The Look of Trust: How Does Physical Embodiment Shape Trust in Human-Agent Interaction?

ANGEL HSING-CHI HWANG, Cornell University, USA

In the present paper, we elaborate on a work-in-progress, a systematic review and qualitative meta-analysis regarding the effect of embodiment on trust for intelligent agents. Our primary interest examines whether a higher degree of embodiment and anthropomorphism facilitate trust-building in human-agent interaction. Meanwhile, we also summarize consensus in theoretical support and common methodological strategies. We conclude with unsolved questions that we continue to qualitatively code the set of literature in our current review, while identifying potential avenues for future research.

CCS Concepts: • **Human-centered computing → HCI theory, concepts and models**; **Collaborative and social computing theory, concepts and paradigms**.

Additional Key Words and Phrases: embodiment, trust, human-agent teamwork, intelligent agent, machine agency

ACM Reference Format:
Angel Hsing-Chi Hwang. 2022. The Look of Trust: How Does Physical Embodiment Shape Trust in Human-Agent Interaction?. 1, 1 (April 2022), 7 pages. https://doi.org/XXXXXXX.XXXXXXX

The human body incorporates, perhaps, one of the richest pieces of messages that one can possibly communicate to others – it reveals one's emotions, social identity, intentions, and more. Physical appearance or *embodiment* of individuals, therefore, serves a central role in trust-building during interpersonal communication and interaction [3, 27]. Indeed, a rich body of literature in social and behavioral psychology is dedicated to understanding how individuals form trust through gestures, non-verbal cues, and even attires. As scholars and practitioners make increased attempts to design and introduce intelligent agents as social entities, embedding them with tangible features and even identities, many interests arise to examine the effect of physical embodiment on trust in human-agent interaction. In this current paper, we describe a work-in-progress, where we conduct a systematic review and qualitative meta-analysis to synthesize relevant work and their insights thus far. In particular, the present review holds the following objectives:

- *(primary)* **To understand the effect of embodiment and anthropomorphism on users' trust toward intelligent agents.** Particularly, we decompose the impact of specific embodied features (e.g., figure, gesture, speech, locomotion) on trust-building in human-agent interaction.
- To synthesize common constructs of trust that have been examined in human-agent interaction research, while delineating how other fields conceptualize trust.
- To summarize existing methodological approaches, particularly teamwork or intra-group scenarios, researchers have adopted to study how users interact with embodied agents.
- To identify research gaps and open-ended questions regarding the physical embodiment of interactive agents.

Author's address: Angel Hsing-Chi Hwang, hh695@cornell.edu, Cornell University, Ithaca, NY, USA.

XXXX-XXXX/2022/4-ART $15.00
https://doi.org/XXXXXXX.XXXXXXX

Using the keywords "trust" and "embodiment," we assemble papers from the Web of Science, ACM Digital Library, IEEE Xplore, and Scopus database. We then go through a screening process to include papers using the following criteria: (1) the papers are empirical studies (including quantitative, qualitative, and mixed methods) investigating how human subjects interact with computer-mediated agents; (2) the papers examine the effect of different forms and/or levels of embodiment on trust through self-report, behavioral, or a combination of both types of measures; (3) during the studies, participants and agents work on given tasks collectively, sufficing the minimal conception of teams [12]. While we plan to conduct further coding to systematically summarize studies included in the present review, we describe key takeaways and reflections as we work on this review.

1 CONSENSUS IN THEORETICAL GROUNDWORK

1.1 Physical and social presence: Why do we trust humans with physical representation?

Across papers examined in this review, the primary interest in their literature review sections discusses why physical embodiment can enhance our trust toward other *humans*. A common theorem suggests that embodiment (e.g., a person embodying an avatar in a virtual environment) elicits the perceived physical and social presence of interactive others [13, 18, 21]. Being present in physical bodies allows others to more efficiently extract information during communication, as human bodies are the most naturalistic forms through which we process cues from our conversational and interactive partners [20, 35]. Such enriched information then serves as the foundation to establish trust. In other words, as humans present themselves through physical bodies, they are effectively offering pieces of "evidence" as they "persuade" others to trust them.

1.2 Trustworthiness as focused constructs of trust

As mentioned in recent theoretical (such as the Multi-Dimensional Measure of Trust (MDMT) framework in [19]) and empirical work (such as an re-examination of trust measures in human-robot interaction in [6]), users' trust toward intelligent agents are not only based on their performance but also on additional moral, intentional, and social aspects. Moreover, as proposed in the Deconstructed Trustee Theory, [36] has found users may assign distinct levels of trust toward the physical body and toward the identity of an agent, demonstrating yet another layer of complexity in investigating the impact of embodiment on trust. Together, scholars have posited that users' trust for agents should be studied as its own unique realm, instead of being directly inferred from existing knowledge in users' trust toward automation systems.

