

Trust Interplay: Robot Performance Influences Cognitive but not Affective Trust

Adi Manor

Media Innovation Lab
Faculty of Data and Decision Sciences
Reichman University; Technion, Israel
adi.manor@milab.idc.ac.il

Avi Parush

Faculty of Data and Decision Sciences
Technion, Israel
aparush@technion.ac.il

Hadas Erel

Media Innovation Lab
Reichman University, Israel
hadas.erel@milab.idc.ac.il

Abstract—Cognitive and affective trust are fundamental elements in human-robot interactions. Previous research suggests that robots' affective behaviors related to affective trust can also impact cognitive trust. In this work, we explored whether the opposite influence also exists and whether robots' cognitive behaviors related to cognitive trust can also impact affective trust. Our results revealed that the robot's cognitive capabilities significantly impacted cognitive trust but did not influence affective trust. Participants in the *High Competence Robot* condition reported higher cognitive trust scores and were more likely to trust the robot's suggestions compared to the *Low Competence Robot* and *Baseline* conditions. We did not observe any differences in affective trust, and almost all participants in all conditions reported low affective trust. We suggest that while robot cognitive performance can effectively build cognitive trust, comprehensive trust development may require affective-related behaviors.

Index Terms—Cognitive trust, affective trust, non-humanoid robot, robot performance.

I. INTRODUCTION

Robots are becoming common, demonstrating a potential to impact various aspects of our daily lives [1]. A well-known aspect that plays a role in shaping our interactions with robots is trust [2]. Trust can foster a positive experience, making individuals feel more comfortable, satisfied, and engaged when interacting with robots [3]. It also enhances robots' acceptance as collaborative partners, reducing the need for constant monitoring and allowing for fluent interactions [4], [5]. When interactions with robots are based on trust, they can support more than just functional needs. Robots can provide companionship, address emotional needs, and even facilitate human interactions [6]–[13].

Trust, a concept extensively studied in psychology and other social sciences, has been traditionally divided into two main types: Cognitive Trust (CT) and Affective Trust (AT) [14], [15]. This fundamental distinction has been applied to human-robot interactions, where CT is based on rational judgments and objective evaluations of the robot's performance, relying on factual knowledge, past performance, and observable behaviors [3], [4]. Factors such as the robot's competence, reliability, and performance are at the basis of CT [2], [5]. CT was shown to influence task management, reliance on the robot's advice, and risk assessment [2], [5], [16]. In contrast, AT is based on feelings of closeness and empathy [14], [15].



Fig. 1. A robot performing a cognitive task next to the participant

When AT is present, individuals feel comfortable sharing their thoughts and emotions with the robot, and the interaction is characterized by openness and attentiveness [17]. AT facilitates the perception of the robot as a partner and an entity that will provide support when needed [1], [5], [18], [19]. AT in HRI can enhance satisfaction, cooperation, and overall engagement in the interaction [5], [19], [20]. While most of the research on trust in HRI has focused on CT [1], [2], there is growing interest in AT due to its significant impact on the quality and nature of human-robot interactions [14], [19], [20].

One aspect of trust that was less studied in the context of Human-Robot Interactions (HRI) is the interplay between the two types of trust. Studies from the psychology domain suggest that CT and AT are not mutually exclusive and may also impact one another [14], [15]. Interestingly, there are indications that some level of AT is necessary for CT to develop [15], [21] and when AT is established, CT is more likely to occur [21]. Trust is often viewed as involving an emotional attitude [22], and accordingly, AT sets the basis for CT. When people feel secure and safe (i.e., high AT), they are more likely to perceive the other person as reliable and trustworthy because their emotional needs are consistently met [21]. Similarly, research suggests that CT can significantly influence the development and maintenance of AT [23], [24]. When individuals consistently demonstrate competence, reliability, and sound judgment (i.e., high CT), AT can thrive. The predictability and dependability associated with high CT creates a sense of security that facilitates the development of emotional bonds [23], [24]. These examples suggest that while CT and AT are distinct constructs, they have a dynamic,

reciprocal relationship where each form of trust can reinforce and strengthen the other [24], [25].

Given that trust is an important aspect of HRI, it's important to investigate whether a similar interplay exists with robots. There are already indications that AT impacts CT in HRI. Studies with robots of different complexity have demonstrated how robotic behaviors known to impact AT can also influence CT. Specifically, robotic listening and attentive behaviors indicating that one can emotionally rely on the robot were also shown to enhance CT [17], [19]. Interestingly, the opposite possibility of AT being impacted by a robot's capabilities (i.e., CT) was not explored.

