). However, determining the extent to which theories that relate to human behavior or human-human interactions should be replicated within the domain of HRI requires a careful evaluation of which studies actually need to be replicated, else we run the risk that literally hundreds of theories would be replicated within the context of HRI.

From the literature review and from recent developments in robotics, we can ask- What are the implications of the HRI theories for future research especially when ethical issues are concerned? As one answer, there is a current trend for robots to be equipped with the latest techniques of artificial intelligence, including large language models. Thus, robots are becoming smarter, displaying more sophisticated social skills, and conversing with individuals in an expanding range of applications. This will necessarily produce gaps in our understanding in how humans interact with increasingly smart robots and will require an examination of current theories, conceptual frameworks, and models to determine whether they might apply to robots equipped with the latest techniques of artificial intelligence. In terms of how to conceptualize theories, while we think the categories that were used to group the theories in Tables 3 and 4 have face validity, the authors would argue that a new overarching category is necessary due to the increasing use of artificial intelligence in HRI. Thus, we conclude from a review of theories guiding research on HRI that there is a current gap in the literature explaining how artificial intelligence influences performance in HRI. This is essentially a "pacing problem" caused by technology advancing so quickly that if theory is not forward-looking enough it may quickly be outpaced by developments in technology.

From the above discussion and summary tables an important question to ask is how to conceptualize and think about the role of theory in guiding HRI research? While the literature search resulted in a summary table consisting of several theories, there are surely more theories in the literature that have been used to guide research on HRI given the goal of researchers to provide a theoretical framework on how humans interact with robots. And given that robots equipped with machine learning algorithms are beginning to display sophisticated social skills, some HRI researchers are applying theory to conceptualize robot behavior that is becoming autonomous in its decision making and behavior; among others, this approach applies to robots, automated cars and digital robo-assistants. From the literature (Dennis et al. 2016) the ethical issues implicated by autonomous systems are well known. Given the many theories guiding HRI research which were presented in summary Table 3 and 4, as a call for action, we advocate for more discussion and efforts aimed at developing conceptual frameworks which integrate theories which focus on a similar topic or make the same predictions to understand the factors thought to influence HRI. As an

example, say a researcher proposes a theory predicting that robots with a humanoid appearance will be evaluated as having a higher level of mental agency than robots that are more mechanical in appearance. Reviewing Tables 3 and 4, the Theory of Mind and Damasio's Theory of Mind/Consciousness combined with the theory related to robot anthropomorphism would be good starting points to explore the idea that the perception of robot agency may to a large extent, depend on robot design, It would be interesting to look carefully at the theories and see what their common features are and how they could be coalesced into a conceptual framework to advance research on HRI, ethics, and robot agency. From this approach, basic research could be conducted to establish how different levels of human appearance in robot design could affect the evaluation of robot agency- which we note is an emerging topic of interest within the HRI community that has implications for ethics. In addition, such a framework would have value for smart robots that interact with humans in tasks which require the robot to have control over its actions and knowledge of its action's consequences.

It is also evident from the review of the literature and theories presented in the summary tables that while much of the research on HRI is empirical based, people who interact with robots can be thought to experience the robot in a social context and through HRI construct a social reality in which the negotiation of a common meaning of reality occurs. Given the importance of epistemology and phenomenology in understanding how humans may perceive and experience robots, more humanistic theories such as Aesthetic Theory or Cultural Theory need to be considered for HRI. Further, the use of qualitative research methods need to be considered more carefully to uncover how individuals actually experience robots in different social contexts. Thus, as a call for action there needs to be greater effort among HRI researchers to employ a humanistic perspective to human interaction with robots, especially as robots are experienced in the wild, implicate ethical issues, and lived with for extended periods of time.

Broadly speaking, many of the theories provided in the summary tables focus on whether we consider robots to be part of an individual's in-group or as an out-group member- the two prominent theories on HRI alluded to earlier, robot anthropomorphism and the Uncanny Valley effect apply here. However, from psychology there is also a theory on this very topic which has been adopted by HRI researchers (Barfield, 2023 a,b), the theory is termed- Social Identity Theory (SIT) (Tajfel, et al. 1979). SIT is used to specify the circumstances under which people think of themselves or others as individuals or as members of the same group (Edwards et al. 2019; Tajfel et al, 1979). Due to advances in artificial intelligence which have occurred over the last five years, robots now interact with individuals as group members in the role of problem solver and decision maker. Thus, SIT would be useful for determining whether the robot is accepted within a group and whether its contribution is given value during group decision making (Cao, Stewart and Leonard, 2008; Edwards et al, 2019). And if we consider theories related to anthropomorphism or the Uncanny Valley effect as fundamental theories guiding research on HRI, then SIT could be used to make predictions about whether group membership affects the process of robot anthropomorphism and the extent to which robots perceived as an out-group member are judged to be within the dip of the Uncanny Valley. This is another example where theories could be combined to create a broad conceptual framework in which to guide HRI research.

