Effects of Perceived Robot Autonomy and Personal Differences on Trust in Human-Robot Interactions

Ali Fallahi[1], Patrick Holthaus[1], Farshid Amirabdollahian[1], and Gabriella Lakatos[1]

Robotics Research Group, University of Hertfordshire, Hatfield, United Kingdom.
{a.fallahi, p.holthaus, f.amirabdollahian2, g.lakatos}@herts.ac.uk

Abstract. Trust is essential in human-robot interaction (HRI), yet the role of various factors shaping trust remains complex. This study investigated how perceived robot autonomy and individual user traits influence trust. Participants interacted with a Pepper robot in one of two conditions differing in the manipulation of perceived autonomy. Results showed that perceived autonomy affected ratings of the robot's sincerity, but did not significantly affect other trust dimensions. Participants' pre-existing attitudes toward robots were associated with trust perceptions, while personality traits showed no significant influence. These findings suggest that user attitudes may play a more critical role than perceived autonomy in shaping trust during short-term HRI, underscoring the need to personalise robot design based on attitudinal differences.

Keywords: HRI · Trust · Percieved Autonomy

1 Introduction

As robots become increasingly integrated into homes, workplaces, and healthcare settings, understanding how humans form trust in these systems is crucial to ensure effective and safe collaboration [3, 15]. In HRI, trust directly influences people's willingness to rely on robotic systems, engage with them, and accept their assistance [26]. While building technically advanced robots is a growing achievement, ensuring they are perceived as trustworthy remains a central challenge [1]. Both under-trust and over-trust can lead to risks such as disengagement, disuse, or blind over-reliance in critical situations [2, 22].

One of the key dimensions that seems to influence trust is a robot's autonomy—the degree to which it operates independently of human control. Robots can be classified along a continuum of autonomy, from tele-operated systems to fully autonomous agents [6]. Autonomous systems are often perceived as intentional agents [16], and these perceptions can significantly impact user trust, even if the robot's actual behaviour remains unchanged [12]. Prior studies suggest that higher autonomy can increase trust, but only when the system remains predictable and transparent [11, 23, 25]. Misperceptions of autonomy—such as assuming a robot is autonomous when it is not—can lead to mismatched expectations and breakdowns in trust. Despite growing interest in this topic, few studies have directly examined how perceived autonomy (rather than actual autonomy) shapes both subjective and behavioural trust responses.

This study focuses on the relationship between perceived autonomy and trust in a social HRI context. This study uniquely focuses on perceived, not actual, robot autonomy, addressing an overlooked aspect in HRI trust research. Specifically, we investigate how merely telling users a robot is autonomous versus remotely controlled influences their perception of trustworthiness and their willingness to engage with it. We also consider the role of individual differences—such as personality traits, attitudes towards robots, and gender—which are known to shape trust in automation [11], yet are underexplored in relation to perceived autonomy. Understanding these relationships has two major implications. First, it can help robot designers determine how to frame or implement autonomy in ways that foster appropriate trust. Second, it provides insight into how experimental narratives and user traits influence outcomes in HRI studies, promoting more standardised and comparable trust assessments across the field. To address these goals, we investigate the following research questions:

RQ1 How might people's perception of robot autonomy influence their trust towards companion robots? **RQ2** How do individual differences (personality traits, attitudes towards robots, gender) influence trust in robots?

2 Background and Motivation

The complex concept of trust remains a challenging and evolving research area within the field of HRI, with many questions still open for investigation [29]. One of the key factors influencing trust in HRI is the perceived autonomy of the robot; how independently it appears to operate and make decisions. An object is considered an agent if it performs a beneficial function for itself or others, showing autonomy [16]. Robots can be categorised based on their level of autonomy into two main types: autonomous robots and tele-operated robots [6]. An autonomous robot operates independently, completing tasks without human intervention, while a tele-operated robot relies on a human operator to achieve its goals [8]. In shared control models, human inputs and the robot's autonomous control are combined to realise the robot's behaviours [21].

