"*What's on your mind?*": Understanding the Development of Multidimensional Trust in Social Robots

Chih-Wei (Charlotte) Ning
Delf University of Technology
Delft, The Netherlands
c.w.ning@tudelft.nl

Carolina Centeio Jorge
Delf University of Technology
Delft, The Netherlands
c.jorge@tudelft.nl

Myrthe L. Tielman
Delf University of Technology
Delft, The Netherlands
m.l.tielman@tudelft.nl

Mark A. Neerincx
Delf University of Technology
Delft, The Netherlands
m.a.neerincx@tudelft.nl

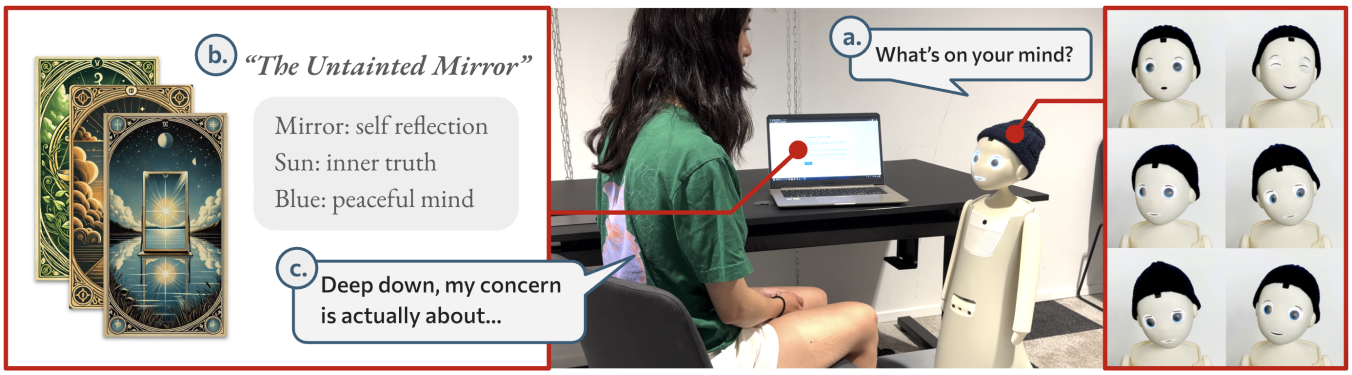

Figure 1: The setting of the Card Divination Task. (a) The robot prompts the user to talk about a personal topic. (b) The user draws a card on the screen, which contains visual symbols to be explained. (c) Based on the card content and the robot's guidance, the user discloses and reflects about their concern.

Abstract

As robots and virtual agents are increasingly envisioned as long-term companions, understanding how trust develops becomes crucial for ensuring safe and appropriate human-robot relationships. This research investigates how affective and cognitive trust evolve in social human-robot interactions. Participants (n=40) engaged in a 2 (social attitude: social, baseline) × 3 (time: t_1, t_2, t_3) mixed-design user study with a social robot, using a novel *Card Divination Task* developed to elicit both cognitive and affective trust dimensions. Results show that cognitive trust develops early while affective trust emerges gradually. Moreover, social cues enhance both cognitive trust, affective trust, and participants' certainty in trust judgment. These findings provide empirical support for the theoretical distinction between trust dimensions and highlight the role of social behavior in shaping trust over repeated interactions.

CCS Concepts

• **Human-centered computing** → **Empirical studies in collaborative and social computing**.

Keywords

Trust, Affective Computing, Collaborative Robots, Social Robots, Humanoid Robots, Human-Robot Interaction, Social Intelligence, Emotional Responses, Trust Management

ACM Reference Format:
Chih-Wei (Charlotte) Ning, Carolina Centeio Jorge, Myrthe L. Tielman, and Mark A. Neerincx. 2026. "*What's on your mind?*": Understanding the Development of Multidimensional Trust in Social Robots. In *Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction (HRI '26), March 16–19, 2026, Edinburgh, Scotland, UK*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3757279.3785556

1 Introduction

There is a paradigm shift in social robotics research, with robots increasingly expected to act as social actors rather than merely functional tools [46]. Advances in AI and large language models have accelerated this shift, making interactions with robots and agents more accessible and personalized, allowing them to evolve from one-time servants to lasting companions. Now, people turn to social agents not only for advice but also for emotional support [34],

HRI '26, Edinburgh, Scotland, UK

ACM ISBN 979-8-4007-2128-1/2026/03
https://doi.org/10.1145/3757279.3785556

which has been reported to reduce stress, anxiety, and loneliness [13, 37]. With the popularity of consumer products such as OpenAI GPT or xAI Companion, this brand-new form of "relationships" is becoming more and more ubiquitous.

In response to the potential emergence of such close, socially oriented relationships, we must ensure that users maintain an appropriate level of trust. In the HRI field, trust is often defined as *an attitude where the trustor feels positive about relying on the trustee, despite being in a situation characterized by uncertainty and vulnerability* [21]. Sufficient trust empowers one to build relationships and to make full use of a system's capabilities, whereas excessive trust introduces the risk of system misuse. By definition, to trust is to willingly place oneself in a potentially risky or dependent position. Within the context of emotionally significant interactions, this risk becomes even more pronounced and may lead to unhealthy attachment. The first step to addressing this concern is to understand how trust is formed and develops in such relationships. We approach this by investigating trust through two relevant perspectives: *multidimensional* and *developmental*.

First, the *multidimensional* lens allows an emphasis on the affective nature of emotionally meaningful relationships. Various multidimensional models (e.g., [10, 29, 30]) have been proposed to capture trust as a nuanced construct, recognizing not only ability-based aspects but also relational components. Yet, while the latter is particularly critical in the social interactions described above, relevant studies remain scarce in the HRI field [33]. We adopt McAllister's model [31] as our theoretical foundation, which explicitly captures the relational aspect through the dimension of *affective trust*: the belief that the trustee genuinely cares about and is emotionally close to the trustor. The other dimension, *cognitive trust*, reflects the belief of the trustee's competence and reliability, corresponding to ability-based aspects.

Second, as relationships and emotional bonds take time to form and strengthen [22, 24, 44], the *developmental* lens becomes a necessary consideration. Aligning with the increasing advocation in the research community [16, 33, 43], we view trust as a dynamic variable rather than a single static measurement, recognizing trust development as an ongoing calibration process shaped by accumulated experiences through repeated interactions [1, 20, 35].

We look into the intersection of the *multidimensional* and *developmental* perspectives of trust, an underexplored research gap in experimental HRI despite prior work in human-human trust (e.g., [24, 44]) and computational models (e.g., [8, 48]). To systematically induce and observe differences across multiple interactions and dimensions, we examine the effect of robot's *social attitude*, which corresponds to our relational focus and has been positively linked to both trust dimensions [14, 39, 41, 42].