As robots continue to gain in social skills and intelligence and take-on different roles within society there will continue to be a need to determine which existing theories from disciplines such as marketing, communication, psychology, and information sciences, are a good fit for HRI. However, many of the theories in Tables 3 and 4, whether redundant with other theories or not, have already been successfully applied to HRI and used to answer fundamental questions about HRI. Among others, such theories include telepresence and social presence theory which have been used to determine how robots influence the user's sense of presence within an environment (see Biocca, Harms, and Burgoon, 2003; Lombard and Ditton, 1997; Short et al., 1976) and the Computers as Social Actors theory which has shown that in many situations people react to robots as they do people (Nass and Brave, 2005; Nass, Steuer, and Tauber, 1994). And as another example of successfully adopting theories from other disciplines to study HRI, as discussed earlier in the chapter, the field of HCI has provided several theories that have been successfully extended to the domain of HRI. For example, Heuer and Stein (2020) when explaining how HCI theories are useful for

designing human-robot interfaces proposed a model which consisted of elements thought to influence HRI based on factors found to be important for HCI such as a person's attitude, experience, and situated knowledge. Other examples of theories which have been successfully used for HRI include Expectancy Violation theory which relates to how people respond to unexpected behaviors by robots (Burgoon, 2015), Communication Privacy Management theory which deals with how people maintain and coordinate privacy boundaries when interacting with robots (Petronio, 2012), and studies which focused on issues thought to influence HRI including robot uncertainty (Berger and Calabrese, 1975) and robot trust (Hancock, et al. 2011). Summarizing, many of the theories presented in Table 3 and 4 consist of a particular theoretical approach or have specific elements which may relate to a researcher's interests; thus, the summary tables can serve as a useful resource for identifying theories which can be used to guide research activities in HRI.

While the theories used to investigate HRI have proven beneficial to the research community, it is also necessary in providing the reader some insights into how to critically evaluate theories guiding HRI research which we will do by presenting guidelines and examples. To begin we note that critically assessing a theory will allow the researcher to determine if the theory offers a valid and reliable explanation for the phenomena it seeks to explain. For example, the Uncanny Valley effect is theorized to occur when an image approaches the dip in the uncanny valley curve and if so, will produce a feeling of unease or even revulsion (Mori, 1970). The data from numerous studies has shown that the theory reliably produces the predicted effect for computer-generated images and for android robots. However, the shooter bias paradigm described by Bartneck and colleagues (2018) predicts that a darker colored robot will be subject to more bias compared to a lighter colored robot. For this theory, the results of prior studies suggest that several unexplored factors may mediate the shooter-bias effect; thus, more data is needed to validate the theory and determine if the predictions of the theory are reliable and generalizable to real-world scenarios. Another guideline for the critical evaluation of a theory is to ensure that the research and knowledge from the theory are based on sound principles and evidence. To determine this, the method section of the paper (explaining the study design, participants, and procedure) should be critically reviewed and the statistical procedures used to test the predictions carefully evaluated to make sure the analysis is valid. Further, in the discussion section of a paper, the discussion should clearly answer the question of whether the theory guiding HRI research was supported from the data. Here it is also important to determine whether the evidence for the theory is consistent with findings from other theories, this should also be addressed in the discussion section of a paper. However, some of the theories presented in Tables 3 and 4 were the first to explore a particular topic in HRI, in this case it may be possible to compare the results to other studies where the theory was used with similar technology. For example, CASA was developed for HCI, but it is reasonable to compare the results found within HCI to the domain of robots given features common to both technologies. In addition, to perform a critical evaluation of a theory we should ask what issue(s) does the theory seek to explain and was the theory successful in explaining the issue(s)? Typically, the introduction to a paper should clearly state the objectives of any theory to be tested in the research.

As a critical review of a theory, it is also important to analyze the logical structure of the theory and to look for fallacies or inconsistencies in the reasoning that supports the theory. The scope of the theory is also important to consider, that is, the range of phenomena that the theory is designed to explain and whether the theory is too narrow or too broad in scope. For example, Lakoff's Likeness Theory (Kumar et al. 2022) explicitly states that it emphasizes the importance of face-saving acts and to avoid impolite behavior. Also, it is important to critically evaluate a theory to consider the testability of the theory. The question is whether the theory can be tested with empirical evidence and are there specific predictions that can be tested to support or refute the theory? For example, self-determination theory (SDT) is used to explains how people's innate psychological needs and growth tendencies motivate their choices in the absence of external influences. SDT suggests that all people have three basic psychological needs that are universally important for psychological well-being, these include autonomy, competence, and relatedness. Finally, it is important to consider the implications of the theory and evaluate whether the theory has implications that are consistent with other well-established theories and whether the implications are supported by available evidence.