The way a robot behaves, including its level of autonomy, can shape trust as a psychological response in users. However, people's expectations of how autonomous these systems are can sometimes be inaccurate [12]. The effects of robot autonomy on human trust, especially concerning how perceptions of robot agency affect trust, have recently been getting the attention of some HRI researchers. In a comprehensive meta-analysis, [11] found that higher robot autonomy tends to increase trust. The authors mentioned that robot reliability and predictability might be affected by a robot's degree of autonomy. They noted, however, that unexpected behaviours at high autonomy levels can severely reduce trust, pointing to the challenge of balancing autonomy with trustworthiness. This is in line with findings of [25], who found that trust differs significantly based on the level of robot autonomy, and it drops when robots perform essential tasks autonomously without human supervision. They proposed the implementation of precisely defined autonomy levels to keep efficiency and trust at acceptable levels. [5] examined how trust develops differently under varying autonomy conditions, a trust-aware partially observable Markov decision process [14], and a myopic decision-making strategy, where the robot acts without considering trust. Their findings showed that medium-level autonomy created the most favourable conditions for trust development in a table-clearing scenario.

As discussed above, high autonomy can lead to user ambiguity and the feeling of diminished control, thus demonstrating the necessity of robot intentionality and transparency to maintain trust. [23] demonstrated that transparent decision-making by autonomous robots led to higher trust compared to robots that made decisions without providing explanations. Although these findings provide valuable insights into the connection between robot autonomy and user trust, there are no insights into how user characteristics like personality traits, demographic factors, and attitudes toward robots might influence these perceptions. Trust in this study is defined both as subjective (perceived sincerity, reliability, competence, ethics) and behavioural (willingness to follow robot requests) in line with existing literature on trust in human-robot interaction [11].

3 Methods

To investigate our research questions, we conducted an in-person study in the University of Hertfordshire's Robot House, a four-bedroom residential home adapted for HRI research. Participants interacted with a Pepper robot in one of two conditions that differed only in perceived autonomy. The study was approved by The University of Hertfordshire Health, Science, Engineering and Technology Ethics Committee with Delegated Authority (SPECS/PGR/UH/05839). Based on previous research, we formulated the following hypotheses:

H1 Effect of perceived autonomy (related to RQ1) We hypothesise that the belief that a robot is autonomous, as opposed to remotely controlled, will affect participants' trust in the robot: **H1.1** Participants who perceive the robot as autonomous will rate it as more trustworthy (subjective evaluation). **H1.2** Participants who perceive the robot as autonomous will show greater willingness to interact with the robot and follow its instructions (behavioural response).

H2 Effect of individual differences (related to RQ2) We hypothesise that participants' individual characteristics will influence how they perceive and respond to the robot: **H2.1** Participants' personality traits will influence their perceptions of the robot's trustworthiness and social attributes. **H2.2** Participants' pre-existing attitudes toward robots will influence their perceptions of and behavioural responses to the robot.

3.1 Experimental Manipulation

This study employed a between-participants experimental design to manipulate people's perception of robot autonomy. Participants were randomly assigned to one of two experimental conditions to investigate potential influences on trust and interaction behaviours.

- *Remotely-controlled* (RC): Participants were told that the robot was being controlled by a human operator. To reinforce the participants' belief that a human was controlling the robot, the observation room with a human controller was shown at the beginning of the experiment and the controller engaged in five "check-up" procedures, one before each task, making adjustments to the robot and announcing, "It's ready for the next scenario."
- *Autonomous* (AU): Participants were told that the robot was acting independently. The control room was not shown, and no "check-up" interventions by the experimenter were performed. Like in the other condition, however, the robot was remotely controlled to ensure consistency of behaviours.

3.2 Participants

A total of 33 participants (12 identified as female, 19 as male, and 1 as genderfluid) took part in the experiment. Recruitment was based on similar HRI studies [22, 24] with similar sample sizes. Their ages ranged from 19 to 40 years (mean age $\bar{x}_{Age} = 26.48$, standard deviation $\sigma_{Age} = 4.6$). Participants were recruited individually on campus via flyers and randomly assigned to one of two experimental conditions AU (16 participants) or RC (17 participants).