This work is guided by the research question: **How does a robot's social attitude affect the development of cognitive and affective trust?** We hypothesize that affective trust emerges gradually through repeated interactions, whereas cognitive trust forms earlier through rational assessment. While this assumption is grounded in interpersonal literature (e.g., [24, 44]) and widely adopted in HRI computational models (e.g., [8, 48]), we aim to provide empirical HRI evidence to address potential differences between users' perceptions of robots and humans, as mentioned in various works (e.g., [2, 3, 11, 19]).

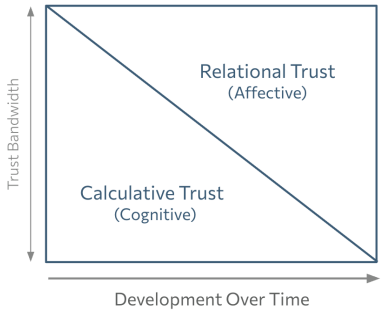

Figure 2: Rousseau et al.'s model [44] suggests that relational trust become more important overtime. (adapted from [24])

The remainder of this paper first reviews relevant literature and then introduces the *Card Divination Task*, a novel interaction designed to elicit both affective and cognitive trust. Next, we describe the user study, in which participants repeatedly interact with a social robot. Finally, we present the results and discussions.

This work contributes to HRI trust research by providing empirical evidence on the distinct development of affective and cognitive trust, as well as the influence of a robot's social attitude across repeated interactions. In addition, we introduce the Card Divination Task, a novel experimental paradigm designed to elicit emotionally meaningful interaction and both dimensions of trust. Together, our findings offer insights for future systems that monitor and calibrate trust, empowering humans in their interactions with robots.

2 Background

2.1 Trust and Multidimensional Models

Trust is a complex concept that has been widely studied across psychology [24], organization management [30], and human-technology interactions [17, 21]. It can be conceptualized as either an internal belief, a decision, or an intended behavior [10]. We follow previous HRI literature and regard trust as an attitude [4, 10, 21].

To account for trust's complex nature, various multidimensional models have been proposed. The Multi-Dimensional Measure of Trust questionnaire assesses trust through *performance-based* (reliable, competent) and *moral-based* dimensions (ethical, transparent, benevolent) [29], while the Socio-Cognitive Model identifies *competence* (can-do) and *willingness* (will-do) as core components [10]. These dimensions consistently cluster around two key aspects: ability-based and relational. Nevertheless, existing HRI literature have predominantly focused on the ability-based, cognitive aspects, with less attention paid to the relational, affective elements [32, 33, 47] despite their importance in forming close bonds [24, 28, 48]. We address this gap by centering emotional factors while still accounting for competence-related aspects.

2.2 Trust Development

Trust is recognized as developmental across disciplines. Psychological models often describe interpersonal trust as a staged process, where cognitive trust dominates in the early stages and affective trust emerges later, as the trustor gets to know the trustee better

over repeated and multifaceted interactions [22, 24]. Rousseau et al. [44] conceptualize trust as a fixed "bandwidth" shared by calculative and relational elements, with rational assessments gradually giving way to emotions (Figure 2). Lewicki et al. [23] indicate that trust maturation depends on interaction frequency, duration, and diversity of shared challenges, with frequency most commonly studied in short, repeated-measured experiments [24].

In human–machine interaction, the 3P model [21] characterizes trust as shaped by dynamic evaluations of performance, purpose, and process via analytic, analogical, and affective pathways. Other stage-based models, such as the three-layered [17] and three-stage frameworks [20], emphasize how trust may start as subjective, dispositional trust, but replaced by learned trust after actual encounters. These models were originally developed for non-social automation and are recently applied to social robots [1, 35]. However, none of these works focus on affective trust or emotionally meaningful interactions.

Another line of HRI research explores computational frameworks to model trust as a dynamic process. Guo and Yang [16] applied Bayesian inference to predict individual trust toward drones, while Ahmad et al. [1] extended the three-layered model [17] to estimate trust in social robots. Yet, such works either treat trust as a single construct or focus narrowly on ability-based trust. In contrast, Urbano et al. [48] and Deljoo et al. [8] both highlight the affective dimension (termed benevolent trust) in social contexts and propose additional social components in computational models. However, their assumption that affective trust develops only through repeated interactions, remains grounded in interpersonal literature and lacks empirical validation in HRI context.

2.3 Affective Trustworthiness, Benevolence, and Social Attitude

A trustor forms trust based on their estimation of the trustee's trustworthiness [30, 48]. With a focus on relational aspects of trust, we seek to manipulate the *affective trustworthiness* of the robot. Achieving this requires first identification of what robot behaviors are perceived as affectively trustworthy.

Among popular HRI trust models, *benevolence* from the Ability, Benevolence and Integrity (ABI) model [30] has been closely linked to affective trustworthiness and long-term, emotionally grounded relationships [4, 39, 48]. However, benevolence is defined as trustee's positive intention [4], which cannot be guaranteed to be perceived by the trustor. Therefore, in practical study designs, this intention is often operationalized as social attitude [4, 14].

Social attitude has been positively associated with both cognitive and affective trust [14, 31, 39, 41, 42], and is typically manipulated through both verbal and non-verbal cues. Non-verbal cues include gaze [4, 5, 46], touch [14], arm gestures [4], and facial expressions [46]. Verbal expressions often take the form of caring or empathetic statements [4, 14, 39, 42], socially oriented topics [5, 42], and memory-based personalization such as referring to the user by their name [4]. We followed these approaches and manipulated social attitude as the independent variable, aiming to observe its impact on multidimensional trust development.

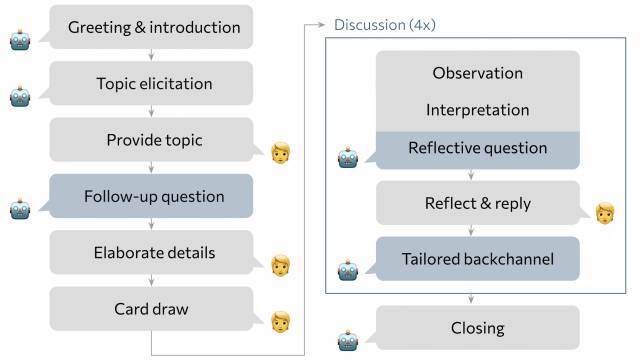

Figure 3: The conversation flow of each round of the card-divination task. Blue blocks are tailored by LLM.

3 Task

We developed a novel interaction task for the specific goal of this research, guided by the following requirements:

- **TR1. Social Settings:** Following the approach of [4, 14], we leverage one-on-one, face-to-face spoken conversations that incorporate social engagement.
- **TR2. Multiple Rounds:** To examine trust development over multiple encounters, the interaction must be a repeatable task with varying content for each round.
- **TR3. Emotional Vulnerability:** To elicit *affective trust*, we focus on emotional vulnerability, as expressing such vulnerability itself constitutes the risk that makes trust meaningful [21, 33]. We encourage self-disclosure in personal concerns as the primary mechanism.
- **TR4. Cognitive Competence:** To elicit *cognitive trust*, the robot demonstrates cognitive competence in terms of knowledge sharing and logical interpretation [42].