Despite the various dimensions of trust in human-agent interaction, we noticed a substantial portion of research in human-agent interaction remained focused on competence, expertise, and intelligence as the bases of trust, e.g., [7, 16, 17]. While these constructs can often be headily implemented in experimental design (which we will discuss further in the next section), they explain trustworthiness, rather than directly addressing the complexity of trust, to a greater extent [1]. Besides, a considerable number of studies examined trust through participants' compliance and reliance on agents during teamwork [10, 28]. Indeed, individuals tend to agree more and turn to people they trust, but it is worth noting that these behaviors reflect an agent's perceived trustworthiness and are consequences of trust – instead of the conception per se [1, 9]. Therefore, when researchers investigate these behaviors, it is critical to assess whether such behavioral patterns results from participants' trust toward agents or other antecedents. For instance, a person may comply more with an agent if they are more inclined to avoid conflicts.

In other disciplines, such as organizational behavior and social psychology, trust is often formed through long-term relationships [12]. That is, one can rarely establish trust firmly through a single occurrence. Instead, it is often developed through multiple, meaningful interactions. Nonetheless, research in human-agent interaction seldom studies individuals' trust for agents beyond a one-time encounter [34]. By contrast, there are certain critical components of trust that have been studied more extensively in this research area. For example, trust

repair is an intriguing yet highly complex element of trust, but previous work using social robots has readily investigated this topic [15, 16, 25].

2 EXISTING METHODOLOGICAL APPROACHES

2.1 Study protocol

The majority of work studying the effect of embodiment on trust in human-agent interaction adopts the following paradigm: participants would work with agents in different conditions of embodiment (either through between- or within-subject design) and respond to post-task questionnaires (including self-report items to survey their perceived trust toward the agents), while their teamwork behaviors and performance may, later on, be analyzed by researchers as behavioral measurements of trust. This protocol is often conducted under controlled settings in a lab environment, while there are also a small number of field studies [28].

2.2 Manipulation of embodiment

To vary forms of embodiment in these studies, some researchers would directly adopt different consumer products (e.g., Amazon Echo) [15, 23] or linearly add embodied features to various agents. For instance, in [14], Kim and colleagues compared three versions of agents: speech only, speech + body figure, and speech + body figure + movement. Regardless of how experimental manipulations were realized, researchers commonly hypothesized certain conditions possess lower or even no embodiment (e.g., an agent on-screen) while others have a higher degree of embodiment (e.g., a social robot with humanoid gestures). However, we notice quite a portion of work did not include manipulation checks to confirm whether different forms of embodiment were indeed perceived as more or less embodied.

2.3 Teamwork tasks and scenarios

During human-agent interaction, participants were typically instructed to work on tasks collaboratively with agent partners. As mentioned above, compliance – particularly when agents offer new paradigms or novel solutions to the teamwork challenges – during teamwork is often used as a behavioral measure to evaluate participants' trust toward agents [10, 11]. Counts, text lengths, and time spent on compliance have all been used as measurements in previous work, but again, careful examination is needed to parse out the influence of participants' characteristics (e.g., openness to adopt new changes and suggestions) on study results. Besides, some work has rigorously manipulated the load of information, quality, and even "personality" (e.g., pro-social vs. egotistical [7]) in agent responses, as bots' behaviors can be better controlled than humans'.

3 CONVERGENCE IN FINDINGS: WHEN EMBODIMENT IS (NOT) EFFECTIVE

In this short workshop paper, we will focus on elaborating two key findings among the results of our systematic review.

3.1 The impact of embodiment may be context-dependent

After synthesizing results across studies, we first remark that though a higher degree of embodiment does often effectively enhance participants' perceived trust toward agents, behavioral measures suggest the effect may be rather nuanced [11, 14, 22, 24, 24, 30, 34]. On one hand, this may again indicate the challenge of capturing trust through one's behaviors. On the other hand, several interaction effects may imply the function of embodiment is more complicated. In particular, numerous studies found the influence of embodiment in human-agent teams is conditioned on various contextual factors. These include, but are not limited to, the context of **tasks** performed during studies (such as whether the teamwork settings require domain-specific knowledge and/or involve high-risk, high-stake scenarios [7, 11, 18]), **devices** used for interaction with agents (e.g., [33] found individuals held

highly different mental models for agents with interconnected systems; therefore, users may view Siri either as the same smart assistant across all devices or as various individual entities), and **modalities** of communication with agents (for instance, [14, 15, 26, 30] studied the impact of embodiment through the voice of agents).