3.3 Experimental Procedure

Participants, upon arrival at the Robot House, were greeted by the experimenter and escorted to the living room area, where they were given a brief explanation of the study. After reviewing and signing a consent form, participants completed two pre-intervention questionnaires (cf. Sect. 3.4). They were then introduced to the study scenario where they were visiting a friend's home while the friend was away, to check if everything was alright. The robotic assistant, Pepper, would stay with them and help them with feeding the owner's cat. Participants were instructed to interact with the robot naturally and comfortably, and that all instructions would be provided by the robot during the interaction. Depending on their previously assigned experimental group, they were told that the robot was either autonomous or remotely controlled. In the RC condition only, the experimenter introduced the observation room. To initiate the interaction, participants were asked to stand on a marked point in the centre of the living room, simulating the moment of having just entered the house. The experimenter then excused themselves and left for the control room to trigger the robot's behaviour in a Wizard-of-Oz style. Pepper approached the participant and began the interaction. The robot then asked the participant to complete four interactive tasks, selected to represent social scenarios reflecting different aspects of trust:

1. **Show ID:** At the beginning of the experiment, the robot asked participants to show an identification document. Participants could choose to comply or decline and were later asked to explain their decision.
2. **Fill Bowl:** The robot instructed participants to help feed and provide water to the cat. Cat food and two bottles were provided: one with red liquid and another containing clear liquid, labelled "diluted cleaning solution." Participants could show one or both bottles to the robot for feedback before making a decision. If participants showed the clear bottle, the robot confirmed the choice; if they showed the red liquid, it issued a warning. If participants presented both, the robot recommended the clear one. These responses were intended to simulate autonomous decision-making and prompt reflection on the robot's judgment and reliability.
3. **Play Sudoku:** The robot invited participants to play a Sudoku puzzle on its screen, and offered help solving it. Participants could either solve the puzzle independently or ask the robot to provide the solution for them.
4. **Join Dance:** The robot invited participants to dance to a music genre of their own or Pepper's choice. This scenario assessed willingness to engage in a light-hearted activity, while potentially eliciting feelings of discomfort.

Following the final interaction, participants were asked to complete two post-intervention questionnaires, evaluating their perceptions of the robot's social characteristics and trustworthiness, respectively (cf. Sect. 3.4). At the end of the session, participants were fully debriefed.

3.4 Dependent variables

Objective (behavioural) measurements We measured participants' willingness to follow the robot's suggestions in each of the four tasks as an established behavioural trust indicator [22,23], detailed in Section 3.3. We established different social situations, where the ID validation task revealed personal information, the water bowl task looked at a pet that might be reliable on the participant's judgements, a casual situation playing Sudoku, and a dancing situation putting people in a less comfortable position.

Subjective (questionnaire) measurements The Ten-Item Personality Inventory (TIPI) [10] was used as a pre-intervention questionnaire to assess participants' personality traits and to help address **RQ2**, as individual differences were expected to influence trust and engagement. TIPI is a brief questionnaire containing 10 items on a 7-point scale, designed to assess the "Big Five" personality traits: Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness to Experience [7, 13]. Likewise, the Negative Attitude Towards Robots (NARS) [20] (14 items, 5-point scale) was employed prior to the interaction to examine general concerns and attitudes toward robots, contributing further insights into **RQ2**. NARS evaluates negative attitudes across three dimensions: Negative Attitudes toward Interaction with Robots, toward Social Influence of Robots, and toward Emotional Interaction with Robots.

Table 1: Wilcoxon Rank-Sum Test Results for TIPI Traits between the robot conditions AU and RC. * indicates a significant difference ($p < 0.05$).

Trait	**W Statistic**	**p-value**	**Mean *AU***	**Mean *RC***
Extraversion	103.0	0.2391	48.44	59.31
Agreeableness	128.5	0.7981	65.62	67.16
Conscientiousness*	68.5	0.0138	69.79	85.29
Emotional stability	136.0	1.0000	60.42	60.78
Openness	93.0	0.1127	76.56	89.22

The Multi-Dimensional Measure of Trust (MDMT) [18, 30] was used as a post-intervention measure to evaluate trust in the robot, thereby addressing **RQ1**. MDMT includes 20 items on an 8-point scale spanning five trust dimensions: Competence, Reliability, Integrity, Sincerity, and Benevolence. The Robot Social Attribute Scale (RoSAS) [4] (18 items. 7-point scale) was also administered post-interaction to assess perceptions of the robot's social characteristics. It includes three subscales: Warmth, Competence, and Discomfort.

4 Results

As the data were not continuous and did not meet the assumption of normality, as confirmed by the Shapiro–Wilk test [27] ($p < .05$), non-parametric methods were applied to all statistical tests. Specifically, Wilcoxon rank-sum tests [19] were used for between-group comparisons of questionnaire data, Fisher's exact tests [9] were used to compare proportions of responses between conditions on binary (Yes/No) data, and Spearman's rank correlation coefficients [28] (ρ) were computed to examine associations among variables.