From the summary tables and literature review we present two main conclusions for the use of theories to guide HRI research and with implications for ethics. The first is that several of the theories presented in Tables 3 and 4 were developed to describe and/or guide HRI research with the goal to design robots that facilitate social interactions with people in different contexts. For that reason, numerous theories from psychology and communication sciences have been adopted by researchers to explore HRI. The second conclusion is that HRI researchers have adopted different theories with the goal to help design the actual physical features of a robot in order to influence how robots should be implemented in society as a form of technology. Here, theories on anthropomorphism and the Uncanny Valley effect have proved especially useful in guiding HRI research. We also conclude that as we go forward, we need to develop broad conceptual frameworks to guide HRI research especially for robots that are gaining in intelligence and social skills based on different AI techniques. Finally, we also need to shift through the numerous theories that thus far have been used to guide HRI research, and from this begin to develop broader and more comprehensive frameworks which include features of different theories which address the same phenomena.

References

Abubshait, A., Weis, P. P., Momen, A., and Wiese, E., Perceptual Discrimination in the Face Perception of Robots is Attenuated Compared to Humans. *Scientific Report*s, Vol. 13, 16708, 2023.

Akbulut, C., Weidinger, L., Manzini, A., Gabriel, I., and Rieser, V., All Too Human? Mapping and Mitigating the Risk from Anthropomorphic AI, *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 7(1), 13–26, 2024.

Akdim, K., Belanche, D., and Flavián, M., Attitudes Toward Service Robots: Analyses of Explicit and Implicit Attitudes Based on Anthropomorphism and Construal Level Theory. *International Journal of Contemporary Hospitality Management*, Vol. 35(8), 2816-2837, 2023.

Allan, D. D., Vonasch, A. J., and Bartneck, C., The Doors of Social Robot Perception: The Influence of Implicit Self-theories, *International Journal of Social Robotics*, Vol. 14 (1), 127-140, 2022 a.

Allan, D. D., Vonasch, A. J., and Bartneck, C., "I Have to Praise You Like I Should?" The Effects of Implicit Self-Theories and Robot-Delivered Praise on Evaluations of a Social Robot, *International Journal of Social Robotics*, Vol. 14(4(, 1013-1024, 2022 b.

Asemi, A., Ko, A., and Nowkarizi, M., Intelligent Libraries: a Review on Expert Systems, Artificial Intelligence, and Robot, *Library Hi Tech*, Vol. 39 (2), 412-434, 2021.

Austin R. R,. McLane, T. M., Pieczkiewicz, D. S., Adam, T., Monsen K. A., Advantages and Disadvantages of Using Theory-Based Versus Data-Driven Models with Social and Behavioral Determinants of Health data, *J Am Med Inform Assoc.,* Vol. 30(11), 818-1825, 2023.

Balkenius, C., and Johansson, B., Almost Alive: Robots and Androids, *Frontiers in Human Dynamics,* Vol. 4*,* Article 703879, 2022.

Barfield J. K., Discrimination and Stereotypical Responses to Robots as a Function of Robot Colorization, *UMAP '21: Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 109–114, 2021a.

Barfield, J. K., Self-Disclosure of Personal Information, Robot Appearance, and Robot Trustworthiness, *30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 67-72, 2021b.

Barfield, J. K., Designing Social Robots to Accommodate Diversity, Equity, and Inclusion in Human-Robot Interaction, *CHIIR '23: Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*, 463–466, 2023a.

Barfield, J. K., A Hierarchical Model for Human-Robot Interaction, *ASIS&T Mid-Year Conference Proceedings*, 2023b.

Barfield, J. K., Towards Diversity, Equity, and Inclusion in Human-Robot Interaction. In: Kurosu, M., Hashizume, A. (eds) *Human-Computer Interaction. HCII 2023,* Lecture Notes in Computer Science, Vol. 14013, Springer, Cham. 2023c.

Bartneck, C., Nomura, T., Kanda, T., Suzuki, T., and Kennsuke, K., Cultural Differences in Attitudes Towards Robots. *Proceedings of the AISB Symposium on Robot Companions: Hard Problems And Open Challenges In Human-Robot Interaction*, Hatfield, 1-4, 2005.

