Human, Machine, or Hybrid? Using Anthropomorphism to Conceptualize Trust in Robots

Completed Research Full Paper

Samia Cornelius Bhatti
University of Michigan
samiaco@umich.edu

Lionel Peter Robert Jr.
University of Michigan
lprobert@umich.edu

Abstract

While robots appear to be more and more human-like in form and function, they are still machines. People can hence perceive them as humans or machines. With varying human-like designs and user perceptions, there is much confusion about how to measure trust in human-robot relationships. While some researchers use human-like trusting beliefs to conceptualize trust, others use machine-like trusting beliefs to do the same. In this paper, we present a conceptual model and related research propositions to help researchers determine the correct conceptualization of trust for human-robot interaction. We propose that anthropomorphism, or perceptions of humanness about the robot, can dictate the conceptualization of trust in human-robot relationships.

Keywords

Human-Robot Interaction, Robots, Humanoid, Trust, Anthropomorphism, Uncanny Valley

Introduction

Robots represent a highly complex technology that aims to simulate human appearance and behavior. They are central to the discussion of the shift of agency and control from humans to technology. They hence invite exploration and perspectives on transformed human-technology relationships (Murray, Rhymer, and Sirmon 2021). They also present new opportunities and challenges associated with human-robot interaction and collaboration. In an age where, robots are increasingly taking over human roles such as decision-making, customer service, and knowledge management, new theoretical and practical questions arise about their design, effectiveness, and sustainability. In this environment, the trust that users develop in robots is key to understanding their use and acceptance.

Trust can be defined as “the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party” (Mayer, Davis, and Schoorman 1995). Although first believed to be a human-human interaction construct, researchers soon concluded that technology could also be trusted. After all, people trusted their cars, their computers, or their word processors (Du, Pradhan, Yang, and Robert Jr 2019). The definition provided by Mayer et al (1995) did not limit the concept of trust to only human-human interaction, thus paving the way for the use of the concept in human-technology interaction (Wang, Qiu, Kim, and Benbasat 2016). As a result, trust, along with perceived usefulness and ease of use (Davis 1989), became a central concept to predicting technology acceptance and use (Hoff and Bashir 2015).

Trust is important for human-technology relationships because of the perceived risk embedded in the acceptance of any new technology, owing to the complexity and non-determinism of technology behavior. This is particularly important in robots, where artificial intelligence can be quite complex for people to understand and predict. Moreover, the prevalence of robots in the workplace, working as teammates,

Bhatti, S., and Robert, L. P. (2023). **Human, Machine, or Hybrid? Using Anthropomorphism to Conceptualize Trust in Robots**, *Proceedings of the 26th Americas Conference on Information Systems*, (**AMCIS 2023**), Aug 10-12, Panama City, Panama.

customer assistants, service attendants, medical assistants, and security guards can result in a dramatic change in organizational structures (Faraj, Pachidi, and Sayegh 2018).

Trust in technology has been conceptualized and measured both as with factors typically associated with human beings, such as ability, integrity, and benevolence (Vance, Elie-Dit-Cosaque, and Straub 2008), as well as factors typically associated with machines, like reliability, functionality, and helpfulness (Mcknight, Carter, Thatcher, and Clay 2011). The choice of which measures to use when is often not quite clear. It can become even more ambiguous when robots appear and behave like human beings (Baker, Phillips, Ullman, and Keebler 2018). In such cases, perhaps a trust measure encompassing human characteristics like integrity, benevolence, and competence should be used. But robots are also machines, and their mechanistic nature can foster trust associated with reliability, functionality, and helpfulness. Therefore, this begs the question: *What measures, then, should researchers employ to study trust in robots?* The use of the wrong set of constructs can result in misleading theoretical implications, incorrect design, and acceptance of conclusions. However, using all of them will lead to the problem of multi-correlation and inflated findings. To answer the question about which set of measures to use needs further exploration and understanding.

In this paper, we present a trust model that takes into consideration both human- and machine-trust in robots. We also propose that using the wrong trust concept can lead to weaker or insignificant relationships with outcome variables and invite researchers to assess the possibilities of using one path or the other to study robot acceptance. We propose that anthropomorphism can dictate the conceptualization of trust and consequently the utility of the robot. However, we assert that anthropomorphism may need to be re-conceptualized from a continuous variable a much lumpier concept (see Bhatti and Robert, 2023). Nonetheless, the first step forward is to propose and examine whether anthropomorphism does dictate which set of constructs to employ.