4.1 Condition Balance Checks

Participants were randomly distributed across the experimental conditions with a balanced gender composition and similar age profiles. The AU condition included 6 female, 1 genderfluid, and 9 male participants, with a mean age of 25.56 years (SD = 4.02). The RC condition comprised 7 female and 10 male participants, with a mean age of 27.35 years (SD = 5.06). To check for potential pre-existing differences between groups and confounding factors, we examined participants' personality traits (TIPI) and attitudes towards robots (NARS). However, personality traits were also investigated across all participants, independent of condition, to address RQ2 (see Sect. 4.3).

On the TIPI scale, only *Conscientiousness* showed a statistically significant difference between conditions, where people in the RC condition showed higher conscientiousness than people in the AU condition ($\bar{x}_{RC} = 85.29$, $\bar{x}_{AU} = 69.79$, $W = 68.5$, $p = 0.0138$). No other traits showed significant differences; for details, refer to Table 1. No statistically significant differences were found between AU and RC across any of the NARS subscales ($p > .05$ for all).

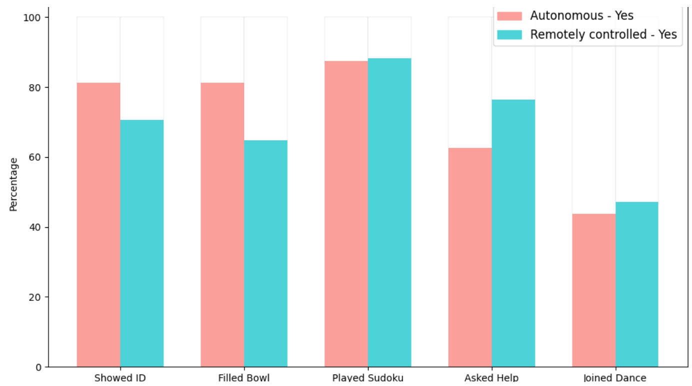


Fig. 1: Bar chart showing the percentage of *Yes* responses to behavioural questions grouped by experimental condition (AU and RC).

4.2 Effect of Condition

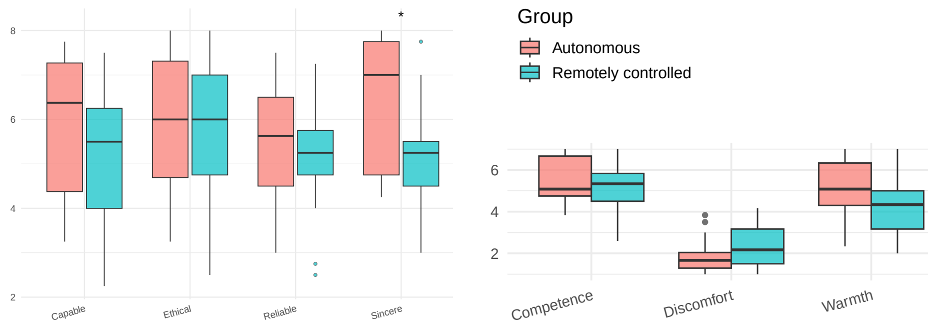
Behavioural Responses To assess whether participants' behavioural responses differed between the AU and RC robot conditions, Fisher's Exact Tests were conducted on five measurements (Yes/No) whether participants engaged in each of the tasks (cf. Sect. 3.3), i.e whether they showed their ID, filled the cat's bowl, played a game of Sudoku with the robot, asked it for help during the game, and joined the robot in its dancing routine. Figure 1 displays these behaviours, grouped by condition.

Response rates are also presented in Table 2, along with odds ratios and p-values from Fisher's Exact Tests. No statistically significant differences were observed between the AU and RC conditions across any of the behaviours.

Table 2: Fisher's Exact Test results and behavioural responses (number and percentage) for tasks across the robot conditions AU and RC.

Activity	**p-value**	**Odds Ratio**	**No − AU**		**Yes − AU**		**No − RC**		**Yes − RC**	
Showed ID	0.6880	1.806	3	18.8%	13	81.2%	5	29.4%	12	70.6%
Filled Bowl	0.4384	2.364	3	18.8%	13	81.2%	6	35.3%	11	64.7%
Played Sudoku	1.0000	0.933	2	12.5%	14	87.5%	2	11.8%	15	88.2%
Asked Help	0.4646	0.513	6	37.5%	10	62.5%	4	23.5%	13	76.5%
Joined Dance	1.0000	0.875	9	56.2%	7	43.8%	9	52.9%	8	47.1%

Subjective Measurements To assess people's trust in and social perception of the robot, Wilcoxon rank-sum tests between the conditions were conducted on both `MDMT` and `RoSAS`. On the `MDMT` scale, a statistically significant difference



(a) Boxplots of MDMT subdimensions.　　(b) Boxplots of RoSAS subscales.