To further understand the potential interplay between CT and AT in HRI, we tested the possibility that a robot's cognitive capabilities would impact both CT and AT (despite the lack of any manipulation related to AT). Specifically, we examined whether high cognitive performance would enhance both CT and AT, while low cognitive performance would decrease both trust dimensions. We conducted an experiment involving a simple robotic object resembling a desk lamp (used with permission [26], see Figure 1). Participants were asked to perform a cognitively demanding activity alongside a robot that was presented as being trained to solve such tasks. The robot scanned the screen where the items were presented and indicated its response by pressing one out of four buttons that corresponded with the possible answers presented on the screen. We assessed participants' CT and AT tendencies toward the robot in three conditions: *High Competence Robot* condition that indicated correct answers, *Low Competence Robot* condition that indicated incorrect answers, and *Baseline* condition that the robot provided no answer.

II. RELATED WORK

Relevant previous studies include HRI studies that explored cognitive and affective trust.

A. Cognitive Trust in HRI

Multiple studies have identified various factors influencing CT development in human-robot relationships [5], [27], with robot performance and reliability emerging as primary factors [5], [28]–[30]. Both cognitive errors (e.g., memory lapses) and physical mistakes (e.g., navigation inaccuracies), can significantly decrease CT levels [31]. A notable finding by Xu and Howard (2018) demonstrated the importance of robots' performance at the beginning of the interaction. Robots that performed a task well maintained high CT throughout subsequent interactions, while initial errors led to persistent decreases in CT [28]. Similarly, Desai et al. (2012) showed that consistency in performance (even imperfect performance) has a great impact on CT, with higher consistency leading to higher performance. Although these studies demonstrate the impact of a robot's performance on CT, they did not explore its possible impact on AT.

B. Affective Trust in HRI

Previous studies indicated that the robot's behavior towards the participant has the greatest impact on AT [32].

For instance, Leite et al. (2013) found that robots demonstrating empathetic behaviors and emotional responsiveness fostered stronger AT, particularly in long-term interactions with children [33]. Anzabi and Umemuro (2023) explored how active listening (demonstrating attention and validating the speaker's emotions and experiences) and empathetic listening (incorporating empathetic gestures into verbal behavior) by a humanoid robot influenced AT tendencies. They showed that these distinct listening behaviors impact and shape AT [19]. Interestingly, Manor et al. (2024) discovered that robotic attentiveness can foster AT even with very simple robots. When a lamp-like robot exhibited attentive behavior, participants reported feeling closer to it and provided high AT ratings [17]. While these studies demonstrate the impact of a robot's attentive behaviors on AT, they did not explore the possibility that CT would impact AT.

C. The interplay between Affective and Cognitive Trust in HRI

Previous studies have also provided initial indications that AT-related factors can impact CT. Anzabi and Umemuro (2023) showed that robot empathetic listening could impact trust formation [19]. When the robot's behavior indicated interest and empathy, both AT and CT increased. They demonstrated how empathetic listening behaviors can enhance both AT and CT. Similarly, Manor et al. (2024) found that a simple non-humanoid robot, performing attentive non-verbal gestures related to AT, can enhance CT. Participants in their study explained that the robot's attentiveness implied that it was also competent and capable of assisting with complex tasks [17]. This surprising result raised the need for further investigation into the opposite direction and testing whether perceptions of a robot's competence and performance can influence affective trust.

III. METHOD

We designed an interaction with a simple robotic object to test whether different robotic cognitive capabilities would impact CT and AT. Participants were asked to complete a cognitively demanding computerized task next to a simple robot that was presented as being trained to solve the same task. At each trial, the robot scanned the task item presented on the screen and presented its answer by pressing one out of four buttons that corresponded with the four possible answers presented on the screen (see Figure 1). The robot's cognitive capabilities were manipulated in three conditions: In the *High Competence Robot* condition, the robot always chose the correct answers; In the *Low Competence Robot* condition, the robot mostly indicated incorrect answers; In the *Baseline* condition, the robot scanned the task items on the screen but did not indicate any answer. The study was approved by the university's ethics committee.