Bartneck, C., Yogeeswaran, K., Ser, Q-M., Woodward, G., Sparrow, R., Wang, S., Eyssel. F., Robots and Racism, *Proceedings of ACM/IEEE International Conference on Human Robot Interaction (HRI '18)*, 196-204, 2018.

Bento, L. F. H., Prates, R. O., and Chaimowicz, L., Using Semiotic Inspection Method to Evaluate a Human-Robot Interface, *7th Joint LA-WEB/CLIHC Conference*, 77-84, 2009.

Berger, C. R., and Calabrese, R. J., Some Explorations in Initial Interaction and Beyond: Toward a Developmental Theory of Interpersonal Communication, *Human Communication Research*, Vol. 1, 99–112, 1975.

Bernotat, J., Eyssel, F., and Sachse, J., Shape It - The Influence of Robot Body Shape on Gender Perception in Robots, *9th International Conference on Social Robotics (ICSR)*, 75-84, 2017.

Bernotat, J., Eyssel, F., and Sachse, J., The (Fe)male Robot: How Robot Body Shape Impacts First Impressions and Trust Towards Robots. *International Journal of Social Robotics*, Vol. 13, 477–489, 2021.

Biocca, F., Harms, C., and Burgoon, J. K., Toward a More Robust Theory and Measure of Presence: Review and Suggested Criteria. *Presence*, 12(5), 456–480, 2003.

Borau, S., Otterbring, T., Laporte, S., and Fosso Wamba, S., The Most Human Bot: Female Gendering Increases Humanness Perceptions of Bots and Acceptance of AI, *Psychology & Marketing,* Vol. 38(7), 1052–1068, 2021.

Bradwell, H. L., Edwards, K. J., Winnington, R., Thill, S., and Jones, R. B., Companion Robots for Older People: Importance of User-Centred Design Demonstrated through Observations and Focus Groups Comparing Preferences of Older People and Roboticists in South West England. *BMJ Open,* Vol. 9(9), e032468, 2019.

Breazeal, C., Function Meets Style: Insights from Emotion Theory Applied to HRI, *IEEE Transactions on Systems, Man, and Cybernetics, Part C- Applications and Reviews,* Vol. 34 (2), 187-194, 2004. .

Burgoon, J. K., Expectancy Violations Theory. In C. R. Berger, M. E. Roloff, S. R. Wilson, J. P. Dillard, J. Caughlin, and D. Solomon (Eds.), *The International Encyclopedia of Interpersonal Communication*, 2015.

Cao, M., Stewart, A., and Leonard, N. E., Integrating Human and Robot Decision-Making Dynamics with Feedback: Models and Convergence Analysis, *47th IEEE Conference on Decision and Contro*l, 1127-1132, 2008.

Cover, J. A., Curd, M., and Pincock, C., *Philosophy of Science: The Central Issues,* Second Edition, W. W. Norton & Company, 2012.

Datey, I., Zheng, W., Walquist, E., Zhou, X., Berishaj, K., Valentine, M., and Zytko, D., Ethical Participatory Design of Social Robots Through Co-design with Children, *32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN),* 1–7), 2023.

de Keyser, A., and Kunz, W. H., Living and Working with Service Robots: a TCCM Analysis and Considerations for Future Research, *Journal of Service Management*, Vol. 33 (2),165-196, 2022.

Demutti, M., D'Amato, V., Recchiuto, C., Oneto, L., and Sgorbissa, A., Assessing Emotions in Human-Robot Interaction Based on the Appraisal Theory, *31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) - Social, Asocial, and Antisocial Robots,* 1435-1442, 2022.

Dennis, L., Fisher, M., Slavkovik, M., and Webster, M., Formal Verification of Ethical Choices in Autonomous Systems, *Robotics and Autonomous Systems*, Vol. 77, 1-14, 2016.

de Melo, C. M., Gratch, J., Marsella, S., and Pelachaud, C., Social Functions of Machine Emotional Expressions, *Proceedings of the IEEE,* Vol. 111(8), 1382–1397, 2023.

de Souza, C. S., Prates, R., Barbosa, S., and Barbosa, S., A Semiotic Engineering Approach to User Interface Design, *Knowledge-Based Systems*, Vol. 14(8), 461-465, 2001.

de Souza, C. S., D., Leitão, C. F., Prates, R., da Silva, E. J., The Semiotic Inspection Method, *IHC '06: Proceedings of VII Brazilian Symposium on Human Factors in Computing Systems,* 148–157, 2006.

Duffy, B. R., Anthropomorphism and the Social Robot, *Robotics and Autonomous Systems*, Vol. 42 (3-4), 177-190, 2003.