Fig. 2: Boxplots of dependent subjective measurements are shown on a scale between 1 and 7, grouped by robot condition (AU left, red and RC right, cyan). Significant differences between the conditions ($p < 0.05$) are indicated by '*'.

was found for the *Sincere* subdimension, with participants in the AU condition reporting higher sincerity ratings than those in the RC condition ($W = 194.0$, $p = 0.0376$, $\hat{x}_{AU} = 6.36$, $\sigma_{AU} = 1.50$; $\hat{x}_{RC} = 5.16$, $\sigma_{RC} = 1.19$). No significant differences were observed for the other trust dimensions (*Reliable*, *Capable*, *Ethical*). No statistically significant differences between the conditions were found in the RoSAS subscales *Competence*, *Warmth*, and *Discomfort*. See Figure 2 for an illustration and Tables 3 and 4 for detailed test results.

Table 3: Wilcoxon Rank-Sum Test Results for MDMT subdimensions between the experimental conditions AU and RC.

Subdimensions	**W Statistic**	**p-value**	**Mean value AU**	**Mean value RC**
Capable	177.0	0.1440	5.85	5.1
Ethical	138.5	0.9424	6.01	5.88
Reliable	162.5	0.3468	5.62	5.13
Sincere	194.0	0.0376	6.36	5.16

Overall, scores were relatively high in all MDMT subdimensions and the *Competence* and *Warmth* subscales of RoSAS, with median ratings at the upper end of the 7-point scale, whereas *Discomfort* was rated on the lower and, as expected.

4.3 Effect of Individual Differences

In addition to testing for group differences between the AU and RC conditions, we also investigated relations between demographics and pre-intervention measurements (personality traits, robot attitude) and dependent variables (trust and social attributes).

Table 4: Wilcoxon Rank-Sum Test Results for RoSAS Subscales between the robot conditions *AU* and *RC*.

Trait	W Statistic	p-value	Mean value AU	Mean value RC
Competence	150.5	0.6135	5.50	5.10
Discomfort	97.0	0.1636	1.90	2.35
Warmth	176.0	0.1542	5.01	4.30

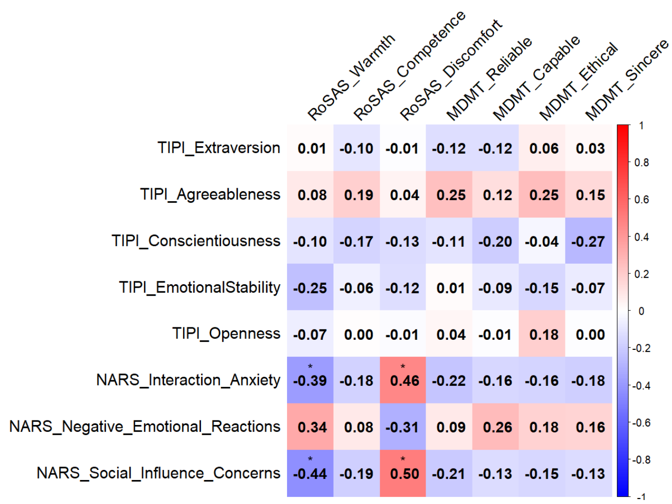


Fig. 3: Spearman correlation heatmap labelled with coefficients between individual difference measures (TIPI and NARS subscales) and robot perception (RoSAS and MDMT subdimensions). Warmer colours indicate stronger positive correlations, while cooler colours indicate stronger negative correlations. Statistically significant correlations ($p < 0.05$) are marked with an asterisk (*).