A. The Task

We used the Raven Progressive Matrices (RPM) task, a well-known task that measures abstract reasoning abilities [34]. The RPM task involves visual items, each presenting a

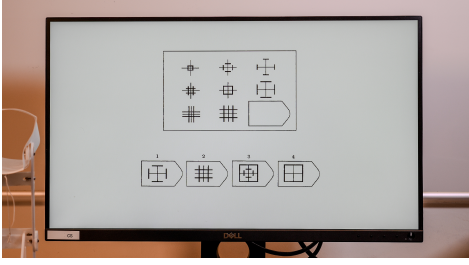


Fig. 2. RPM item presented on the screen

3x3 matrix of shapes with one shape missing. Participants are asked to identify the pattern governing the matrix and select the appropriate missing shape from several alternative options (see Figure 2). Traditionally, RPM task items present eight possible answers (options for the missing shape). We adapted the task to our study and presented only four possible answers for each item to ensure that participants could clearly notice the robot's answers. Additionally, to mitigate the influence of fatigue, we used a partial selection of the items typically used in the RPM (12 instead of 18). The progressive increase in the difficulty levels of the items allowed for flexibility in designing the robot's performance. At each trial, after the robot and participant provided an answer a feedback was presented (Accurate/Inaccurate) indicating separately the accuracy of the robot's answer and the participant's answer (except for the baseline condition which included only the participant's feedback).

B. Participants

30 undergraduate students from the university participated in the study (21 females, 9 males; $M = 24.1$, $SD = 2.3$). All participants completed an informed consent form and received course credits for their participation.

C. Experimental settings

The experiment took place in a quiet room. The robot was placed on the table, approximately 55 cm to the left of the participant, with its upper part directed toward the screen. The robot was powered by the Butter Robotics MAS platform [35]. Pre-scripted robot commands were executed by a web-based interface. A small camera was placed in the room.

D. Experimental design

The between-participant experimental design included three conditions. The robot's competence was designed as follows:

High Competence Robot condition - The robot chose the correct answer for all 12 items.

Low Competence Robot condition - The robot mostly chose incorrect answers (items 4-6,8-12). To prevent participants from perceiving the robot as malfunctioning (which may lead to low trust overall, [5]), the robot provided correct answers to the simpler items (first three items) and the seventh item.

Baseline condition - The robot only scanned the screen but did not choose any answer throughout the task.

Participants were *randomly* assigned to conditions using a matching technique balancing gender and Negative Attitudes Toward Robots (NARS) [36] to avoid a-priori differences.

E. Measures

We used two measures to assess CT and AT.

1) *Affective and Cognitive trust scale*: Participants completed the Anzabi and Umemuro's (2023) adaptation of the McAllister's [14] trust scale. The seven-point Likert scale (1 - "strongly disagree", 7 - "strongly agree") includes nine items measuring CT and seven items measuring AT.

2) *Open questions*: To further evaluate cognitive and affective trust, at the end of the experiment, the researcher conducted a short interview and introduced two direct questions: (1) "Would you trust the robot's suggested solution to a difficult task?"; (2) "Would you trust the robot to be there for you and support you when needed?".

F. Procedure

Participants received the NARS and Demographic questionnaires by mail. Upon arrival at the lab, they were asked to perform a cognitive task next to a robot that was being trained to perform the task. Participants were directed to the experiment room and were asked to sit in front of a computer screen, with the robot next to them. They were then left to complete the 12 RPM items, with the robot performing it next to them, activated according to the relevant condition. When participants completed the RPM they completed the questionnaires and answered the researcher's direct questions. Finally, the researcher debriefed the participants and ensured they left with an overall positive experience.

IV. FINDINGS

The robot's competence impacted CT but not AT.

1) *Cognitive and Affective trust scale*: The one-way ANOVA analysis, performed for the CT and AT scales, revealed an effect for the CT scale, $F(2, 27) = 5.39$, $p = 0.01$, $\eta_p^2 = 0.342$. Post-hoc multiple comparisons using Scheffe's method indicated that participants' CT scores in the *High Competence Robot* condition were higher than those in the *Low Competence Robot* condition, $p = 0.004$ (see Figure 3, Left). No differences were found for AT, $F(2, 27) = 0.5$, $p = 0.61$, $\eta_p^2 = 0.004$ (see Figure 3, Right).