Edwards, C., Edwards, A., Stoll, B., Lin, X., and Massey, N., Evaluations of an Artificial Intelligence Instructor's Voice: Social Identity Theory in Human-Robot Interactions, *Computers in Human Behavior*, Vol. 90, 357-362, 2019.

Epley, N., Waytz, A., and John, T., On Seeing Human: A Three-Factor Theory of Anthropomorphism, *Psychological Review*, Vol. 114(4), 864-886, 2007.

Eyssel, F., and Hegel, F., (S)he's Got the Look: Gender Stereotyping of Robots, *Journal of Applied Social Psychology*, Vol. 42(9), 2213-2230, 2012.

Eyssel F., and Kuchenbrandt, D., Social Categorization of Social Robots: Anthropomorphism as a Function of Robot Group Membership, *Brit J Soc Psychol,* Vol. 51, 724–731, 2012.

Eyssel, F., Kuchenbrandt, D., Hegel, F., and De Ruiter, L, Activating Elicited Agent Knowledge: How Robot and User Features Shape the Perception of Social Robots, *IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication,* 851–857, 2012b.

Eyssel, F., and Loughnan, S., It Don't Matter if You're Black or White? Effects of Robot Appearance and User Prejudice on Evaluations of a Newly Developed Robot Companion, 422-433, In: Herrmann G., Pearson M. J., Lenz A., Bremner, P., Spiers A., & Leonards U. (eds) Social Robotics, ICSR 2013, *Lecture Notes in Computer Science*, Vol. 8239. Springer, Cham, 2013.

Fox, J., and Gambino, A., Relationship Development with Humanoid Social Robots: Applying Interpersonal Theories to Human/Robot Interaction, *Cyberpsychology Behavior and Social Networking,* Vol. 24(5), 294-299, 2021.

Fraune, M. R., Our Robots, Our Team: Robot Anthropomorphism Moderates Group Effects in Human-Robot Teams**,** *Frontiers in Psychology,* Vol. 11, 2020.

Friedman, K., Theory Construction in Design Research: Criteria: Approaches, and Methods, *Design Studies,* Vol. 24(6), 507-522, 2003.

Gonsior, B., Buß, M., Sosnowski, S., Wollherr, D., Kühnlenz, K., and Buss, M., Towards Transferability of Theories on Prosocial Behavior from Social Psychology to HRI, *IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*, 101-103, 2012.

Gray, K., and Wegner, D. M., Feeling Robots and Human Zombies: Mind Perception and the Uncanny Valley, *Cognition*, Vol. 125(1), 125-130, 2012.

Grootendorst, M., BERTopic: Neural Topic Modeling with a Class-based TF-IDF Procedure. arXiv preprint arXiv:2203.05794, 2022.

Hancock, P. A., Billings, D. R., and Schaefer, K. E., Can You Trust Your Robot? *Ergonomics in Design,* Vol. 19(3), 24-29, 2011.

Hantula, D. A., What Performance Management Needs Is a Good Theory: A Behavioral Perspective, *Industrial and Organizational Psychology- Perspectives on Science and Practice*, Vol. 4 (2), 194-197, 2011.

Henkel, Z., Baugus, K., Bethel, C. L., and May, D. C., User Expectations of Privacy in Robot Assisted Therapy, *Paladyn, Journal of Behavioral Robotics,* Vol. 10(1), 140–159, 2019.

Heuer, T., and Stein, J., From HCI to HRI: About Users, Acceptance and Emotions, *2nd International Conference on Human Systems Engineering and Design (IHSED) - Future Trends and Applications*, 149-153, 2020.

Higgins, E. T., Beyond Pleasure and Pain, *Am. Psychol*., Vol. 52, 1280-1300, 1997.

Higgins, E. T., Value From Regulatory Fit, *Current Directions in Psychological Science*, Vol. 14(4), 209–213, 2005.

Hirvonen, N., Multas, A-M., Nygard, T., and Huotari, M-L., Cognitive Authority: A Scoping Review of Empirical Research, *J Assoc Inf Sci Technol*, 1-28, 2024.

Hmelo, C.E., Gotterer, G.S. & Bransford, J.D. A theory-driven approach to assessing the cognitive effects of PBL. *Instructional Science* **25**, 387–408 (1997)
Hoorn, J. F., Theory of Communication: II. Befriending a Robot Over Time, *International Journal of Humanoid Robotics*, Vol. 17(6), 1-25, 2020a.

Hoorn, J. F., Theory of Robot Communication: I. The Medium is the Communication Partner, *International Journal of Humanoid Robotics*, Vol. 7 (6), 1-21, 2020b.

Jones, D., and Gregor, S., The Anatomy of a Design Theory, *Journal of the Association for Information Systems*, Vol. 8(5), 2—7.