Trust in Robots

Trust in a robot is a characteristic of a human user relative to the robot in question (You and Robert Jr 2018). It can be a belief or an attitude (Lee and See 2004), an affective response (Coeckelbergh 2012), a sense of willingness (Mayer, Davis, and Schoorman 1995), a mutual understanding (Azevedo, Raizer, and Souza 2017), or an act of reliance (Wang, Humphrey, Liao, and Zheng 2018) of a human regarding a robot. It is a multidimensional construct that is a result of some stimuli and leads to reliance on a robot.

Extant literature has explored trust in robots with respect to automated systems (Hancock, Billings, Schaefer, Chen, De Visser, and Parasuraman 2011; Hoff and Bashir 2015), intelligent agents (Israelsen and Ahmed 2019) or technology in general (Mcknight, Carter, Thatcher, and Clay 2011). But trust in robots does not necessarily align with trust in automated systems or intelligent agents. A “robot” is an embodied agent with physical manifestation (Baker, Phillips, Ullman, and Keebler 2018). An automated system on the other hand may be a computerized process without any physical form. Similarly, an intelligent agent could exist on a computer in the form of an algorithm. Robots also differ from other kinds of technology due to their proximity and nature of interaction with human beings. They interact with human beings in a much more intimate (human-like) manner than other types of technologies. This robot-other technology distinction can inform our understanding of how the conceptualization of trust may differ for robots as compared to other technologies. For one, the embodiment and physical presence of a robot make the design of the robot a key consideration for the formation of trust. Secondly, robots are usually deployed in dynamic environments where they work alongside human beings, suggesting that the ability to socialize with these human agents will also inform the development of trust. Hence, the conceptualization of trust in robots is a result of design and user perceptions of that design.

Human, Machine, or Hybrid?

Robots can be characterized as a machine, human, or human-machine hybrid (humanoids) depending upon their technology design. While human-like and machine-like trusting beliefs can be found in the extant literature, little is known about conceptualizing trust for human-machine hybrids. Furthermore, considering that a robot is a machine embedded with human-like characteristics, it is often hard to decide which trust concepts should be used to conceptualize trust in human-robot interactions.

The three human-like trusting beliefs prevalent in literature are competence, integrity, and benevolence (Mayer, Davis, and Schoorman 1995). Competence is the belief that a person has the skills to have influence in a specific domain. Integrity is the belief that a person follows a set of accepted principles and rules. And benevolence is the belief that the person will want to do good to the trustor regardless of incentives and motives. These trusting beliefs assume that the trustee has the power to make ethical decisions. Since technologies usually do not have such power, some researchers have developed alternative constructs to measure trust in machines. McKnight et al (2011) use reliability, functionality, and helpfulness as direct corollaries to human-like trusting beliefs to measure trust in machines. Reliability is similar to integrity and is the belief that the technology will consistently perform properly. Functionality is similar to competence and defines the ability of the technology to do what it is meant to do. Helpfulness is similar to benevolence and suggests that technology is able to provide adequate help.

With technology becoming more and more human-like, the boundary between what is human and what is not keeps blurring, especially in the case of robots, that can look and behave like human beings. Does that suggest that human-like trusting beliefs can be used to measure trust in robots? This is a question that has no clear answer, but using the wrong type of trusting beliefs to measure trust in robots may result in lower path coefficients between trust variables and outcomes leading to inaccurate conclusions about design, trust, and acceptance. While the literature is not clear on the appropriateness of using one set of trusting beliefs as compared to the other, it seems like the influence of trust on outcome variables is driven by users' perceptions of humanness. In the following section, we explain how human-like attribution can lead people to develop different trusting beliefs.

Attribution of Human-Likeness

A number of theories and perspectives, including the computers as social actors paradigm (CASA), social response theory, and theory of anthropomorphism explain the formation of human-like perceptions towards technology (Epley, Waytz, and Cacioppo 2007). CASA dictates that human beings apply social rules to their interaction with computers, as a response to a social situation (Nass, Steuer, and Tauber 1994). The authors further explain that human-computer interaction is "fundamentally social", and simplistic human-like cues in technology design can trigger familiarity and acceptance. This highlights two important stimuli: a social situation and a human-like design. Social response theory reinforces this view and states that people respond to technology with higher social cues as though it were human (Gefen and Straub 2003). It suggests that whenever technology exhibits a human-like appearance or behavior, the users are more likely to attribute human-like characteristics to them. For example, people make kind comments to a computer that demonstrates courtesy (Wang, Baker, Wagner, and Wakefield 2007), and find technology with more human-like characteristics to be more trustworthy (Qiu and Benbasat 2009).