3.2 Anthropomorphism and perceived agency

We further examined whether embodying in humanoid features enhance users' trust toward agents (i.e., whether human-likeness elicit trust) and found several studies suggesting humanoid embodiment, in fact, harms trust for agents in teamwork [7, 26, 31], with exceptions for voice agents [4, 5]. To be specific, recent work has found that participants showed a significantly higher degree of trust for agents with human-like voices than those with synthetic voices. These relevant studies offered at least two possible explanations: The first is based on the Uncanny Valley, suggesting that individuals may have distrust a highly embodied agent if it is very human-like, but not quite to the extent that makes users comfortable (i.e., the level of anthropomorphism is at the low point of the Uncanny Valley). Second, perhaps users do not simply trust humanoid agents because they are like humans, but because highly humanoid features may imply the technical sophistication of an intelligent agent (e.g., an agent with a human voice is perceived as more "high tech"). This resonates with Sundar's [32] theoretical model for human-AI interaction (HAII-TIME), suggesting that users may view AI as a cue, while an AI system that can "think and act" like humans is often perceived with greater machine agency. In other words, a mere pursuit of anthropomorphism may not necessarily be effective on trust-building in human-agent interaction, but perhaps, communicating technical capacity through humanoid cues can be a promising means.

4 UNRESOLVED QUESTIONS AND CHALLENGES

Informed by the key learnings we have obtained so far, in this last section, we summarize additional inquiries which we either attempt to address through further coding the pool of literature in our present review or identify as research gaps in existing work, informing avenues for future research.

- **Context of tasks:** As mentioned above, a single factor of embodiment or anthropomorphism can not effectively explain users' trust for intelligent agents. Rather, their perception is often dependent on situated context. In our review, we will continue to systematically examine and categorize the various tasks and scenarios that have been investigated in previous work. Besides the level of risks involved in team scenarios, we will further examine the circumstances when either users have compatible knowledge with agents or not. Recent work in explainable AI has found the effectiveness of explainability may depend on the difference in knowledge levels of humans and of AI [2]. We ask whether trust-building can be more challenging when users are equally knowledgeable on a task topic.
- **Affordances of interaction:** Along the same vein, the medium through which users communicate with intelligent agents serves as yet another contextual factor in human-AI teams. Different devices and platforms set various levels of affordances and modalities for users to communicate, interact, and evaluate their conversational partners, which can directly or indirectly influence the extent to which they trust their machine teammates. Beyond comparing agents with visual-only, audio-only, or combined visual-audio cues, virtual agents in mixed reality platforms (e.g., virtual and/or augmented reality) can embody through a wide variety of non-verbal cues as well (e.g., gestures, movements). The impact of physical embodiment through multi-modal cues on users' trust toward agents remain an active and intriguing research area.
- **Group size:** A recent HRI review [29] has pointed out that studies examining the use of robots beyond one-on-one interaction (i.e., one robot vs. one human) remain scarce. Similarly, as we conduct the current review, we have encountered very few studies assessing teams including multi-humans and/or multi-agents, despite these scenarios may better represent the reality [23]. Among one of the very few examples, [30] showed that human behaviors were less affected by agents when they were with other humans. In

this regard, the presence of others is likely to introduce an additional layer of complexity to trust in a human-agent team, and we encourage future work to attempt this challenging problem space.

- **Trust in the long run:** While building trust among humans takes time and frequent interactions, the majority of existing research only investigates trust for embodied agents in one-off settings. While conducting longitudinal studies can be highly demanding for the time and resources of researchers, as an alternative, we propose that future empirical studies can at least incorporate multiple team interactions in their study design to evaluate users' perception and behaviors when they encounter agents in subsequent teamwork.
- **Single out the effect of trust:** While we discuss the various pitfalls in how trust is studied in some of the previous work, we also see the importance to consider and measure individual differences in order to account for their effect on trust. For instance, as users' personalities can play a key role in their tendency to comply, it is critical that future research include measures of personality, so that we can examine their moderating effect separately from the impact of embodiment and the design of agents as such.
- **User expertise:** Last but not least, as it is commonly believed that increased familiarity with technology would change one's behaviors [8], we are curious whether users' trust for (embodied) agents have increased/decreased as they have become more and more robot- or AI-savvy, given that as they now have more opportunities to interact with these agents through consumer products.