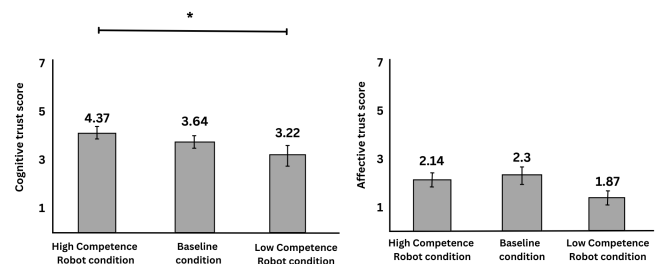


Fig. 3. The CT scale scores (Left) and AT scale scores (Right)

TABLE I
DISTRIBUTION OF PARTICIPANTS' CT TENDENCIES IN DIFFERENT CONDITIONS

Condition	Cognitive Trust Tendencies	
	Yes	No
High Competence Robot	7	3
Baseline	2	8
Low Competence Robot	1	9

2) *CT and AT in the open questions*: Participants' responses to the direct questions were analyzed using a chi-square test of independence. The analysis was conducted separately for each question and included the frequency of yes/no answers across the different conditions. A significant dependence was found between CT and the robotic conditions, $\chi^2(2) = 9.3$, $p < 0.01$. Participants in the *High Competence Robot* condition trusted the robot's cognitive abilities more than participants in the *Low Competence Robot* and *Baseline* conditions (see Table I). In the AT question, there was no dependency, and almost all participants across all conditions reported no affective trust, $\chi^2(2) = 2.07$, $p = 0.35$ (see Table II).

V. DISCUSSION

Our preliminary findings demonstrated that a robot's cognitive performance impacts CT and not AT, indicating an asymmetric interplay between the constructs. While previous studies found that AT-related robotic behaviors can impact both AT and CT, we did not find an indication of an impact in the opposite direction. The CT-related robotic behavior (i.e., cognitive performance) impacted only CT and did not influence AT. Participants who interacted with a highly competent robot reported significantly higher CT than those with a low-competence robot. AT was similar regardless of the robot's competence. While we acknowledge that the lack of effect may result from a floor effect, this is highly unlikely, as previous studies with the same robot indicated that the robot's attentive behavior can significantly increase participants' AT [17].

Our findings suggest that even when robots demonstrate high competence, the pathways to developing emotional bonds and AT may require explicit social and affective behaviors. The mere demonstration of capability is insufficient for developing social-emotional trust. The symmetry between CT and AT commonly observed in human interactions is not as direct in HRI. While people use the robot's cognitive capabilities to adjust their CT, they do not find this information relevant for adjusting their AT. However, they do find a robot's affective behavior relevant and informative for adjusting their CT (see [17], [19]). One possible explanation for this asymmetric interplay between AT and CT may relate to people's expectations. Expectations from robots were suggested as one of the main factors determining the robot's social influence [37]. It is possible that since people have high expectations from a robot's cognitive capabilities [38]–[40] its performance has a local impact that does not extend to other domains. However, people

TABLE II
DISTRIBUTION OF PARTICIPANTS' AT TENDENCIES IN DIFFERENT CONDITIONS

Condition	Affective Trust Tendencies	
	Yes	No
High Competence Robot	1	9
Baseline	0	10
Low Competence Robot	0	10

do not expect a robot (which is considered a "machine") to be attentive and caring. Thus, when a robot exceeds expectations by demonstrating AT-related behaviors, its impact extends to other domains and influences CT. This asymmetric interplay between CT and AT, which differs from what is typically observed in human-human interactions [14], [15] supports the more general need to map robots' social impact, as it cannot be directly inferred by human-human interactions [37]. The different expectations associated with interactions with robots may lead to social effects that should be well understood and considered when designing interactions with robots.

We note that the lack of increase in AT can not be associated with a ceiling effect as AT was low in the baseline group. We intentionally chose a simple robot that would not be naturally associated with AT, allowing for testing if the robot's competence would increase it. Importantly, previous studies with the same robot clearly indicated that AT-related robotic behaviors (e.g., attentiveness) can drastically enhance AT, with participants reporting they would trust the simple robot to be there for them in times of need. We also acknowledge the limitation associated with the small sample size and the need for future evaluation of the lack of AT with more participants. However, the low effect size found in our study supports the possibility that AT is not sensitive to robots' competence.

Another aspect of our findings concerns the impact of the robot's competence on CT. We extend the common finding of increased CT in interactions with competent robots [28], [31] and demonstrate that this strong relationship can be also applied to interactions with very simple robots that do not speak and communicate only via very simple gestures.