The theory of anthropomorphism proposed by Epley and colleagues (2007) explains that human-like design in technology acts as a situational catalyst to human-like attribution. Human-like attribution or *anthropomorphism* dictates that the neural systems that are used to make judgments about other human beings are also activated when making judgments about non-human agents (Iacoboni, Lieberman, Knowlton, Molnar-Szakacs, Moritz, and Fiske 2004). This manner of mental processing facilitates humans in comprehension and understanding as well as the development of social connections, by attributing familiar qualities to familiar entities (Epley, Waytz, and Cacioppo 2007; Cornelius and Leidner 2021). As such, anthropomorphism is the attribution of human-like knowledge representations to nonhuman entities based on past knowledge acquired through observation of, and interaction with, other human beings, and observed to understand the world in a better way, build social connections, and enhance belonging.

In the IS literature, anthropomorphism is often discussed in the context of technology use and acceptance. From a design perspective, anthropomorphism is discussed as the deployment of humanness in technology design in the form of appearance and behavior (Duffy 2003), and from a user perspective it is studied as an attribution process that can take place mindfully or mindlessly (Kim and Sundar 2012). Human-like attribution is often catalyzed in the presence of human-like design, suggesting that human-like appearance and behavior can motivate users to perceive technology as human-like. Perceptions of technology humanness have been associated with increased familiarity, use, and acceptance (Nass, Steuer, and Tauber 1994).

In summary, computers as social actors and the social response theory explain user's human-like response to social cues in design and situation, and the literature on anthropomorphism discusses the deployment of human-like cues in technology as well as the attribution of human-like characteristics to technology as a result of those cues. The literature suggests that human-likeness in robots leads to the attribution of human-like characteristics to it, which can dictate the trusting beliefs that users form in a robot. In essence, we propose that the conceptualization of trust in robots is driven by anthropomorphism or attribution of humanness to robots. We posit that differences in human-like design can lead to different perceptions of humanness as a result of high, low or mixed levels of human-like attribution: machine, human, or hybrid. These perceptions can then change the trusting beliefs as well as their influence on outcome variables. In this study, we examine whether people perceive robots varying in human-like design differently, that is, anthropomorphize them differently, and if these differences in humanness perceptions change how different trusting beliefs influence outcome variables associated with intention and behavior (for example, intention to use, reliance, satisfaction, or enjoyment).

Conceptual Model and Propositions

In this section, we develop our conceptual model and relevant research propositions. As presented in the literature above we propose that differences in human-like design in robots will lead to varying perceptions of humanness. Or in other words, people will attribute varying degrees of human-likeness to robots based on differences in human-like design. We also suggest that the conceptualization of trust will be more humanlike when robots are perceived as human-like, and more machine-like when robots are perceived as less human-like. Lastly, we explain how the uncanny valley can influence the formation and conceptualization of trust in hybrid robots. Below we present our conceptual model.