3.1 The Card Divination Task

Based on the requirements, we designed the *Card Divination Task*. The social robot Navel is framed as a "research assistant" collaborating with the user to explore a newfound divination deck. As illustrated in Figure 3, an interaction round starts with Navel greeting the user and inviting them to propose and elaborate on a personal topic, followed by the participant drawing a card. Next, the two parties discuss the symbolic content and possible implications of the card. Each robot turn consists of: (1) an observation about the visual symbols on the card, (2) an interpretation of the observation, and (3) a reflective question that either invites the user's opinion on the interpretation or encourages them to relate the interpretation to their situation. After each user response, the robot provides an LLM-generated backchannel tailored to the context.

As depicted in Figure 1, the interaction takes the form of face-to-face spoken conversations (**TR1**) and can be repeated with distinct cards and topics (**TR2**). Throughout the conversation, Navel constantly encourages the user to disclose personal experiences and concerns (**TR3**), while sharing knowledge about card symbolism (e.g., the mirror is relevant to self reflection in many cultures) and

Table 1: Example utterances illustrating how the robot's speech differs between social and baseline conditions. Each row corresponds to a design principle, with the same informational content conveyed at varying levels of social attitude.

Principle	Social Condition	Baseline Condition
Engaging pronouns	Now, let's look at the card together.	Now, please look at the card.
Acknowledgment	These are deep and courageous reflections, {name}!	Your reflections are noted.
Subjective opinion	The calm water makes me think of a peaceful mind.	The calm water may be associated with a peaceful mind.
Relational comment	Thank you for being my research buddy!	Thank you for your participation.
Cross-conversation memory	Speaking of balance: it's not just about {current topic}, but also about how it fits with {previous topics}.	Regarding balance, it may not just about {current topic}, but also about how different areas of your life interact.

logical interpretation that links to the user's situation (e.g., you might want to reflect on your inner desire) (**TR4**). In this case, trust in the robot relates to trusting it to give useful and truthful information about the divination ritual (cognitive trust), and to react emphatically to the disclosed vulnerability (affective trust).

Although users appear to "draw" a card at random, the sequence of cards and corresponding scripts are in fact fixed and identical for all participants. Only the connections between the card and the user's input are personalized. The downside is that trust measurement might be influenced by the specific card rather than the interaction order alone. However, we chose not to counterbalance the cards due to the infeasible number of participants required. Moreover, fixed card order allows a coherent storyline across participants. To attempt to distinguish trust in the card from trust in the robot, trust in the card was measured separately and explicitly next to trust in the robot. Finally, according to the *Barnum Effect* [9], people tend to perceive ambiguous interpretations as unique and personal [9, 12]. Therefore, a fixed set of stimuli can still offer meaningful interaction in a controlled experiment.

3.2 Conditions

The robot demonstrates different level of social attitude across two conditions. The robot displays vivid facial expressions and uses warm, supportive language in the *social* condition, actively engaging with the user and retaining cross-conversation memory by referring to their name or previous topics. In contrast, the *baseline* robot maintains a neutral face and communicates in a distant, objective tone without memory. Table 1 provides sample utterances illustrating the manipulation principles between groups.

3.3 Platform

We developed a custom software system with three components: (1) A central, local web server which manages the overall flow and natural language processing (OpenAI GPT-4[1]). (2) A web-based UI which provides instructions, visual stimuli, and interactive buttons. (3) The social robot "Navel" [46], which interacts with users through facial expressions and spoken dialogue, supported by speech recognition (Azure[2]) and text-to-speech (Navel SDK[3]) APIs.

[1]https://openai.com/index/gpt-4-research/
[2]https://azure.microsoft.com/en-us/products/ai-services/ai-speech
[3]https://doc.navelrobotics.com/

4 Methods

The Card Divination Task was used in a 2×3 mixed design study. The between-subject factor is the robot's social attitude (*social* vs. *baseline*). The within-subject factor is the repeated measure taken after each conversation rounds (time: t_1, t_2, t_3).

4.1 Hypotheses

We established the following hypotheses baes on literrature:

- **H(a):** Affective trust takes more interaction rounds to develop than cognitive trust.
- **H(b.1):** Social attitude has a positive effect on the cognitive trust in a social robot.
- **H(b.2):** Social attitude has a positive effect on the affective trust in a social robot.
- **H(b.3):** The positive effect of social attitude on affective trust emerges later comparing to that on cognitive trust.
- **H(c.1):** Users' certainty about their judgment in cognitive trust increases over multiple rounds.
- **H(c.2):** Users' certainty about their judgment in affective trust increases over multiple rounds.

4.2 Participants

A total of 40 participants (19 women, 20 men, 1 non-binary) were recruited through personal networks. Most of them were university students or recent graduates, aged 21-35. All participants demonstrated sufficient English proficiency to comfortably engage in spoken conversations, and were evenly assigned to two experimental groups based on demographics characteristics. The experiment protocol was reviewed and approved by the Human Research Ethics Committee of Delft University of Technology (ID: 5502).

4.3 Measurements

Measurements were collected at the start of the experiment, after each of the three rounds, and after the whole study.

4.3.1 *Demographics and Control Variable.*

- **Demographics:** We collected participants' gender, age, and prior experience with social robots and conversational agents (interaction frequency and typical contexts).
- **Paranormal Belief:** Participants' beliefs in supernatural phenomena are considered a potential influencing factor due to the task's divinative framing. For instance, a skeptical individual might report low trust only because they distrust any claims associated with "magic". We thus applied the

Precognition subscale of the Revised Paranormal Belief Scale [45]. An item *"Astrology is a way to accurately predict the future"* was adapted to *"Tarot is a way to accurately reveal guidance about the future."* to fit the task better.

4.3.2 Post-Interaction Survey.

- **Subjective Topic Intimacy:** This exploratory measure captured how personally meaningful each conversation was. Participants are asked to rate how intimate the topic they discussed was (0–100%).
- **Trust in the Card:** Participants evaluated their impression of the card deck before evaluating their trust in the robot, in order to isolate different trust sources. This measure also allowed exploratory analysis.
- **Cognitive and Affective Trust in the Robot:** We followed the approach of Anzabi and Umamero [4], utilizing an adapted version of the pre-existing cognitive-affective trust questionnaire [15]. Terms such as "this person" were replaced with "Navel", and "work" with "interact" to match the study context. Averaged scores for cognitive and affective dimensions were calculated separately.
- **Certainty in Cognitive/Affective Trust Assessment:** Beside trust itself, we also ask *"How certain are you about your ratings to the statements above? (0-100%)"*, immediately after completing the cognitive or affective trust questionnaires. Inspired by stage-based models [17, 20], this metric represents a transition from dispositional to learned trust.