In sum, our findings provide an initial understanding of the complex relationship between the different trust constructs in HRI. While robots' competence can support CT even with simple robots, it does not seem to enhance AT. With previous research indicating that AT-related manipulations can impact both trust dimensions, our findings suggest that the great emphasis on CT in HRI studies may need re-balancing. The development of robots' social-emotional capabilities may offer more comprehensive benefits for human-robot trust, as these features can simultaneously foster both AT and CT, while robotic competence seems to impact CT alone.

REFERENCES

- [1] J. Han, J. Davids, H. Ashrafian, A. Darzi, D. S. Elson, and M. Sodergren, "A systematic review of robotic surgery: From supervised paradigms

- to fully autonomous robotic approaches,” *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 18, no. 2, p. e2358, 2022.
- [2] B. C. Kok and H. Soh, “Trust in robots: Challenges and opportunities,” *Current Robotics Reports*, vol. 1, pp. 297–309, 2020.
 - [3] S. Naneva, M. Sarda Gou, T. L. Webb, and T. J. Prescott, “A systematic review of attitudes, anxiety, acceptance, and trust towards social robots,” *International Journal of Social Robotics*, vol. 12, no. 6, pp. 1179–1201, 2020.
 - [4] J. J. Lee, B. Knox, J. Baumann, C. Breazeal, and D. DeSteno, “Computationally modeling interpersonal trust,” *Frontiers in psychology*, vol. 4, p. 56004, 2013.
 - [5] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, and R. Parasuraman, “A meta-analysis of factors affecting trust in human-robot interaction,” *Human factors*, vol. 53, no. 5, pp. 517–527, 2011.
 - [6] O. Zuckerman, D. Walker, A. Grishko, T. Moran, C. Levy, B. Lisak, I. Y. Wald, and H. Erel, “Companionship is not a function: The effect of a novel robotic object on healthy older adults’ feelings of ‘being seen’,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–14.
 - [7] N. Massa, P. Bisconti, and D. Nardi, “The psychological implications of companion robots: a theoretical framework and an experimental setup,” *International Journal of Social Robotics*, vol. 15, no. 12, pp. 2101–2114, 2023.
 - [8] T. Shibata and K. Wada, “Robot therapy: a new approach for mental healthcare of the elderly—a mini-review,” *Gerontology*, vol. 57, no. 4, pp. 378–386, 2011.
 - [9] S. Gillet, M. Vázquez, S. Andrist, I. Leite, and S. Sebo, “Interaction-shaping robotics: Robots that influence interactions between other agents,” *ACM Transactions on Human-Robot Interaction*, 2024.
 - [10] H. Erel, D. Trayman, C. Levy, A. Manor, M. Mikulincer, and O. Zuckerman, “Enhancing emotional support: The effect of a robotic object on human-human support quality,” *International Journal of Social Robotics*, pp. 1–20, 2021.
 - [11] O. Sadka, A. Jacobi, A. Grishko, U. Lumnitz, B. Megidish, and H. Erel, ““by the way, what’s your name?”: The effect of robotic bar-stools on human-human opening-encounters,” in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 2022, pp. 1–6.
 - [12] O. Sadka, A. Parush, O. Zuckerman, and H. Erel, “All it takes is a slight rotation: Robotic bar-stools enhance intimacy in couples’ conflict,” in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–7.
 - [13] D. Rifinski, H. Erel, A. Feiner, G. Hoffman, and O. Zuckerman, “Human-human-robot interaction: Robotic object’s responsive gestures improve interpersonal evaluation in human interaction,” *Human-Computer Interaction*, vol. 36, no. 4, pp. 333–359, 2021.
 - [14] D. J. McAllister, “Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations,” *Academy of management journal*, vol. 38, no. 1, pp. 24–59, 1995.
 - [15] D. Johnson and K. Grayson, “Cognitive and affective trust in service relationships,” *Journal of Business research*, vol. 58, no. 4, pp. 500–507, 2005.
 - [16] F. Saleem, Y. Z. Zhang, C. Gopinath, and A. Adeel, “Impact of servant leadership on performance: The mediating role of affective and cognitive trust,” *Sage Open*, vol. 10, no. 1, p. 2158244019900562, 2020.
 - [17] A. Manor, A. Parush, and H. Erel, “Attentiveness: A key factor in fostering affective and cognitive trust with non-humanoid robots,” in *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*. IEEE, 2024, pp. 469–476.