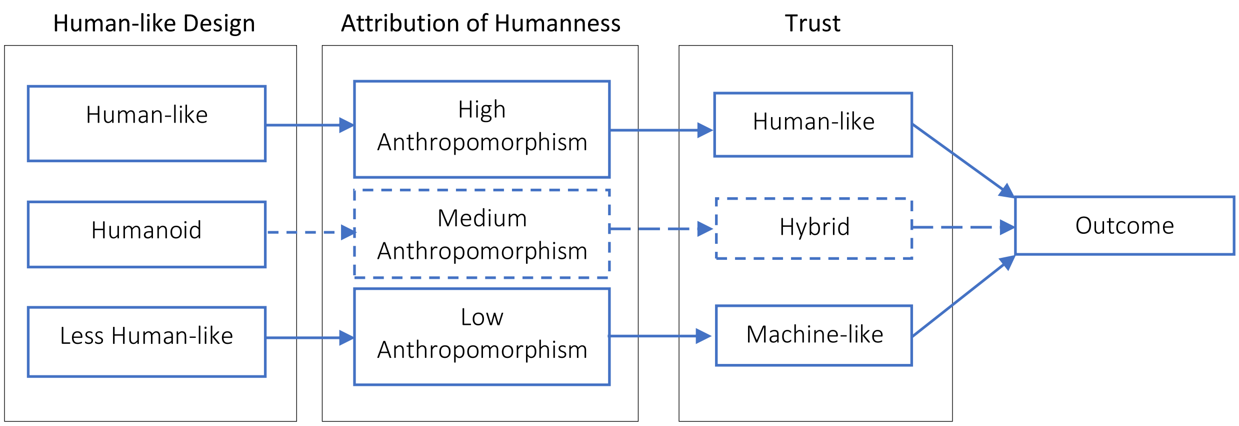

Figure 1: Conceptual Model

Technology Design and Anthropomorphism

Anthropomorphism traditionally has been associated with the attribution of human-like characteristics as a user tendency, anthropomorphism from a designer's perspective is to create stimuli to motivate such thinking (Epley, Waytz, and Cacioppo 2007). This means that designers of technological artifacts build on human beings' chronic tendency to anthropomorphize nonhuman entities, to improve interaction with technology by embedding them with human-like form and/or function (Złotowski, Proudfoot, Yogeeswaran, and Bartneck 2015). Designers delve into such design implementation because anthropomorphism has been associated with building long-term relationships with technology (Bartneck and Forlizzi 2004) and holds great promise for advancing human-technology interaction (Sproull, Subramani, Kiesler, Walker, and Waters 1996). Hence, by embedding technology with human-like characteristics, designers provide the stimulus for anthropomorphism.

There are a number of ways designers embed human-like design stimuli in robots to motivate anthropomorphism. The conventional way to do so is by configuring the robot's physical appearance (Hancock, Billings, Schaefer, Chen, De Visser, and Parasuraman 2011). Attribution of human-likeness often takes place at the first encounter, and the decision to trust a robot can hinge upon its physical appearance (Złotowski, Sumioka, Nishio, Glas, Bartneck, and Ishiguro 2016). Robots with a human-like appearance are often perceived as more human-like and considered more likable and trustworthy. For example, Castro-Gonzalez and colleagues (2016) investigate whether robots will full body versus just arms is more likable. They also assess if human-like versus machine-like movement in robots leads to a difference in human-like perceptions. Their study shows that movement in robots is positively associated with anthropomorphism and likeability. Gong (2008) in his experiment with interface agents manipulate facial characteristics to depict human, robot, or humanoid appearances to assess differences in anthropomorphism and resulting social responses. The study shows that increases in human-likeness result in increasing human perceptions, as well as perceptions of trustworthiness. Similarly, Seymour et al (2021) manipulate levels of realism in a digital avatar interacting with people in a simulated VR environment to conclude that observers rated the more human-like avatar to be more trustworthy and preferred it as a virtual agent.

Function or behavior is the second way in which robots can be deployed with human-likeness. Responsiveness, adaptiveness, pro-sociality, and natural language processing are all functions that add to the intelligence of the robots and are known to increase human-like behaviors towards them. Extant research has shown that as a robot becomes more intelligent, it becomes easier for users to perceive them as fellow teammates (Glikson and Woolley 2020). Bickmore and colleagues (2013) tested the effectiveness of a robot museum guide and found that its responsiveness had a positive impact on engagement, enjoyment, learning, and trust.

In summary, variations in the human-like technology design of a robot can result in various degrees of human-like attributions to the robot, such that human-like robots are perceived as more human-like (high anthropomorphism) as compared to those that are humanoid or less human-like (medium/low anthropomorphism). And so, we propose:

P1a: Individuals will perceive more human-like robots to be more human-like (high anthropomorphism) as compared to robots that are humanoid or less human-like (medium/low anthropomorphism).

We also know from extant research that people's perception of technology can change over time (Bhatti and Robert 2023). Researchers have also found people to perceive technology as less human when it makes a mistake or as they understand it better (Dzindolet, Peterson, Pomranky, Pierce, and Beck 2003; Bhatti and Robert 2023). These findings suggest that although more human-like robots will be perceived as more human-like, these perceptions from human-like attribution, and hence anthropomorphism, can change over time as people interact more with a robot. Therefore we propose that humanness perceptions or degrees of anthropomorphism can change over time.

P1b: Individuals will change their humanness perceptions (anthropomorphism) towards a robot over time.

Anthropomorphism and Trusting Beliefs

The next set of propositions explain the influence of trusting beliefs on outcome variables. We propose that the kind of trusting beliefs an individual forms about a robot depend on the degree of the robot's anthropomorphism. The cognitive match between human-like perception and trusting beliefs is supported by the CASA paradigm, the social response theory, as well as literature on anthropomorphism. These perspectives suggest that people attribute human-like characteristics (human-like trusting beliefs) to technology that is perceived to be more human-like. For example, Zhang et al (2010) tested different features of a service robot with elderly participants and found that more human-like features of the robot were associated with more emotional trust. Trust research also suggests that cognitive consistency is important in trust relationships: maintaining consistency between trusting beliefs and perceptions requires much less cognitive load (McKnight, Cummings, and Chervany 1998).