4.3.3 Post-Study Open Questions. At the end of the whole experiment, participants reflected on the overall experience and answered the following open questions:

(1) How did your impression of *Navel's understanding of you* (i.e., if Navel got to know you better) change over the course of rounds, if at all?
(2) How did your *affective trust* in Navel change as the rounds progressed, if at all? Did you feel more or less comfortable to emotionally rely on Navel over time?
(3) How did your *cognitive trust* in Navel change as the rounds progressed, if at all? Did your perception of Navel's ability and reliability improve or decline over the rounds?
(4) Is there any additional feedback for the whole experiment?

4.4 Procedure

The entire study took around 45 minutes. Each participant arrived individually and was welcomed into a quiet, isolated room. The researcher provided a brief overview of the study, invited the participant to sign the informed consent form, and collect their demographics as well as paranormal beliefs. Next, the participant engaged in a brief trial round designed to familiarize them with the conversational rhythm and the user interface. The trial consisted of three short dialogue rounds. In each round, the robot initiated a small-talk prompt (e.g., *"How was your day?"*), and repeated whatever the user responded to confirm that their speech had been recognized. The participant could opt to repeat the trial round if they need more practice.

The main experiments included three interaction rounds of the Card Divination Task. In each round, the participants were invited to share a personal concern, draw a card, and reflect with the robot as elaborated in Section 3.1. Since the conversation could be sensitive or personal, the researcher left the room private for the participant and the robot to ensure a comfortable setting. However, the researcher continuously monitored the process through a backend terminal to ensure that any technical issues could be addressed promptly. Each conversation lasted 4-5 minutes. After each round, the researcher re-entered the room to provide the post-interaction survey and to check in with the participant.

After all three rounds were completed, the participant answered the post-study questionnaire. Then, the researcher ethically clarify the nature of the task by explaining that the symbolic cards were neither magical nor tailored. Participants were able to ask questions or share feedback. Finally, they were thanked for their participation with a small snack and guided to leave.

4.5 Data Analysis

We performed mean-based statistical tests on the four major metrics: *cognitive trust*, *affective trust*, *certainty in cognitive trust assessment*, and *certainty in affective trust assessment*. Parametric assumptions were assessed with Levene's test (homogeneity) and Shapiro-Wilk test (normality in each group). If the assumptions are confirmed, we use mixed-design ANOVA to examine the effects of *social attitude* and *time*; otherwise, we used the non-parametric Wald-type test. Pearson correlations were conducted to examine relationships among exploratory metrics (*paranormal belief*, *topic intimacy*, and *trust in the card*), as well as their associations with *round number* and *cognitive*/*affective trust* in the robot. Finally, qualitative feedback was analyzed using thematic analysis.

5 Results

This section first examines demographic balance and correlations among exploratory metrics. Then, we present results for cognitive/affective trust and certainty (Figure 4), followed by a summary of qualitative feedback.

5.1 Participants

Chi-square and Kruskal-Wallis tests showed no significant differences in either gender (p = .431), age group (p = .928), or prior experience with robots/agents (p = .967) across conditions, suggesting balanced demographic assignments. A t-test also revealed balanced level of paranormal belief across groups (p = .686).

5.2 Correlation Analysis

Pearson correlation coefficients among trust-related variables are shown in Figure 5. Cognitive and affective trust in the robot were strongly correlated in both groups. While strongly associated with cognitive trust in both conditions, trust in card's association with affective trust was notably stronger in the social group than in the baseline group. No correlation was observed between round number and trust in card, suggesting minimal confounding despite card content is by design tied to round number.

In the social group, paranormal belief showed small to moderate correlations with cognitive trust, trust in card, and topic intimacy. However, this correlation was not significant in the baseline group or overall sample. We further fitted two linear mixed models: (1)

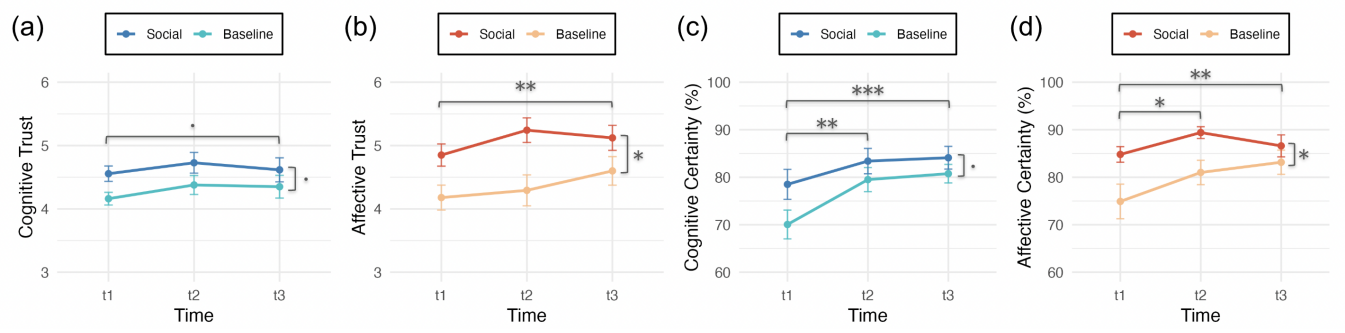

Figure 4: The results of mean (a) cognitive trust, (b) affective trust, (c) certainty of cognitive trust assessment and (d) certainty of affective trust assessment across conditions and multiple interaction rounds. ($^{***}: p \leq .001, ^{**}: p \leq .01, ^{*}: p \leq .05, \cdot: p \leq .1$)

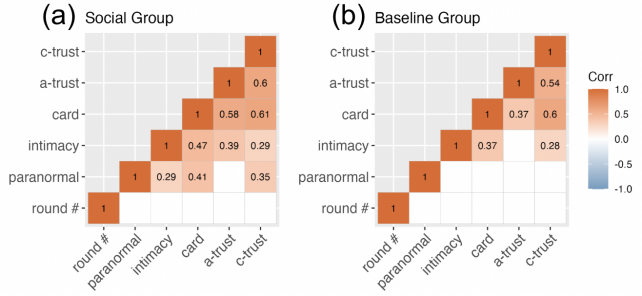

Figure 5: Pearson correlation coefficients among *cognitive trust*, *affective trust*, *trust in card*, *topic intimacy*, *paranormal belief*, and *round number*. Results in (a) social and (b) baseline groups are shown separately. Only statistically significant ($p < .05$) correlations are presented.

M_0 : *cognitive trust* ~ *social attitude* * *time* + (1 | *id*) and (2) M_1 : *cognitive trust* ~ *social attitude* * *time* + *paranormal* + (1 | *id*).

A likelihood ratio test showed that M_1 have neither better model fit ($\chi^2 = 1.207, p = .272$) nor smaller AIC (M_0 : 207.05, M_1 : 207.85). Therefore, paranormal belief was excluded in subsequent analyses comparing either groups or rounds.