Age-related effects were not analysed due to similar age distributions across conditions No significant gender differences were found across the four MDMT trust subdimensions (Wilcoxon rank-sum tests, all $p > .05$). Spearman correlations and p-values between individual traits (TIPI, NARS) and participants' perceptions of the robot (RoSAS, MDMT) are shown in Figure 3, with warmer colours indicating stronger positive and cooler colours stronger negative correlations. Coefficients are displayed within each cell, with statistically significant results ($p < 0.05$) marked by an asterisk (*). We report only statistically significant correlations ($p < .05$); values not reaching significance are only shown in Figure 3. Among the NARS subscales, *Social Influence Concerns* was positively correlated with *RoSAS Discomfort* ($\rho = 0.50$, $p = .0029$) and negatively correlated with *RoSAS Warmth* ($\rho = -0.44$, $p = .0109$). Similarly, *Interaction Anxiety* positively correlated with *RoSAS Discomfort* ($\rho = 0.46$, $p = .0069$) and negatively correlated with *RoSAS Warmth* ($\rho = -0.39$, $p = .0253$).

5 Discussion and Limitations

Four of the five personality traits were balanced across groups, but participants in the *RC* condition showed higher *Conscientiousness*, a trait linked to rule-following and trust in structured systems [13]. This imbalance may have influenced behavioural responses independently of the autonomy framing. Participants' negative attitudes towards robots did not significantly differ between conditions, suggesting that trust-related effects are unlikely to stem from pre-existing biases. Gender and age were evenly distributed, though the limited diversity may restrict generalisability. Overall, the groups were comparable in personality and attitudes, except for conscientiousness.

H1.1 proposed that participants who believed the robot was *AU* would perceive it as more trustworthy than those who believed it was *RC*. Results from post-interaction trust scales suggest partial support for this hypothesis. Specifically, participants in the *AU* condition rated the robot more positively in terms of sincerity, suggesting that perceived autonomy influenced certain social-intentional evaluations. Other trust-related attributes, like competence, reliability, and warmth, did not differ between conditions. This suggests that while autonomy framing influenced some social judgments, it did not consistently enhance trust across all dimensions. These results align with studies showing that perceived autonomy can trigger anthropomorphic responses, though such effects may depend on context, task, and user expectations [12,17,31]. Although agency was not directly measured, the increase in perceived sincerity in the *AU* condition may indicate that participants implicitly attributed greater agency to the robot [32]. This reinforces the role of framing in shaping how humans evaluate social robots. Although anthropomorphism is known to influence perceived agency and thus potentially trust in HRI, this factor was beyond the scope of this study. Future research should include explicit assessments of anthropomorphism to better understand potentially confounding effects on trust evaluations.

H1.2 proposed that participants who believed the robot was autonomous would be more likely to follow its instructions and engage with it during tasks. However, the findings did not support this hypothesis. Participants in both conditions showed generally high levels of compliance, with no clear differences in behaviour across tasks such as showing ID, requesting help, or completing challenges. These results suggest that merely telling participants a robot is autonomous may not be enough to change behaviour. Prior research shows that observable behaviours—like adaptability or visible errors—build trust more effectively than verbal framing [22]. In our study, the robot's identical behaviour across conditions likely reduced the impact of the autonomy manipulation. Moreover, participants may have complied with the robot's requests simply because they perceived it as part of their role in a short-term experimental setting, rather than due to genuine trust. As highlighted by Salem et al. [23], participants in brief lab-based interactions often comply even with illogical or faulty robot instructions, indicating potential overtrust or perceived obligation. These findings underline the limitations of short-term studies in assessing behavioural trust and suggest that more ecologically valid, long-term interactions are necessary to capture authentic user responses. Future studies could explore alternative cues of autonomy to better understand when and how perceived autonomy translates into behavioural trust. **H2.1** and **H2.2** focused on the influence of personality traits and pre-existing attitudes toward robots on how participants perceived the robot after interaction. The findings provide partial support for these hypotheses. Participants' negative attitudes toward robots, particularly discomfort in interacting with them, were associated with more negative evaluations of the robot's warmth and sociability. These results suggest that individuals with more negative attitudes or anxiety toward robots tend to see them as less warm and more discomforting. While this correlation between NARS and RoSAS is expected due to their conceptual overlap, it reinforces the role of affective attitudes in shaping social perceptions of robots. Notably, no significant correlations were found between personality or robot attitudes and MDMT trust dimensions, indicating that individual differences may influence social perceptions but not necessarily trust in the robot's abilities or intentions. In contrast, personality traits such as extraversion or agreeableness did not appear to meaningfully influence participants' perceptions. This implies that attitudinal factors (as measured by NARS) may be more predictive of subjective trust in robots than general personality traits. These results highlight the importance of accounting for users' preconceptions about robots when evaluating their reactions to autonomous systems. Moreover, results suggest that personalising robots based on personality traits may be less effective than focusing on user attitudes. These findings underline the importance of addressing user biases and robot-related concerns in designing socially acceptable robotic systems, especially when the goal is to foster trust and cooperation. Limitations include the short-term lab setting and that autonomy manipulation was based on framing and contextual cues, while robot behaviours were intentionally kept constant to control for confounds. Future studies could include observable autonomy primers to further strengthen the manipulation.