The robot's appearance is particularly associated with human emotions (Krämer, Lucas, Schmitt, and Gratch 2018). A robot's attractiveness and its visual similarity to humans can evoke unconscious emotional reactions, similar to perceptions of warmth, benevolence, and positive intentions that can influence human trust and behavior (Khan and Sutcliffe 2014). Similar results have been reported by Zhang and colleagues (2010), who find that a human-like robot's face evokes perceptions of warmth and benevolence. Human-like embodiment (e.g., like arms, facial features, and smooth movement) and physical presence of a robot are also associated with an increase in emotional trust (Bartneck, Suzuki, Kanda, and Nomura 2007). We thus predict that when forming human-like perceptions, or in cases of high anthropomorphism, people will develop human-like trusting beliefs more significantly as compared to machine-like trusting beliefs, and these beliefs will impact outcome variables more strongly as compared to machine-like beliefs. This will be true both within the level of anthropomorphism in a robot and across the levels of anthropomorphism among robots.

P2: For robots with high anthropomorphism, human-like trusting beliefs will have a stronger influence on outcome variables than machine-like trusting beliefs.

For robots that are developed to be less human-like in appearance or behavior, i.e., they have more machine-like physical features like industrial wheels and lack of facial features, people are less likely to attribute human-like characteristics (Epley, Waytz, and Cacioppo 2007). Trust in such robots is presumed to be developed with the help of mechanistic trusting beliefs: reliability, functionality, and helpfulness (Mcknight, Carter, Thatcher, and Clay 2011). Extant literature has also called this kind of trust cognitive trust (Glikson and Woolley 2020).

Mechanistic or machine-like trusting beliefs have been found to deteriorate in cases where a robot makes an error. For example, Dzindolet et al (2003) found that errors committed by an automated decision aid significantly reduced the reliability of the aid. Researchers have also found that erroneous functions have a stronger effect on trust than correct functions (Manzey, Reichenbach, and Onnasch 2012). When directly measuring perceived trust, de Visser and Parasuraman (2011) report that people's trust in automation was directly associated with reliability.

It is hard to imagine people endowing attributes like benevolence and integrity to robots that are more machine-like and less human-like. Therefore, we propose that for robots that are not human-like, beliefs like reliability, functionality, and helpfulness will be significantly stronger than integrity, benevolence, and competence. Consequently, these machine-like trusting beliefs will have a stronger influence on outcome variables.

P3: For robots with low anthropomorphism, machine-like trusting beliefs will have a stronger influence on outcome variables than human-like trusting beliefs.

Lastly, we propose that human likeness is not a two-dimensional construct, and neither is anthropomorphism. A robot can be human-like and machine-like, but it can also be a human-machine hybrid. The idea of a humanness continuum may sound daunting because of the complexity associated with quantifying degrees of human likeness, but it has already been explored in the past with the help of hypotheses like the uncanny valley. The uncanny valley hypothesis explains that human-like characters that are not completely human-like appear eerie and lead to withdrawal and unacceptance (Mori 1970). Extant research on the uncanny valley hypothesis with human-like technology design iterates one basic claim: nonhuman features deployed in more human-like characters are unsettling (MacDorman, Green, Ho, and Koch 2009). Eeriness is associated with less familiarity (Tinwell, Grimshaw, and Williams 2010), low attractiveness (Cornelius, Leidner, and Bina 2023), and low trust ratings (Weisman and Peña 2021). Hybrid robots may fall into the uncanny valley, suggesting that people either do not trust them or trust them less than human or machine-like robots.

One of the reasons we have provided for the alignment of human-like perception and trusting beliefs is that people prefer cognitive consistency in trust relationships. When a robot has both human-like and machine-like features and is perceived as a hybrid, both human-like and machine-like cognitive representations are activated in the brain. For the successful application of these representations to the robot, they must be integrated. This process can be cognitively straining to the user and can cause cognitive dissonance leading to less favorable attitudes and withdrawal (MacDorman and Ishiguro 2006; Kätsyri, Förger, Mäkäräinen, and Takala 2015).

Lastly, extant literature informs us that people like to make categories of trusting beliefs and match them with categories of human-likeness (human-like versus machine-like) (McKnight, Cummings, and Chervany 1998). However, people may find it difficult to put hybrid robots in a category. Research has shown that confusion in category identification can make it hard for people to make perceptions about an entity (Ho and MacDorman 2010). This suggests that people may refuse to develop human-like or machine-like trusting beliefs about a hybrid robot. Consequently, there will be a less or insignificant impact on outcome variables.