5.3 Cognitive Trust

Mixed-design ANOVA revealed no significant interaction effect ($F = .265$, $p = .768$, $ges = .002$). The main effect of time was marginally significant ($F = 2.447$, $p = .093$), but the small effect size ($ges = .014$) suggests limited practical relevance. The main effect of social attitude was also marginally significant ($F = 3.089$, $p = .087$) with a medium effect size ($ges = .059$), indicating a trend where participants in the social condition reported higher cognitive trust ($M = 4.63$, $SD = .710$) than those in the baseline condition ($M = 4.30$, $SD = .652$).

5.4 Affective Trust

Although the baseline group violated the normality assumption, we proceeded with the mixed-design ANOVA, as it is robust to such violations and the residuals were normally distributed. There was no significant interaction effect ($F = 1.814$, $p = .170$, $ges = .010$), but a significant main effect of time ($F = 3.511$, $p = .010^{**}$) with a small to medium effect size ($ges = .026$). Results of pairwise t-tests with Bonferroni correction indicated that affective trust at t_3 ($M = 4.86$, $SD = .972$) was significantly higher than at t_1 ($M = 4.51$, $SD = .889$; $p = .010^{**}$, $d = .370$). No significant differences were found between t_1 and t_2 ($p = .107$, $d = .249$) or t_2 and t_3 ($p = 1.000$, $d = .089$). The significant main effect of social attitude ($F = 7.491$, $p = .009^{**}$, $ges = .136$) shows that participants in the social group reported higher affective trust ($M = 5.07$, $SD = .848$) than those in the baseline group ($M = 4.36$, $SD = .996$).

5.5 Certainty in Cognitive Trust Assessment

Similar to affective trust level, data in the baseline group violated the normality assumption. Nevertheless, we proceeded with the parametric analysis given the normal distribution of residuals and the robustness of mixed ANOVA. There was no significant interaction effect ($F = 1.183$, $p = .312$, $ges = .010$), but a significant main effect of time ($F = 11.967$, $p = .001^{***}$) with a medium to large effect size ($ges = .090$). Post-hoc pairwise t-tests with Bonferroni correction revealed that certainty significantly increased: (1) from t_1 to t_2 ($t = 3.66$, $p = .002^{**}$) with medium effect size ($d = .543$), and (2) from t_1 to t_3 ($t = 4.33$, $p < .001^{***}$) with medium effect size ($d = .637$). No significant difference was found between t_2 and t_3 ($t = .605$, $p = .549$, $d = .089$). The main effect of social attitude is marginal ($F = 2.837$, $p = .100$). However, the small to medium effect size ($ges = .049$) suggests that such effect could still be meaningful and that participants in the social group tend to report higher confidence in cognitive trust judgment.

5.6 Certainty in Affective Trust Assessment

The non-parametric Wald-type test revealed no significant interaction effect ($W = 3.832$, $p = .241$). A significant main effect of time was observed ($W = 13.755$, $p = .001^{***}$). Post-hoc Wilcoxon signed-rank tests with Bonferroni correction revealed a significant increase in certainty from t_1 to t_2 ($V = 113$, $p = .002^{**}$) and from t_1 to t_3 ($V = 184$, $p = .035^{*}$). Difference between t_2 and t_3 ($V = 316$, $p = 1.000$) was not significant. The main effect of social attitude was also significant ($W = 5.967$, $p = .015^{*}$), showing that participants

interacting with a socialized robot reported higher certainty in their assessment of affective trust.

5.7 Qualitative Results

We conducted an exploratory thematic analysis on the open-ended responses, focusing on positive/negative factors on trust as well as reasons of trust increasing, decreasing, or remaining.

5.7.1 Cognitive Trust. Both positive and negative factors are more relevant to *what* the robot said instead of *how* Navel spoke. Those who reported high cognitive trust described Navel as adaptive, knowledgeable, and appreciated the quality of advice or card interpretation. Conversely, those who have lower cognitive trust found the advice unhelpful or overly vague. From a developmental perspective, many ($n = 13$) participants reported that their cognitive trust remained relatively stable, whether consistently positive (e.g., *"professional from the start"*), neutral or negative. Some ($n = 7$) participants, however, noted an increase as they became more familiar with the interaction and more confident in Navel's performance.

5.7.2 Affective Trust. Positive factors include robot's empathetic and supportive attitude, making participants feel "heard" and safe to self disclose. Negative factors often stemmed from technical limitations or script design (e.g., interruption, lengthy utterance, short round). Notably, social manipulations such as robot's memory, referring users' name, and showing emotions are recognized as positive factors. On the other hand, many ($n = 6$) in the baseline group indicated that they had low affective trust in Navel because it acted too "official". Regarding development, several ($n = 5$) indicated how positive experiences helped foster trust (e.g., *"I was a bit scared Navel would judge me, but after the first round that fear disappear"*). However, others noted that despite positive impression overall, the interaction was too short and shallow to develop emotional bonds.

6 Discussion

This section first interprets the findings on trust and certainty development, revisits the hypotheses, and explores the relationships among multiple dimensions as well as other contextual factors. Then, we discuss ethical implications, limitations, and directions for future work.

6.1 Summary of Findings

6.1.1 Cognitive Trust Remains Stable. The level of cognitive trust did not show a significant change over rounds, suggesting that repeated interactions do not substantially alter participants' evaluations of the robot. In our experiment, each round followed the same procedure in which the robot demonstrated consistent capacity. As cognitive trust is typically grounded in rational assessments of agents' competence and reliability [31, 39], participants may have formed their judgments early and found no need to update them in the absence of new evidence. As one participant has pointed out: *"It (the level of cognitive trust) did not change much. Navel felt professional from the very beginning."*

6.1.2 Affective Trust Increases Gradually. The main effect of time indicated an increase of affective trust across interaction rounds, whereas post-hoc analyses revealed a gradual pace: significant difference emerged only between t_1 and t_3, but not in between. This offers empirical HRI evidence that relational trust is fostered through repeated interactions, consistent with interpersonal literature [22, 24, 44] and supports assumptions in computational models where benevolence is conceptualized as time-dependent [8, 48]. With positive experience, people may feel increasingly comfortable in a relationship and naturally develop affective trust. As one participant stated: *"The affective trust comes up when I realize Navel gives emotional support."*

6.1.3 Certainty In Trust Assessment Increases. For both dimensions, the main effect of time showed a clear increase in participants' certainty of trust judgments, supporting **H(c.1)** and **H(c.2)**. Drawing from the 3-layered model [17], users may enter an interaction with pre-existing, uncertain trust attitudes (i.e., dispositional layer). Through repeated encounters, however, trust shifts toward the dynamic-learned layer, in which users continually calibrate trust by comparing observed behavior with initial expectations [20]. The increase we found may reflect this transition.