Jones, K. S., Niichel, M. K., and Armstrong, M. E., Robots Exhibit Human Characteristics: Theoretical and Practical Implications for Anthropomorphism Research, *13th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 137-138, 2018.

Kühne, R., and Peter, J., Anthropomorphism in Human-Robot Interactions: A Multidimensional Conceptualization, *Communication Theory*, Vol. 33 (1), 42-52, 2023.

Kumar, S., Itzhak, E., Edan, Y., Nimrod, G., Same-Fleischmann, V., and Tractinsky, N., Politeness in Human-Robot Interaction: A Multi-Experiment Study with Non-Humanoid Robots, *International Journal of Social Robotics,* Vol. 14 (8), 1805-1820, 2022.

Kwon, M., Biyik, E., Talati, A., Bhasin, K., Losey, D. P., and Sadigh, D., When Humans Aren't Optimal: Robots that Collaborate with Risk-Aware Humans, *ACM/IEEE International Conference on Human-Robot Interaction*, 43-52, 2020.

Leichtmann, B., Nitsch, V., and Mara, M., Crisis Ahead? Why Human-Robot Interaction User Studies May Have Replicability Problems and Directions for Improvement, *Frontiers in Robotics and AI,* Vol. *9,* 2022.

Letheren K., Jetten J., Roberts, J., and Donovan, J., Robots Should be Seen and Not Heard Horizontal Ellipsis Sometimes: Anthropomorphism and AI Service Robot Interactions, *Psychology & Marketing*, Vol. 38(12), 2393-2406, 2021.

Li, M., Guo, F., Ren, Z., and Duffy, V. G., A Visual and Neural Evaluation of the Affective Impression on Humanoid Robot Appearances in Free Viewing, *International Journal of Industrial Ergonomics,* Vol. 88, 103159, 2022..

Lombard, M., and Ditton, T., At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, Vol. 3(2), JCMC321, 1997.

Love, T., Philosophy of Design: A Meta-Theoretical Structure for Design Theory, *Design Studies,* Vol. 21(3), 293-313, 2000.

Liu, Y., Yang, D., Chang, D., Jiang, P., Pan, Y., and Liu, Z., The Effects of robot Anthropomorphic Characteristics on Employees' Anti-robot Sabotage. *Scientific Reports,* Vol. 15, Article 22064, 2025,

Lu, L., Zhang, P., and Zhang, T. T., Leveraging "Human-Likeness" of Robotic Service at Restaurants, *International Journal of Hospitality Management*, Vol. 94, 1-9, 2021.

Manthiou, A., Klaus, P., Kuppelwieser, V. G., and Reeves, W., Man vs Machine: Examining the Three Themes of Service Robotics in Tourism and Hospitality, *Electronic Markets*, Vol. 31(3), 511-527, 2021.

Martini, M. C., Buzzell, G. A., and Wiese, E., Agent Appearance Modulates Mind Attribution and Social Attention in Human-Robot Interaction. In: Tapus, A., André, E., Martin, J. C., Ferland, F., Ammi, M. (eds) Social Robotics. *ICSR 2015*. Lecture Notes in Computer Science(), vol 9388. Springer, Cham., 2015.

Marvel, J. A., Bagchi, S., Zimmerman, M., Aksu, M., Antonishek, B., Wang, Y., Mead, R., Fong, T., and Amor, H. B., Test Methods and Metrics for Effective HRI in Real-World Human-Robot Teams, *HRI '20: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction,* 652–653, 2020.

Miller, L. F., Human Rights of Users of Humanlike Care Automata. *Human Rights Review*, Vol. 21(2), 181–205, 2020,

Mori, M., The Uncanny Valley, *Energy*, Vol. 7(4), 33-35, 1970.

Nakane, M., Young, J. E., and Bruce, N., More Human than Human?: A Visual Processing Approach to Exploring Believability of Android Faces, *HAI '14: Proceedings of the Second International Conference on Human-Agent Interaction,* 377–381, 2014.

Nass, C., and Brave, S., *Wired for Speech,* Cambridge, MA, MIT Press, 2005.

Nass, C., Steuer, J., and Tauber, E. R., Computers are Social Actors, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 72–78, 1994.

Nilsen, P., Making Sense of Implementation Theories, Models and Frameworks. *Implementation Sci*., Vol. 10, 2015.

Osawa, H., Yamada, S. Social Modification Using Implementation of Partial Agency Toward Objects. *Artif Life Robotics,* Vol. 16, 78–81, 2011.

Otterbacher, J and Talias, M., S/he's too Warm/Agentic! The Influence of Gender on Uncanny Reactions to Robots, *12th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 214-223, 2017.