6 Conclusion and Future Work

This study explored how perceived robot autonomy and user differences influence trust in human–robot interaction. Perceived autonomy increased sincerity ratings but had limited impact on other trust measures or behaviours. Instead, pre-existing negative attitudes toward robots—particularly negative emotional reactions and social influence concerns—were stronger predictors of trust outcomes. These findings highlight the complex nature of HRI trust, shaped by both system design and user traits. To advance this understanding, future work should adopt more diverse and ecologically valid study designs, incorporating longitudinal methods and broader participant samples. In particular, qualitative materials already collected—such as open-ended responses and video recordings—offer rich opportunities for further analysis of interaction dynamics, user reasoning, and non-verbal behaviour. Leveraging these insights may reveal subtleties in trust development not captured by quantitative metrics alone.

References

1. Amirabdollahian, F., Dautenhahn, K., Dixon, C., Eder, K., Fisher, M., Koay, K.L., Magid, E., Pipe, T., Salem, M., Saunders, J., et al.: Can you trust your robotic assistant? Social Robotics (2013)
2. Aroyo, A.M., De Bruyne, J., Dheu, O., Fosch-Villaronga, E., Gudkov, A., Hoch, H., Jones, S., Lutz, C., Sætra, H., Solberg, M., et al.: Overtrusting robots: Setting a research agenda to mitigate overtrust in automation. Paladyn, Journal of Behavioral Robotics **12**(1), 423–436 (2021). https://doi.org/10.1515/pjbr-2021-0029
3. Broadbent, E., Stafford, R., MacDonald, B.: Acceptance of healthcare robots for the older population: Review and future directions. International journal of social robotics **1**, 319–330 (2009). https://doi.org/10.1007/s12369-009-0030-6
4. Carpinella, C.M., Wyman, A.B., Perez, M.A., Stroessner, S.J.: The robotic social attributes scale (rosas) development and validation. In: Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction. pp. 254–262 (2017). https://doi.org/10.1145/2909824.3020208
5. Chen, M., Nikolaidis, S., Soh, H., Hsu, D., Srinivasa, S.: Trust-aware decision making for human-robot collaboration: Model learning and planning. ACM Transactions on Human-Robot Interaction (THRI) **9**(2), 1–23 (2020). https://doi.org/10.1145/3359616
6. Choi, J.J., Kim, Y., Kwak, S.S.: The autonomy levels and the human intervention levels of robots: The impact of robot types in human-robot interaction. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication. pp. 1069–1074. IEEE (2014). https://doi.org/10.1109/ROMAN.2014.6926394
7. Costa, P.T., McCrae, R.R.: A five-factor theory of personality. Handbook of personality: Theory and research **2**(01), 1999 (1999)
8. Cui, J., Tosunoglu, S., Roberts, R., Moore, C., Repperger, D.W.: A review of teleoperation system control. In: Proceedings of the Florida conference on recent advances in robotics. pp. 1–12. Citeseer (2003). https://doi.org/10.1016/S0004-3702(98)00023-X
9. Fisher, R.A.: Statistical methods for research workers. In: Breakthroughs in statistics: Methodology and distribution, pp. 66–70. Springer (1970). https://doi.org/10.1007/978-1-4612-4380-9_6
10. Gosling, S.D., Rentfrow, P.J., Swann Jr, W.B.: A very brief measure of the big-five personality domains. Journal of Research in personality **37**(6), 504–528 (2003). https://doi.org/10.1016/S0092-6566(03)00046-1
11. Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y., De Visser, E.J., Parasuraman, R.: A meta-analysis of factors affecting trust in human-robot interaction. Human factors **53**(5), 517–527 (2011). https://doi.org/10.1177/0018720811417254
12. Holthaus, P., Fallahi, A., Förster, F., Menon, C., Wood, L., Lakatos, G.: Agency Effects on Robot Trust in Different Age Groups. In: International Conference on Human-Agent Interaction (HAI 2024). ACM, Swansea, UK (2024). https://doi.org/10.1145/3687272.3690903
13. John, O.P., Srivastava, S., et al.: The big-five trait taxonomy: History, measurement, and theoretical perspectives (1999)
14. Kaelbling, L.P., Littman, M.L., Cassandra, A.R.: Planning and acting in partially observable stochastic domains. Artificial intelligence **101**(1-2), 99–134 (1998)
15. Lewis, M., Sycara, K., Walker, P.: The role of trust in human-robot interaction. Foundations of trusted autonomy pp. 135–159 (2018). https://doi.org/10.1007/978-3-319-64816-3_8