Post-hoc results showed that participants started with relatively low certainty, became substantially more confident after t_2, and displayed only a modest increase thereafter. Similar to cognitive trust development, an explanation is that round 3 provided too little new information to update trust judgments. Once users felt they had sufficiently explored the system, their certainty stabilized.

6.1.4 Social Attitude Boosts Trust and Certainty. Social attitude have positive effects on both affective (significantly) and cognitive (marginally) trust, consistent with prior findings [4, 14, 39, 41] and support both **H(b.1)** and **H(b.2)**. However, without significant interaction, we cannot determine when this effect occurs and thus reject **H(b.3)**.

Our manipulation of social attitude led to significantly higher affective trust, aligning with the common association between benevolence and affective trust in parallel to competence and cognitive trust [4, 39]. Qualitative results resonant as many participants explicitly linked their affective trust to Navel's empathetic attitude. The marginal effect suggests higher cognitive trust in the social group even when robot's competence was held constant. Pralat et al.'s proposed a possible argument: according to the Media as Social Actors (MASA) paradigm [27], social cues can enhance perceptions of a social actor's characteristics, including competence [39]. In other words, the same level of competence may have been better recognized when accompanied by social behaviors.

Though no hypotheses were formulated for certainty, similar trends emerged: social attitude increased certainty significantly in affective trust and marginally in cognitive trust. Again, according to MASA [27], social cues may have supported users' interpretations and enabled more confident trust assessments.

6.1.5 Affective and Cognitive Trust Are Distinct, But Related. Our results revealed different developmental patterns between trust dimensions, where cognitive trust tends to establish early and affective trust emerges more gradually. This trend aligns with prior works [8, 22, 44, 48] and supports **H(a)**. From the perspective of 3-layered model [17], cognitive trust may be categorized in the initial learned layer (i.e., prior experience) after the first round, while affective trust may still belong to the dynamic learned layer (i.e., ongoing interaction).

The effects of social attitude also differed across the two dimensions. First, for both trust level and assessment certainty, the positive effect was significant for the affective dimension but only marginal for the cognitive one. Second, group differences were smaller for cognitive trust. Third, participants referred to social cues more frequently when discussing affective trust. Taken together, these results suggest that while social cues enhance trust in both dimensions, they are more strongly associated with affective trust, resonant with prior research [4, 39]. Overall, these differences between dimensions indicate that cognitive and affective trust are formed via distinct psychological mechanisms [17, 21, 35].

However, the strong correlation between cognitive and affective trust still reveal certain level of dependency, suggesting that users may integrate competence-based and emotion-based evaluations when forming overall trust judgment. As a result, while it is important to treat trust as a multidimensional construct, the dimensions should not be regarded independent.

6.1.6 Social Framing Shapes Perception of Context. The stronger interrelations among variables observed in the social condition (see Fig. 5) suggest that social cues might have prompted participants to apply interpersonal heuristics on anthropomorphized agents, as proposed in the CASA paradigm [40]. Topic intimacy was positively associated with affective trust only in the social group, suggesting that participants might have adjusted disclosure level based on the depth of emotional bonds. Similarly, paranormal belief was linked to multiple factors only in the social group, where participants seemed to evaluate both the robot and the ritual holistically. For instance, skeptics might think the robot was trying to "promote" mysticism and extend their distrust from the cards to the robot. In the baseline group, however, users may treat the robot as a neutral medium and evaluate the ritual independently.

6.2 Ethical Implication

While our task was designed to elicit trust, the goal was not maximizing it. Instead, trust should be maintained at an appropriate level to prevent both system disuse (from under-trust) and misuse (from over-trust) [33]. Trust inherently entails risk. In our Card Divination Task, risks include users overestimating Navel's capabilities, forming inappropriate cognitive trust, and acting on advice the system cannot generate responsibly. To mitigate this, we clarified afterwards that the stimuli were largely predefined and not based on any supernatural force. Excessive affective trust raises other threats, such as social rejection or betrayal if personal information were passed to third parties without consent [6]. We addressed this by explicit informed consent and careful pseudonymizations of conversation transcripts. Future research should identify distinct risks tied to each trust dimension, and develop safeguards through consent and mitigation strategies respectively.

6.3 Limitation and Future Work

This project presents several limitations and opens directions for future research. First, the semi-fixed script without fallback mechanisms limited interaction quality. Participants sometimes struggled to re-engage when the conversation proceeded unexpectedly. Simple dialogue management strategies, such as repeating the last utterance or clarifying intent, could improve the interaction.

A major limitation is the lack of counterbalance of card stimuli, tying the card content to the round number. Despite being an intentional choice due to practical considerations, this introduces difficulty in disentangling the effects of the card content from those of *time*. Though the correlation analysis revealed no significant association between trust in card and round number, the strong correlation between trust in card and trust in robot suggests that card content may still have influenced trust ratings. Should prospective research apply similar experimental design, round content should be counterbalanced to strengthen the validity.

The study relied solely on self-report questionnaire. While this follows the mainstream approaches of evaluating trust as internal attitude rather than demonstrated behavior [33, 49], incorporating behavioral metrics could provide richer insights. Future analyzes of our archived transcripts can complement current results, for example, through measures of self-disclosure [6, 18, 26], topic depth [36], content intimacy [6, 7, 26, 38], and sentiment [25].

Though motivated by lasting HRI relationships, our study consisted of only three consecutive rounds within less than an hour, limiting the generalizability of our findings to longer-term relationships that develop over days, weeks, or even years. According to Lewicki et al. [23], formation of interpersonal trusting relationships depends on interaction frequency, duration, and diversity. In line with most laboratory studies [24], this work focuses on frequency and reveals interesting trust development pattern despite the short duration. We hope these findings prompt reflection on how long-term interpersonal trust theories translate to repeated human-robot interactions, especially given that some HRI computational models [8, 48] are grounded in these theories without corresponding empirical evidence. For prospective researchers, we recommend adopting longitudinal design to further investigate how multidimensional trust evolve over extended duration.

Finally, we manipulated only the robot's social attitude but held competence constant. As prior works often links competence and benevolence respectively to cognitive and affective trust [4, 14], incorporating multiple levels of competence would allow more direct comparisons to literature and a richer understanding of how the two dimensions jointly develop.

7 Conclusion

This research investigated how multidimensional trust in a social, emotionally supportive robot evolves over repeated interactions. Guided by interpersonal literature, we hypothesized that affective trust requires more interaction rounds to develop than cognitive trust. Through a user study with 40 participants and a novel Card Divination Task, we provided empirical evidence that this hypothesis applies not only in human-human relationships but also in human-robot interactions. Furthermore, we found that social attitude enhanced both users' trust level and their certainty in trust assessment, especially in the affective dimension. These findings suggest that cognitive and affective trust have distinct developmental patterns while remaining interdependent. As robots and virtual agents increasingly enter our social lives, this work offers a foundation for sustaining appropriate trust in hybrid human-agent societies, empowering humans in their interactions with robots.