Papagni, G., and Koeszegi, S., A Pragmatic Approach to the Intentional Stance: Semantic, Empirical and Ethical Considerations for the Design of Artificial Agents, *Minds & Machines,* Vol. 31(4), 505–534, 2021.

Paterson, M., Why Robot Embodiment Matters: Questions of Disability, Race and Intersectionality in the Design of Social Robots, *Medical Humanities*, Vol. 50(4), 694–704, 2025.

Perugia, G., Boor, L., van der Bij, L., Rikmenspoel, O., Foppen, R., and Guidi, S., Models of (Often) Ambivalent Robot Stereotypes: Content, Structure, and Predictors of Robots' Age and Gender Stereotypes, *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction,* 428–436, 2023.

Perugia, G., Guidi, S., Bicchi, M., and Parlangeli, O., The Shape of Our Bias: Perceived Age and Gender in the Humanoid Robots of the ABOT Database, *17th ACM/IEEE International Conference on Human-Robot Interaction (HRI),* 110–119), 2022.

Petronio, S., *Boundaries of Privacy: Dialectics of Disclosure*, Suny Press, 2012.

Popper, K. R., *The Logic of Scientific Discovery*, Martino Fine Books, 2014.

Prewett, M. S., Johnson, R. C., Saboe, K. N., Ellipott, L. R., and Coovert, M. D., Managing Workload in Human-Robot Interaction: A Review of Empirical Studies, *Computers in Human Behavior*, Vol. 6 (5), 840-856, 2010.

PRISMA Statement, 2024, accessed 8-15-2024, available at: https://www.prisma-statement.org/.

Rabe, T., Callis, A., Zheng, Z., Heard, J., Bailey, R., and Alm, C., Theory of Mind Assessment with Human-Human and Human-Robot Interactions, *Human Computer Interaction Thematic Area Conference Held as Part of the 24th International Conference on Human-Computer Interaction (HCII),* 564-579, 2022.

Reinhardt, J., Hillen, L., and Wolf, K., Embedding Conversational Agents into AR: Invisible or with a Realistic Human Body? *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*, 299–310, 2020.

Salvini, P., Laschi, C., and Dario, P., Design for Acceptability: Improving Robots' Coexistence in Human Society, *International Journal of Social Robotics,* Vol. 2(4), 451–460, 2010.

Schein, C., and Gray, K., The Eyes are the Window to the Uncanny Valley, Mind Perception, Autism and Missing Souls, *Interaction Studies,* Vol. 16 (2),173-179, 2015.

Seymour, M., Yuan, L., Dennis, A., and Riemer, K, Crossing the Uncanny Valley? Understanding Affinity, Trustworthiness, and Preference for More Realistic Virtual Humans in Immersive Environments, *52nd Hawaii International Conference on System Sciences (HICSS)*, 1748-1758, 2019.

Shalley, C. E., Writing Good Theory: Issues to Consider, *Organizational Psychology Review*, Vol. (3), 258-264, 2012.

Short, J., Williams, E., and Christie, B., Theoretical Approaches to Differences between Media, *Social Psychology of Telecommunications,* 61–66, London: Wiley, 1976.

Singh, A., D'Arcy, M., Cohan, A., Downey, D., & Feldman, S., SciRepEval: A Multi-Format Benchmark for Scientific Document Representations. *Conference on Empirical Methods in Natural Language Processing*, 2022.

Sparrow, R., Do Robots Have Race?: Race, Social Construction, and HRI*, IEEE Robotics & Automation Magazine*, Vol. 27 (3), 44-150, 2020.

Sparrow, R., Robotics Has a Race Problem, *Science Technology & Human Values,* Vol. *45* (3), 538-560, 2020.

Strait, M. K., Floerke, V. A., Ju, W., Maddox, K., Remedios, J. D., Jung, M. F., and Urry, H. L., Understanding the uncanny: Both atypical features and category ambiguity provoke aversion toward humanlike robots. *Frontiers in Psychology*, Vol. 8, 1366, 2017.

Tajfel, H., Turner, J. C., Austin, W. G., and Worchel, S., An Integrative Theory of Intergroup Conflict, *Organizational Identity: A Reade*r, 56-65, 1979.

Teh, N., Hu, S. Y., and Soh, H., A Theoretical Framework for Large-Scale Human-Robot Interaction with Groups of Learning Agents, *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 489-493, 2021.

Teo, G., Wohleber, R., Lin, J., and Reinerman-Jones, L, The Relevance of Theory to Human-Robot Teaming Research and Development, *7th International Conference on Applied Human Factors and Ergonomics/International Conference on Human Factors in Robots and Unmanned Systems,* 175-185, 2017.