16. Luck, M., d'Inverno, M., et al.: A formal framework for agency and autonomy. In: Icmas. vol. 95, pp. 254–260 (1995)
17. Malle, B.F., Scheutz, M., Arnold, T., Voiklis, J., Cusimano, C.: Sacrifice one for the good of many? people apply different moral norms to human and robot agents. In: Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction. pp. 117–124 (2015). https://doi.org//10.1145/2696454.2696458
18. Malle, B.F., Ullman, D.: A multidimensional conception and measure of human-robot trust. In: Trust in human-robot interaction, pp. 3–25. Elsevier (2021)
19. Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics pp. 50–60 (1947)
20. Nomura, T., Kanda, T., Suzuki, T., Kato, K.: Prediction of human behavior in human–robot interaction using psychological scales for anxiety and negative attitudes toward robots. IEEE transactions on robotics **24**(2), 442–451 (2008). https://doi.org/10.1109/TRO.2007.914004
21. Pan, J., Eden, J., Oetomo, D., Johal, W.: Exploring the effects of shared autonomy on cognitive load and trust in human-robot interaction. arXiv preprint arXiv:2402.02758 (2024). https://doi.org/10.48550/arXiv.2402.02758
22. Robinette, P., Li, W., Allen, R., Howard, A.M., Wagner, A.R.: Overtrust of robots in emergency evacuation scenarios. In: 2016 11th ACM/IEEE international conference on human-robot interaction (HRI). pp. 101–108. IEEE (2016). https://doi.org/10.1109/HRI.2016.7451740
23. Salem, M., Lakatos, G., Amirabdollahian, F., Dautenhahn, K.: Towards safe and trustworthy social robots: ethical challenges and practical issues. In: Social Robotics: 7th International Conference, ICSR 2015, Paris, France, October 26-30, 2015, Proceedings 7. pp. 584–593. Springer (2015). https://doi.org/10.1007/978-3-319-25554-5_58
24. Salem, M., Lakatos, G., Amirabdollahian, F., Dautenhahn, K.: 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. pp. 141–148 (2015). https://doi.org/10.1145/2696454.2696497
25. Schaefer, K.E., Chen, J.Y., Szalma, J.L., Hancock, P.A.: A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. Human factors **58**(3), 377–400 (2016). https://doi.org/10.1177/001872081663422
26. Shahrdar, S., Menezes, L., Nojoumian, M.: A survey on trust in autonomous systems. In: Intelligent Computing: Proceedings of the 2018 Computing Conference, Volume 2. pp. 368–386. Springer (2019). https://doi.org/10.1007/978-3-030-01177-2_27
27. Shapiro, S.S., Wilk, M.B.: An analysis of variance test for normality (complete samples). Biometrika **52**(3-4), 591–611 (1965)
28. Spearman, C.: The proof and measurement of association between two things. (1961). https://doi.org/10.1037/11491-005
29. Ueno, T., Sawa, Y., Kim, Y., Urakami, J., Oura, H., Seaborn, K.: Trust in human-ai interaction: Scoping out models, measures, and methods. In: CHI Conference on Human Factors in Computing Systems Extended Abstracts. pp. 1–7 (2022). https://doi.org/10.1145/3491101.3519772
30. Ullman, D., Malle, B.F.: Mdmt: Multi-dimensional measure of trust (2019)

31. Waytz, A., Epley, N., Cacioppo, J.T.: Social cognition unbound: Insights into anthropomorphism and dehumanization. Current Directions in Psychological Science **19**(1), 58–62 (2010). https://doi.org/10.1177/0963721409359
32. Waytz, A., Gray, K., Epley, N., Wegner, D.M.: Causes and consequences of mind perception. Trends in cognitive sciences **14**(8), 383–388 (2010). https://doi.org/10.1016/j.tics.2010.05.006