Acknowledgments

This publication is part of the project 'Hybrid Intelligence: augmenting human intellect' (https://hybrid-intelligence-centre.nl,) with project number 024.004.022 of the research programme 'Gravitation' which is (partly) financed by the Dutch Research Council (NWO).

References

[1] Muneeb Ahmad, Abdullah Alzahrani, Simon Robinson, and Alma Rahat. 2023. Modelling Human Trust in Robots During Repeated Interactions. In *International Conference on Human-Agent Interaction*. ACM, Gothenburg Sweden, 281–290. doi:10.1145/3623809.3623892

[2] Gene M Alarcon, August Capiola, Izz Aldin Hamdan, Michael A Lee, and Sarah A Jessup. 2023. Differential biases in human-human versus human-robot interactions. *Applied Ergonomics* 106 (2023), 103858. doi:10.1016/j.apergo.2022.103858

[3] Gene M Alarcon, Joseph B Lyons, Izz aldin Hamdan, and Sarah A Jessup. 2024. Affective responses to trust violations in a human-autonomy teaming context: humans versus robots. *International Journal of Social Robotics* 16, 1 (2024), 23–35. doi:10.1007/s12369-023-01017-w

[4] Naeimeh Anzabi and Hiroyuki Umemuro. 2023. Influence of Social Robots' Benevolence and Competence on Perceived Trust in Human-Robot Interactions. *JES Ergonomics* 59, 6 (Dec. 2023), 258–273. doi:10.5100/jje.59.258

[5] F. Babel, J. Kraus, L. Miller, M. Kraus, N. Wagner, W. Minker, and M. Baumann. 2021. Small talk with a robot? the impact of dialog content, talk initiative, and gaze behavior of a social robot on trust, acceptance, and proximity. *International Journal of Social Robotics* 13 (2021), 1485–1498. Issue 6. doi:10.1007/s12369-020-00730-0

[6] Franziska Burger, Joost Broekens, and Mark A. Neerincx. 2016. A Disclosure Intimacy Rating Scale for Child-Agent Interaction. In *Intelligent Virtual Agents*, David Traum, William Swartout, Peter Khooshabeh, Stefan Kopp, Stefan Scherer, and Anton Leuski (Eds.). Springer International Publishing, Cham, 392–396. doi:10.1007/978-3-319-47665-0_40

[7] Yuya Chiba and Akinori Ito. 2024. Speaker Intimacy Estimation in Chat-Talks Based on Verbal and Non-Verbal Information. *IEEE Access* 12 (2024), 184592–184606. doi:10.1109/ACCESS.2024.3507945

[8] Ameneh Deljoo, Tom van Engers, Leon Gommans, and Cees de Laat. 2018. Social Computational Trust Model (SCTM): A Framework to Facilitate Selection of Partners. In *2018 IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS)*. 45–54. doi:10.1109/INDIS.2018.00008

[9] D. H. Dickson and I. W. Kelly. 1985. The 'Barnum Effect' in Personality Assessment: A Review of the Literature. *Psychological Reports* 57, 2 (1985), 367–382. doi:10.2466/pr0.1985.57.2.367

[10] R. Falcone and C. Castelfranchi. 2004. Trust dynamics: how trust is influenced by direct experiences and by trust itself. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, 2004. AAMAS 2004.* 740–747. doi:10.5555/1018410.1018824

[11] Andrea Ferrario, Michele Loi, and Eleonora Viganò. 2020. In AI We Trust Incrementally: a Multi-layer Model of Trust to Analyze Human-Artificial Intelligence Interactions. *Philosophy Technology* 33 (09 2020). doi:10.1007/s13347-019-00378-3

[12] Catherine Fichten and Betty Sunerton. 1983. Popular Horoscopes and the "Barnum Effect". *Journal of Psychology - J PSYCHOL* 114 (05 1983), 123–134. doi:10.1080/00223980.1983.9915405

[13] N. Gasteiger, K. Loveys, M. Law, and E. Broadbent. 2021. Friends from the future: a scoping review of research into robots and computer agents to combat loneliness in older people. *Clinical Interventions in Aging* Volume 16 (2021), 941–971. doi:10.2147/cia.s282709

[14] Ioanna Giorgi, Francesca Ausilia Tirotto, Oksana Hagen, Farida Aider, Mario Gianni, Marco Palomino, and Giovanni L. Masala. 2022. Friendly But Faulty: A Pilot Study on the Perceived Trust of Older Adults in a Social Robot. *IEEE Access* 10 (2022), 92084–92096. doi:10.1109/ACCESS.2022.3202942

[15] Kent Grayson. 2005. Cognitive and Affective Trust in Service Relationships. Journal of Business Research. *Journal of Business Research* 58 (04 2005), 500–507. doi:10.1016/S0148-2963(03)00140-1

[16] Y. Guo and X. J. Yang. 2020. Modeling and predicting trust dynamics in human–robot teaming: a bayesian inference approach. *International Journal of Social Robotics* 13 (2020), 1899–1909. Issue 8. doi:10.1007/s12369-020-00703-3

[17] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors* 57, 3 (2015), 407–434. doi:10.1177/0018720814547570

[18] Regina Jucks, Gesa A. Linnemann, Franziska M. Thon, and Maria Zimmermann. 2016. *Trust the Words: Insights into the Role of Language in Trust Building in a Digitalized World*. Springer International Publishing, Cham, 225–237. doi:10.1007/978-3-319-28059-2_13

[19] Jurgis Karpus, Adrian Krüger, Julia Tovar Verba, Bahador Bahrami, and Ophelia Deroy. 2021. Algorithm exploitation: Humans are keen to exploit benevolent AI. *iScience* 24, 6 (2021), 102679. doi:10.1016/j.isci.2021.102679

[20] Johannes Kraus. 2020. *Psychological processes in the formation and calibration of trust in automation.* Ph. D. Dissertation. Ulm University. doi:10.18725/OPARU-32583

[21] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46, 1 (2004), 50–80. doi:10.1518/hfes.46.1.50_30392 PMID: 15151155.

[22] Roy Lewicki and Barbara Bunker. 1994. Trust in relationships: A model of development and decline. (01 1994).

[23] Roy Lewicki, Daniel McAllister, and Robert Bies. 1998. Trust And Distrust: New Relationships and Realities. *The Academy of Management Review* 23 (07 1998). doi:10.2307/259288

[24] Roy J. Lewicki, Edward C. Tomlinson, and Nicole Gillespie. 2006. Models of Interpersonal Trust Development: Theoretical Approaches, Empirical Evidence, and Future Directions. *Journal of Management* 32, 6 (2006), 991–1022. doi:10.1177/0149206306294405

[25] Mengyao Li, Isabel M Erickson, Ernest V Cross, and John D Lee. 2024. It's Not Only What You Say, But Also How You Say It: Machine Learning Approach to Estimate Trust from Conversation. *Human Factors* 66, 6 (2024), 1724–1741. doi:10.1177/00187208231166624 PMID: 37116009.