In summary, we suggest that for robots that have both human-like and machine-like design characteristics, both human-like and machine-like trusting beliefs will be weaker as compared to human-like and less human-like robots.

P4: For robots with medium anthropomorphism, both human-like and machine-like trusting beliefs will have a weaker impact on outcome variables as compared to more human-like (high anthropomorphism) and less human-like robots (low anthropomorphism).

Discussion

Our conceptual model for trust in HRI highlights the role of anthropomorphism in determining the development and conceptualization of trust and its impact on outcome variables. We also highlight that degree of human-like design alone should not be taken as a determinant to trust conceptualization, because user's humanness perceptions towards a particular design can change over time. Thus, we emphasize the dynamic nature of anthropomorphism as it drives trust formation from human-like design. In doing so, we encourage both cross-sectional and longitudinal research for trust conceptualization.

With our conceptual model and propositions we present a number of paths to trust conceptualization that can be taken up by researchers to explore and study. We specially encourage researchers to explore trust constructs that may be relevant to the conceptualization of trust for hybrid robots. Although we propose that human and machine trusting beliefs will be weaker for hybrid robots, testing this proposition may bring to light interesting findings about how to conceptualize trust for such entities. It would also be worthwhile to test the external validity of the conceptual model by investigating the derived propositions in various contexts where trust is a central construct for robot acceptance and use. Two popular contexts are security robots and robots in medicine. Furthermore, social responses associated with human-human inter-personal exchange are often culture- and situation-dependent (Madhavan and Wiegmann 2007). Hence, testing the model's generalizability for different cultures and situations will also help strengthen the validity of this work.

A number of limitations in this work can also be taken up for further research. Although we have used popular trusting beliefs from extant literature, we have come across other trusting beliefs in the literature like utility and performance. Similarly, different types of trust like institution-based trust and interpersonal trust may be more relevant to a particular context or situation. We encourage researchers to use our model to test other trusting beliefs as well as kinds of trust that they believe are relevant to their research and outcome variables. Our study lacks quantitative data analysis to test the propositions. While we hope to conduct additional analysis in the future to determine the accuracy of the propositions, we invite researchers to take up our propositions for experimentation and testing. We especially encourage the analysis of longitudinal data to test the propositions.

The primary goal of the model is to help determine which trusting beliefs should be used to measure trust in human-robot interactions, and for that we believe anthropomorphism can provide the answer. The first step hence is to find out if anthropomorphism in robots does indeed drive differences in trust conceptualization. If that is true, then regardless of what measures are used for human versus system-like trusting beliefs, we belief that human-like attribution (high anthropomorphism) will lead to a stronger impact of human-like trusting beliefs on outcome variables as compared to all else, and machine-like attributions (low anthropomorphism) will lead to a stronger impact of machine-like trusting beliefs on outcome variables as compared to all else.

Conclusion

The study aims to propose, and invited researchers to confirm, that trusting beliefs can differ from human-like perceptions. While acknowledged in extant research, differential trusting beliefs lack standardization and categorization. Moreover, the conceptualization and application of trust concepts remain unclear and dispersed. We hope that with our research we bring to light some important considerations for trust and robotics researchers. While we endeavor to provide some trust categorization with respect to design, we believe that the conceptualization of trust is as dynamic as the attribution of human-like characteristics to robots. Hence, it would make sense to revisit and test the given propositions as human-like design evolves and user perceptions of humanness change.

REFERENCES

Azevedo, C. R., Raizer, K., and Souza, R. 2017. "A vision for human-machine mutual understanding, trust establishment, and collaboration," in *IEEE Conference on Cognitive and Computational Aspects of Situation Management*, pp. 1-3.

Baker, A. L., Phillips, E. K., Ullman, D., and Keebler, J. R. 2018. "Toward an understanding of trust repair in human-robot interaction: Current research and future directions," *ACM Transactions on Interactive Intelligent Systems* (8:4), pp. 1-30.

Bartneck, C., and Forlizzi, J. 2004. "A design-centred framework for social human-robot interaction," in *3th IEEE international workshop on robot and human interactive communication*, pp. 591-594.

Bartneck, C., Suzuki, T., Kanda, T., and Nomura, T. 2007. "The influence of people's culture and prior experiences with Aibo on their attitude towards robots," *AI and Society* (21), pp. 217-230.