5 CONCLUDING THOUGHTS

As we review relevant literature on this present topic, we see that the complex, context-dependent role of embodiment reveals the constantly variant nature in this research area. Not only has the technical capacity of intelligent agents advanced drastically in recent years, but users' viewers toward them may also change quickly in accordance. Together, this may suggest that what we have learned from interactions among human-human teams may not necessarily be applicable to understanding human-agent teamwork. As sophisticated users gradually consider agents more as tools rather than social interactants [8], we may depart from previous research paradigms in seek of new frameworks specifically dedicated to understanding human-machine interaction.

REFERENCES

[1] Nava Ashraf, Iris Bohnet, and Nikita Piankov. 2006. Decomposing trust and trustworthiness. *Experimental economics* 9, 3 (2006), 193–208.

[2] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. 2021. Does the whole exceed its parts? the effect of ai explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

[3] Ray Bull and Nichola Rumsey. 2012. *The social psychology of facial appearance.* Springer Science & Business Media.

[4] Emna Cherif and Jean-François Lemoine. 2014. The Impact of Recommendation Agents' Type of Voice on Perceived Social Presence, Trust and Users Intentions on an Insurance Website. *Trends in Practical Applications of Heterogeneous Multi-agent Systems. The PAAMS Collection* (2014), 139–148.

[5] Emna Cherif and Jean-François Lemoine. 2017. Human vs. Synthetic Recommendation Agents' Voice: The Effects on Consumer Reactions. In *Marketing at the Confluence between Entertainment and Analytics*, Patricia Rossi (Ed.). Springer International Publishing, Cham, 301–310.

[6] Meia Chita-Tegmark, Theresa Law, Nicholas Rabb, and Matthias Scheutz. 2021. Can you trust your trust measure?. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. 92–100.

[7] Filipa Correia, Samuel Gomes, Samuel Mascarenhas, Francisco S Melo, and Ana Paiva. 2020. The dark side of embodiment-teaming up with robots vs disembodied agents. *Proceedings of Robotics: Science and Systems. Corvalis, Oregon, USA* (2020).

[8] Chad Edwards, Autumn Edwards, Brett Stoll, Xialing Lin, and Noelle Massey. 2019. Evaluations of an artificial intelligence instructor's voice: Social Identity Theory in human-robot interactions. *Computers in Human Behavior* 90 (2019), 357–362.

[9] Anthony M Evans and Joachim I Krueger. 2009. The psychology (and economics) of trust. *Social and Personality Psychology Compass* 3, 6 (2009), 1003–1017.

[10] Kerstin S Haring, Kelly M Satterfield, Chad C Tossell, Ewart J De Visser, Joseph R Lyons, Vincent F Mancuso, Victor S Finomore, and Gregory J Funke. 2021. Robot authority in human-robot teaming: Effects of human-likeness and physical embodiment on compliance. *Frontiers in Psychology* 12 (2021).