 - [18] M. Lewis, K. Sycara, and P. Walker, “The role of trust in human-robot interaction,” *Foundations of trusted autonomy*, pp. 135–159, 2018.
 - [19] N. Anzabi and H. Umemuro, “Effect of different listening behaviors of social robots on perceived trust in human-robot interactions,” *International Journal of Social Robotics*, pp. 1–21, 2023.
 - [20] N. ANZABI and H. UMEMURO, “Influence of social robots’ benevolence and competence on perceived trust in human-robot interactions,” *The Japanese Journal of Ergonomics*, vol. 59, no. 6, pp. 258–273, 2023.
 - [21] P. Fonagy and E. Allison, “The role of mentalizing and epistemic trust in the therapeutic relationship,” *Psychotherapy*, vol. 51, no. 3, p. 372, 2014.
 - [22] J. Simon, *The Routledge handbook of trust and philosophy*. Routledge, 2020.
 - [23] J. D. Lewis and A. Weigert, “Trust as a social reality,” *Social forces*, vol. 63, no. 4, pp. 967–985, 1985.
 - [24] R. C. Mayer, J. H. Davis, and F. D. Schoorman, “An integrative model of organizational trust,” *Academy of Management Review*, vol. 20, no. 3, pp. 709–734, 1995.
 - [25] D. M. Rousseau, S. B. Sitkin, R. S. Burt, and C. Camerer, “Not so different after all: A cross-discipline view of trust,” *Academy of Management Review*, vol. 23, no. 3, pp. 393–404, 1998.
 - [26] G. Hoffman, O. Zuckerman, G. Hirschberger, M. Luria, and T. Shani-Sherman, “Design and evaluation of a peripheral robotic conversation companion,” in *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2015, pp. 3–10.
 - [27] T. Gompei and H. Umemuro, “Factors and development of cognitive and affective trust on social robots,” in *Social Robotics: 10th International Conference, ICSR 2018, Qingdao, China, November 28-30, 2018, Proceedings 10*. Springer, 2018, pp. 45–54.
 - [28] J. Xu and A. Howard, “The impact of first impressions on human-robot trust during problem-solving scenarios,” in *2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 2018, pp. 435–441.
 - [29] B. Nesses, D. A. Robb, J. Lopes, and H. Hastie, “Transparency in hri: Trust and decision making in the face of robot errors,” in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021, pp. 313–317.
 - [30] N. Salomons, M. Van Der Linden, S. Strohkorb Sebo, and B. Scassellati, “Humans conform to robots: Disambiguating trust, truth, and conformity,” in *Proceedings of the 2018 acm/ieee international conference on human-robot interaction*, 2018, pp. 187–195.
 - [31] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, “Would you trust a (faulty) robot? effects of error, task type and personality on human-robot cooperation and trust,” in *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*, 2015, pp. 141–148.
 - [32] K. M. Lee, Y. Jung, J. Kim, and S. R. Kim, “Designing for social presence in online communities,” *ACM Transactions on Computer-Human Interaction*, vol. 11, no. 4, pp. 481–500, 2004.
 - [33] I. Leite, C. Martinho, and A. Paiva, “Social robots for long-term interaction: A survey,” *International Journal of Social Robotics*, vol. 5, no. 2, pp. 291–308, 2013.
 - [34] J. C. Raven and J. H. Court, *Raven’s progressive matrices and vocabulary scales*. Oxford Psychologists Press Oxford, 1998.
 - [35] B. Megidish. (2017) Butter robotics. [Online]. Available: <https://butter-robotics.web.app/>
 - [36] T. Nomura, T. Kanda, and T. Suzuki, “Experimental investigation into influence of negative attitudes toward robots in human-robot interaction,” *AI & Society*, vol. 20, no. 2, pp. 138–150, 2006.
 - [37] H. Erel, M. Vázquez, S. Sebo, N. Salomons, S. Gillet, and B. Scassellati, “Rosi: A model for predicting robot social influence,” *ACM Transactions on Human-Robot Interaction*, 2024.
 - [38] J. DelPreto, J. I. Lipton, L. Sanneman, A. J. Fay, C. Fourie, C. Choi, and D. Rus, “Helping robots learn: a human-robot master-apprentice model using demonstrations via virtual reality teleoperation,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 10 226–10 233.
 - [39] P. Robinette, W. Li, R. Allen, A. M. Howard, and A. R. Wagner, “Overtrust of robots in emergency evacuation scenarios,” in *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)*. IEEE, 2016, pp. 101–108.
 - [40] A. Sharkey and N. Sharkey, “We need to talk about deception in social robotics!” *Ethics and Information Technology*, vol. 23, pp. 309–316, 2021.