[26] Mike Ligthart, Timo Fernhout, Mark A. Neerincx, Kelly L. A. van Bindsbergen, Martha A. Grootenhuis, and Koen V. Hindriks. 2019. A Child and a Robot Getting Acquainted - Interaction Design for Eliciting Self-Disclosure. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems* (Montreal QC, Canada) *(AAMAS '19)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 61–70.

[27] M. Lombard and K. Xu. 2021. Social responses to media technologies in the 21st century: the media are social actors paradigm. *Human-Machine Communication* 2 (2021), 29–55. doi:10.30658/hmc.2.2

[28] Bertram F. Malle and Daniel Ullman. 2021. Chapter 1 - A multidimensional conception and measure of human-robot trust. In *Trust in Human-Robot Interaction*, Chang S. Nam and Joseph B. Lyons (Eds.). Academic Press, 3–25. doi:10.1016/B978-0-12-819472-0.00001-0

[29] Bertram F. Malle and Daniel Ullman. 2023. Measuring Human-Robot Trust with the MDMT (Multi-Dimensional Measure of Trust). arXiv:2311.14887 [cs.RO] https://arxiv.org/abs/2311.14887

[30] R. C. Mayer, J. H. Davis, and F. D. Schoorman. 1995. An integrative model of organizational trust. *The Academy of Management Review* 20 (1995), 709. Issue 3. doi:10.2307/258792

[31] D. J. McAllister. 1995. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal* 38 (1995), 24–59. Issue 1. doi:10.2307/256727

[32] Siddharth Mehrotra. 2021. Modelling Trust in Human-AI Interaction *(AAMAS '21)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1826–1828.

[33] Siddharth Mehrotra, Chadha Degachi, Oleksandra Vereschak, Catholijn M. Jonker, and Myrthe L. Tielman. 2023. A Systematic Review on Fostering Appropriate Trust in Human-AI Interaction. arXiv:2311.06305 [cs.HC] https://arxiv.org/abs/2311.06305

[34] Jingbo Meng and Nancy Dai. 2021. Emotional Support from AI Chatbots: Should a Supportive Partner Self-Disclose or Not? *Journal of Computer-Mediated Communication* 26 (05 2021). doi:10.1093/jcmc/zmab005

[35] L. Miller, J. Kraus, F. Babel, and M. Baumann. 2021. More than a feeling—interrelation of trust layers in human-robot interaction and the role of user dispositions and state anxiety. *Frontiers in Psychology* 12 (2021). doi:10.3389/fpsyg.2021.592711

[36] Seiya Mitsuno, Midori Ban, Hiroshi Ishiguro, and Yuichiro Yoshikawa. 2024. Deepening Conversations Over Time: A Chatbot with a Topic Depth Estimation Model for Gradually Engaging in Deeper Chats. In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*. 1354–1361. doi:10.1109/RO-MAN60168.2024.10731430

[37] Xin Yi Or, Yu Xuan Ng, and Yong Shian Goh. 2025. Effectiveness of social robots in improving psychological well-being of hospitalised children: A systematic review and meta-analysis. *Journal of Pediatric Nursing* 82 (2025), 11–20. doi:10.1016/j.pedn.2025.01.032

[38] Jiaxin Pei and David Jurgens. 2020. Quantifying Intimacy in Language. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 5307–5326. doi:10.18653/v1/2020.emnlp-main.428

[39] Nele Pralat, Carolin Ischen, and Hilde Voorveld. 2025. Feeling Understood by AI: How Empathy Shapes Trust and Influences Patronage Intentions in Conversational AI. In *Chatbots and Human-Centered AI*, Asbjørn Følstad, Symeon Papadopoulos, Theo Araujo, Effie L.-C. Law, Ewa Luger, Sebastian Hobert, and Petter Bae Brandtzaeg (Eds.). Springer Nature Switzerland, Cham, 234–259. doi:10.1007/978-3-031-88045-2_15

[40] Byron Reeves and Clifford Nass. 1996. The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK* 10, 10 (1996), 19–36.

[41] Fabian Reinkemeier, Philipp Spreer, and Waldemar Toporowski. 2021. *Voice Assistants in Voice Commerce: The Impact of Social Cues on Trust and Satisfaction.* 130–135. doi:10.1007/978-3-030-86797-3_9

[42] Minjin Rheu, Ji Youn Shin, Wei Peng, and Jina Huh-Yoo and. 2021. Systematic Review: Trust-Building Factors and Implications for Conversational Agent Design. *International Journal of Human–Computer Interaction* 37, 1 (2021), 81–96. doi:10.1080/10447318.2020.1807710

[43] Jimin Rhim, Sonya S. Kwak, Angelica Lim, and Jason Millar. 2023. The dynamic nature of trust: Trust in Human-Robot Interaction revisited. arXiv:2303.04841 [cs.CY] https://arxiv.org/abs/2303.04841

[44] Denise Rousseau, Sim Sitkin, Ronald Burt, and Colin Camerer. 1998. Not So Different After All: A Cross-discipline View of Trust. *Academy of Management Review* 23 (07 1998). doi:10.5465/AMR.1998.926617

[45] Jerome Tobacyk. 2004. A Revised Paranormal Belief Scale. *International Journal of Transpersonal Studies* 23 (01 2004). doi:10.1037/t14015-000

[46] Claude Toussaint, Philipp T Schwarz, and Markus Petermann. 2023. Navel - a social robot with verbal and nonverbal communication skills. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 463, 4 pages. doi:10.1145/3544549.3583898

[47] Anna-Sophie Ulfert, Eleni Georganta, Carolina Centeio Jorge, Siddharth Mehrotra, and Myrthe Tielman and. 2024. Shaping a multidisciplinary understanding of team trust in human-AI teams: a theoretical framework. *European Journal of Work and Organizational Psychology* 33, 2 (2024), 158–171. doi:10.1080/1359432X.2023.2200172

[48] Joana Urbano, Ana Paula Rocha, and Eugénio Oliveira. 2013. The Impact of Benevolence in Computational Trust. In *Agreement Technologies*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Carlos Iván Chesñevar, Eva Onaindia, Sascha Ossowski, and George Vouros (Eds.). Vol. 8068. Springer Berlin Heidelberg, Berlin, Heidelberg, 210–224. doi:10.1007/978-3-642-39860-5_16 Series Title: Lecture Notes in Computer Science.

[49] Magdalena Wischnewski, Nicole Krämer, and Emmanuel Müller. 2023. Measuring and Understanding Trust Calibrations for Automated Systems: A Survey of the State-Of-The-Art and Future Directions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 755, 16 pages. doi:10.1145/3544548.3581197

Received 2025-09-30; accepted 2025-12-01