Bhatti, S., and Robert, L. 2023. "What Does It Mean to Anthropomorphize Robots? Food For Thought for HRI Research," in *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction.*, https://doi.org/10.1145/3568294.3580119.

Bickmore, T. W., Vardoulakis, L. M., and Schulman, D. 2013. "Tinker: a relational agent museum guide," *Autonomous agents and multi-agent systems* (27), pp. 254-276.

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

Coeckelbergh, M. 2012. "Can we trust robots?" *Ethics and information technology* (14), pp. 53-60.

Cornelius, S., and Leidner, D. 2021. "Acceptance of anthropomorphic technology: a literature review," in *Proceedings of the 54th Hawaii International Conference on System Sciences*, pp. 6422-6431.

Cornelius, S., Leidner, D. E., and Bina, S. 2023. "Significance of Visual Realism–Eeriness, Credibility, and Persuasiveness of Virtual Influencers," in *Proceedings of the 56th Hawaii International Conference on System Sciences*, pp. 3421-3430.

Davis, F. D. 1989. "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, pp. 319-340.

de Visser, E., and Parasuraman, R. 2011. "Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload," *Journal of Cognitive Engineering and Decision Making* (5:2), pp. 209-231.

Du, N. H., Pradhan, A. K., Yang, X. J., and Robert Jr, L. P. 2019. "Look who's talking now: Implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload," *Transportation research part C: emerging technologies* (104), pp. 428-442.

Duffy, B. R. 2003. "Anthropomorphism and the social robot," *Robotics and autonomous systems* (42:3-4), pp. 177-190.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., and Beck, H. P. 2003. "The role of trust in automation reliance," *International journal of human-computer studies* (58:6), pp. 697-718.

Epley, N., Waytz, A., and Cacioppo, J. T. 2007. "On seeing human: a three-factor theory of anthropomorphism," *Psychological Review* (114:4), pp. 864.

Faraj, S., Pachidi, S., and Sayegh, K. 2018. "Working and organizing in the age of the learning algorithm," *Information and Organization* (28:1), pp. 62-70.

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., and Sasse, M. A. 2003. "The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 529-536.

Gefen, D., and Straub, D. 2003. "Managing user trust in B2C e-services," *e-Service* (2:2), pp. 7-24.

Glikson, E., and Woolley, A. W. 2020. "Human trust in artificial intelligence: Review of empirical research," *Academy of Management Annals* (14:2), pp. 627-660.

Gong, L. 2008. "How social is social responses to computers? The function of the degree of anthropomorphism in computer representations," *Computers in Human Behavior* (24:4), pp. 1494-1509.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., and Parasuraman, R. 2011. "A meta-analysis of factors affecting trust in human-robot interaction," *Human factors* (53:5), pp. 517-527.

Ho, C. C., and MacDorman, K. F. 2010. "Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices," *Computers in Human Behavior* (26:6), pp. 1508-1518.

Hoff, K. A., and Bashir, M. 2015. "Trust in automation: Integrating empirical evidence on factors that influence trust," *Human factors* (57:3), pp. 407-434.

Iacoboni, M., Lieberman, D., M., Knowlton, B. J., Molnar-Szakacs, I., Moritz, M., and Fiske, A. P. 2004. "Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline," *Neuroimage* (21:3), pp. 1167-1173.

Israelsen, B. W., and Ahmed, N. R. 2019. ""Dave... I can assure you... that it's going to be all right..." A definition, case for, and survey of algorithmic assurances in human-autonomy trust relationships," *ACM Computing Surveys* (51:6), pp. 1-37.

Kätsyri, J., Förger, K., Mäkäräinen, M., and Takala, T. 2015. "A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness," *Frontiers in Psychology* (6), pp. 390.

Khan, R. F., and Sutcliffe, A. 2014. "Attractive agents are more persuasive," *International Journal of Human-Computer Interaction* (30:2), pp. 142-150.

Kim, Y., and Sundar, S. S. 2012. "Anthropomorphism of computers: Is it mindful or mindless?" *Computers in Human Behavior* (28:1), pp. 241-250.

Krämer, N. C., Lucas, G., Schmitt, L., and Gratch, J. 2018. "Social snacking with a virtual agent–On the interrelation of need to belong and effects of social responsiveness when interacting with artificial entities," *International Journal of Human-Computer Studies* (109), pp. 112-121.

Lee, J. D., and See, K. A. 2004. "Trust in automation: Designing for appropriate reliance," *Human Factors* (46:1), pp. 50-80.

MacDorman, K. F., and Ishiguro, H. 2006. "The uncanny advantage of using androids in cognitive and social science research," *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems* (7:3), pp. 297-337.