[11] Sarita Herse, Jonathan Vitale, Meg Tonkin, Daniel Ebrahimian, Suman Ojha, Benjamin Johnston, William Judge, and Mary-Anne Williams. 2018. Do you trust me, blindly? Factors influencing trust towards a robot recommender system. In *2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 7–14.
[12] Martin Hoegl and Hans Georg Gemuenden. 2001. Teamwork quality and the success of innovative projects: A theoretical concept and empirical evidence. *Organization science* 12, 4 (2001), 435–449.
[13] Younbo Jung and Kwan Min Lee. 2004. Effects of physical embodiment on social presence of social robots. *Proceedings of PRESENCE* (2004), 80–87.
[14] Kangsoo Kim, Luke Boelling, Steffen Haesler, Jeremy Bailenson, Gerd Bruder, and Greg F Welch. 2018. Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 105–114.
[15] Dimosthenis Kontogiorgos, Sanne van Waveren, Olle Wallberg, Andre Pereira, Iolanda Leite, and Joakim Gustafson. 2020. Embodiment effects in interactions with failing robots. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–14.
[16] ES Kox, JH Kerstholt, TF Hueting, and PW De Vries. 2021. Trust repair in human-agent teams: the effectiveness of explanations and expressing regret. *Autonomous Agents and Multi-Agent Systems* 35, 2 (2021), 1–20.
[17] Philipp Kulms and Stefan Kopp. 2016. The effect of embodiment and competence on trust and cooperation in human–agent interaction. In *International Conference on Intelligent Virtual Agents*. Springer, 75–84.
[18] Michal Luria, Samantha Reig, Xiang Zhi Tan, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. 2019. Re-Embodiment and Co-Embodiment: Exploration of social presence for robots and conversational agents. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 633–644.
[19] Bertram F Malle and Daniel Ullman. 2021. A multidimensional conception and measure of human-robot trust. In *Trust in Human-Robot Interaction*. Elsevier, 3–25.
[20] David Matsumoto, Mark G Frank, and Hyi Sung Hwang. 2012. *Nonverbal communication: Science and applications*. Sage Publications.
[21] Brian E Mennecke, Janea L Triplett, Lesya M Hassall, Zayira Jordán Conde, and Rex Heer. 2011. An examination of a theory of embodied social presence in virtual worlds. *Decision Sciences* 42, 2 (2011), 413–450.
[22] Dongfang Niu, Jacques Terken, and Berry Eggen. 2018. Anthropomorphizing information to enhance trust in autonomous vehicles. *Human Factors and Ergonomics in Manufacturing & Service Industries* 28, 6 (2018), 352–359.
[23] Anastasia K Ostrowski, Jenny Fu, Vasiliki Zygouras, Hae Won Park, and Cynthia Breazeal. 2021. Speed Dating with Voice User Interfaces: Understanding How Families Interact and Perceive Voice User Interfaces in a Group Setting. *Frontiers in Robotics and AI* 8 (2021), 730992–730992.
[24] Raul Paradeda, Mojgan Hashemian, Carla Guerra, Rui Prada, João Dias, and Ana Paiva. 2017. Fides: How emotions and small talks may influence trust in an embodied vs. non-embodied robot. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. 1673–1675.
[25] Summer Rebensky, Kendall Carmody, Cherrise Ficke, Daniel Nguyen, Meredith Carroll, Jessica Wildman, and Amanda Thayer. 2021. Whoops! Something Went Wrong: Errors, Trust, and Trust Repair Strategies in Human Agent Teaming. In *International Conference on Human-Computer Interaction*. Springer, 95–106.
[26] Samantha Reig, Jodi Forlizzi, and Aaron Steinfeld. 2019. Leveraging robot embodiment to facilitate trust and smoothness. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 742–744.
[27] Harry T Reis, Ladd Wheeler, Nancy Spiegel, Michael H Kernis, John Nezlek, and Michael Perri. 1982. Physical attractiveness in social interaction: II. Why does appearance affect social experience? *Journal of Personality and Social Psychology* 43, 5 (1982), 979.
[28] Kazuki Sakai, Yutaka Nakamura, Yuichiro Yoshikawa, and Hiroshi Ishiguro. 2021. Effect of Robot Embodiment on Satisfaction With Recommendations in Shopping Malls. *IEEE Robotics and Automation Letters* 7, 1 (2021), 366–372.
[29] Sarah Sebo, Brett Stoll, Brian Scassellati, and Malte F Jung. 2020. Robots in groups and teams: a literature review. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–36.
[30] Ameneh Shamekhi, Q Vera Liao, Dakuo Wang, Rachel KE Bellamy, and Thomas Erickson. 2018. Face Value? Exploring the effects of embodiment for a group facilitation agent. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.
[31] Rebecca Stower, Natalia Calvo-Barajas, Ginevra Castellano, and Arvid Kappas. 2021. A meta-analysis on children's trust in social robots. *International Journal of Social Robotics* 13, 8 (2021), 1979–2001.
[32] S Shyam Sundar. 2020. Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII). *Journal of Computer-Mediated Communication* 25, 1 (2020), 74–88.
[33] Nathan L Tenhundfeld, Hannah M Barr, HO Emily, and Kristin Weger. 2021. Is My Siri the Same as Your Siri? An Exploration of Users' Mental Model of Virtual Personal Assistants, Implications for Trust. *IEEE Transactions on Human-Machine Systems* (2021).
[34] Anouk van Maris, Hagen Lehmann, Lorenzo Natale, and Beata Grzyb. 2017. The influence of a robot's embodiment on trust: A longitudinal study. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. 313–314.
[35] Kai Vogeley and Gary Bente. 2010. "Artificial humans": Psychology and neuroscience perspectives on embodiment and nonverbal communication. *Neural Networks* 23, 8-9 (2010), 1077–1090.

[36] Tom Williams, Daniel Ayers, Camille Kaufman, Jon Serrano, and Sayanti Roy. 2021. Deconstructed Trustee Theory: Disentangling Trust in Body and Identity in Multi-Robot Distributed Systems. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. 262–271.