Terada, K., and Takeuchi, C., Emotional Expression in Simple Line Drawings of a Robot's Face Leads to Higher Offers in the Ultimatum Game, *Frontiers in Psycholog*y, Vol. 8, 724, 2017.

Thepsoonthorn, C., Ogawa, K., and Miyake, Y., The Exploration of the Uncanny Valley from the Viewpoint of the Robot's Nonverbal Behaviour, *International Journal of Social Robotics*, Vol. 3 (6), 1443-1455, 2021.

Tojib, D., Sujan, R., Ma, J., and Tsarenko, Y., How Does Service Robot Anthropomorphism Affect Human Co-Workers? *Journal of Service Management*, Vol. 34(4), 750–769, 2023.

Trafton, J. G., Raymond, P., and Khemlani, S., The Power of Theory, *ACM Transactions on Human-Robot Interaction*, Vol. 10(1), 111-112, 2021.

Ullrich, D and Diefenbach, S., Truly Social Robots Understanding Human-Robot Interaction from the Perspective of Social Psychology, *12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)*, Vol 2, 39-45, 2017.

Wagner, A. R., and Arkin, R. C., Analyzing Social Situations for Human–Robot Interaction, *Interaction Studies,* Vol. 9(2), 277–300, 2008.

Walther, J. B., Interpersonal Effects of Computer-Mediated Interaction: A Relationship Perspective, *Communication Research*, Vol. 19, 52-90, 1992.

Wan, X., and Chen, H., A., Research on the Influence Mechanism of Humanization Degree of Service Robots on User Misbehavior, *Management Decision*, 2024.

Wang, N., Li, Z., Shi, D., Chen, P., and Ren, X., Understanding Emotional Values of Bionic Features for Educational Service Robots: A Cross-Age Examination Using Multi-Modal Data, *Advanced Engineering Informatics,* Vol. 6*2*(Part D), 102956, 2024.

Wang, P. X., Kim, S., and Kim, M., Robot Anthropomorphism and Job Insecurity: The Role of Social Comparison, *Journal of Business Research,* Vol. 164, Article 114003, 2023.

Weng, H., and Hirata, Y., Design-Centered HRI Governance for Healthcare Robot, *Journal of Healthcare Engineering*, 1-8, 2022.

Winkle, K., and Mulvihill, N., Anticipating the Use of Robots in Domestic Abuse: A Typology of Robot Facilitated Abuse to Support Risk Assessment and Mitigation in Human-Robot Interaction, *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction,* 781–790, 2024.

Wong, A., and Wong, J., How Anthropomorphic Features in Service Robots Foster Co-Creation and Wellbeing Among Gen Z Consumers?, 2025, *Ingentia Connect*, https://www.ingentaconnect.com/content/mcb/yc/2025/00000026/00000003/art00003;jsessionid=2lct81heqdct.x-ic-live-03.

Wu, J., Lyu, X., Wang, Y., Liu, T., Zhao, S., and Xue, L., Combining Design Neurocognition Technologies and Neural Networks to Evaluate and Predict New Product Designs: A Multimodal Human–Computer Interaction Study, *Electronics,* Vol. 14(6), 1128, 2025.

Wullenkord, R., Eyssel, F., Societal and Ethical Issues in HRI, *Curr Robot Rep* 1, 85–96, 2020.

Xie, L., and Lei, S., The Nonlinear Effect of Service Robot Anthropomorphism on Customers' Usage Intention: A Privacy Calculus Perspective, *International Journal of Hospitality Management,* Vol. 107, 103312, 2022

Xie, Y., Zhu, K., Zhou, P., and Liang, C., How Does Anthropomorphism Improve Human-AI Interaction Satisfaction: A Dual-Path Model, *Computers in Human Behavior*, Vol. 148, 107878, 2023.

Yan, L., Qiling, X., and Wu, S., The effects of voice emotions on users' willingness to pay decision-making process of automated delivery robots: An ERP study. In S. H. Sheu (Ed.), Industrial Engineering and Industrial Management (IEIM 2024) (pp. 112–128). Springer, 2024.

You, Z., Fayaz Ahmad, S., Yan, F., Irshad, M., Garayev, M., and Bani Ahmad Ayassrah, A. Y. A, Investigating the Impact of Safety, Cultural and Character Traits Issues in the Adoption of Humanized Robots in Education, *Humanities and Social Sciences Communications,* Vol. 12, Article 976, 2025.

Zhang, Y., Cao, Y., Proctor, R. W., and Liu, Y., Emotional Experiences of Service Robots' Anthropomorphic Appearance: A Multimodal Measurement Method, *Ergonomics,* Vol. 66(12), 2039–2057, 2023.