MacDorman, K. F., Green, R. D., Ho, C. C., and Koch, C. T. 2009. "Too real for comfort? Uncanny responses to computer generated faces," *Computers in human behavior* (25:3), pp. 695-710.

Madhavan, P., and Wiegmann, D. A. 2007. "Similarities and differences between human–human and human–automation trust: an integrative review," *Theoretical Issues in Ergonomics Science* (8:4), pp. 277-301.

Manzey, D., Reichenbach, J., and Onnasch, L. 2012. "Human performance consequences of automated decision aids: The impact of degree of automation and system experience," *Journal of Cognitive Engineering and Decision Making* (6:1), pp. 57-87.

Mayer, R. C., Davis, J. H., and Schoorman, F. D. 1995. "An integrative model of organizational trust," *Academy of Management Review* (20:3), pp. 709-734.

Mcknight, D. H., Carter, M., Thatcher, J. B., and Clay, P. F. 2011. "Trust in a specific technology: An investigation of its components and measures," *ACM Transactions on management information systems* (2:2), pp. 1-25.

McKnight, D. H., Cummings, L. L., and Chervany, N. L. 1998. "Initial trust formation in new organizational relationships," *Academy of Management Review* (23:3), pp. 473-490.

Mori, M. 1970. "Bukimi no tani (the uncanny valley)," *Energy* (7:4), pp. 33-35.

Murray, A., Rhymer, J. E., and Sirmon, D. G. 2021. "Humans and technology: Forms of conjoined agency in organizations," *Academy of Management Review* (46:3), pp. 552-571.

Nass, C., Steuer, J., and Tauber, E. R. 1994. "Computers are social actors," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 72-78.

Nowak, K. L., and Rauh, C. 2005. "The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction," *Journal of Computer-Mediated Communication* (11:1), pp. 153-178.

Qiu, L., and Benbasat, I. 2009. "Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems," *Journal of management information systems* (25:4), pp. 145-182.

Seymour, M., Yuan, L., Dennis, A., and Riemer, K. 2021. "Have We Crossed the Uncanny Valley? Understanding Affinity, Trustworthiness, and Preference for Realistic Digital Humans in Immersive Environments," Journal of the Association for Information Systems (22:3).

Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., and Waters, K. 1996. "When the interface is a face," *Human-computer interaction* (11:2), pp. 97-124.

Tinwell, A., Grimshaw, M., and Williams, A. 2010. "Uncanny behaviour in survival horror games," *Journal of Gaming and Virtual Worlds* (2:1), pp. 3-25.

Vance, A., Elie-Dit-Cosaque, C., and Straub, D. W. 2008. "Examining trust in information technology artifacts: the effects of system quality and culture," *Journal of management information systems*, (24:4), pp. 73-100.

Wang, L. C., Baker, J., Wagner, J. A., and Wakefield, K. 2007. "Can a retail web site be social?" *Journal of marketing* (71:3), pp. 143-157.

Wang, W., Qiu, L., Kim, D., and Benbasat, I. 2016. "Effects of rational and social appeals of online recommendation agents on cognition-and affect-based trust," *Decision Support Systems* (86), pp. 48-60.

Wang, Y., Humphrey, L. R., Liao, Z., and Zheng, H. 2018. "Trust-based multi-robot symbolic motion planning with a human-in-the-loop," *ACM Transactions on Interactive Intelligent Systems* (8:4), pp. 1-33.

Weisman, W. D., and Peña, J. F. 2021. "Face the uncanny: the effects of doppelganger talking head avatars on affect-based trust toward artificial intelligence technology are mediated by uncanny valley perceptions," *Cyberpsychology, behavior, and social networking* (24:3), pp. 182-187.

You, S., and Robert Jr, L. P. 2018. "Human-robot similarity and willingness to work with a robotic co-worker," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 251-260.

Zhang, T. K., B., S. M., Mosaly, P., and Hodge, L. 2010. "Service robot feature design effects on user perceptions and emotional responses," *Intelligent service robotics* (3), pp. 73-88.

Złotowski, J., Proudfoot, D., Yogeeswaran, K., and Bartneck, C. 2015. "Anthropomorphism: opportunities and challenges in human–robot interaction," *International journal of social robotics* (7), pp. 347-360.

Złotowski, J., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., and Ishiguro, H. 2016. "Appearance of a robot affects the impact of its behaviour on perceived trustworthiness and empathy," *Paladyn, Journal of Behavioral Robotics* (7:1).