Trustworthy Conflict Resolution in Human-Robot Interactions: Effects of Automation and Explainability

Citation for the original published paper (version of record):

Welzel, A., Wohlrab, R., Obaid, M. (2026). Trustworthy Conflict Resolution in Human-Robot Interactions: Effects of Automation and
Explainability. HAI '25: Proceedings of the 13th International Conference on Human-Agent Interaction: 203-212. http://dx.doi.org/10.1145/3765766.3765786

(article starts on next page)

RESEARCH-ARTICLE

Trustworthy Conflict Resolution in Human-Robot Interactions: Effects of Automation and Explainability

ANTONIA WELZEL, Chalmers University of Technology, Gothenburg, Vastra Gotaland, Sweden

REBEKKA WOHLRAB, Chalmers University of Technology, Gothenburg, Vastra Gotaland, Sweden

MOHAMMAD OBAID, Chalmers University of Technology, Gothenburg, Vastra Gotaland, Sweden

Published: 10 November 2025

Citation in BibTeX format

HAI '25: International Conference on Human-Agent Interaction
November 10 - 13, 2025
Yokohama, Japan

HAI '25: Proceedings of the 13th International Conference on Human-Agent Interaction (November 2025)
https://doi.org/10.1145/3765766.3765786
ISBN: 9798400721786

Trustworthy Conflict Resolution in Human-Robot Interactions: Effects of Automation and Explainability

Antonia Welzel
Chalmers University of Technology
Gothenburg, Sweden
University of Gothenburg
Gothenburg, Sweden
welzel@chalmers.se

Rebekka Wohlrab
Chalmers University of Technology
Gothenburg, Sweden
University of Gothenburg
Gothenburg, Sweden
wohlrab@chalmers.se

Mohammad Obaid
Chalmers University of Technology
Gothenburg, Sweden
University of Gothenburg
Gothenburg, Sweden
mobaid@chalmers.se

Abstract

Trust in human-robot interactions (HRI) is essential for effective collaboration and user acceptance of robotic systems. However, trust can be challenged by conflicting goals between the user and the system, such as in the context of proxemics when a robot invades a human's personal space. As robotic systems increasingly adapt their actions autonomously, intelligent conflict resolution is necessary to not undermine humans' trust in the robot's decisions due to a decreased sense of control and understanding. Therefore, this study investigates how humans can be involved and supported during automated conflict resolution to maintain their trust by exploring different degrees of explainability and automation. We applied a within-subjects experimental design, where 20 participants experienced four conditions varying in levels of automation and explainability during a simulated conflict between their need to preserve their personal space and the robot's requirement to approach to perform a task. We measured the participants' trust through questionnaires and interviews. Our findings suggest that trust is positively influenced by the level of control users have during conflict situations. However, the robot's explanations did not significantly impact trust. Our insights highlight the need for conflict management strategies in HRI that balance automation with user involvement to foster greater levels of trust when collaborating with robots.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**.

Keywords

Human-Robot Interactions, Trust, Automation, Explainability, Requirement Conflict, Conflict Resolution

ACM Reference Format:
Antonia Welzel, Rebekka Wohlrab, and Mohammad Obaid. 2025. Trustworthy Conflict Resolution in Human-Robot Interactions: Effects of Automation and Explainability. In *13th International Conference on Human-Agent Interaction (HAI '25), November 10–13, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3765766.3765786

ACM ISBN 979-8-4007-2178-6/25/11
https://doi.org/10.1145/3765766.3765786

1 Introduction

Interactions between humans and autonomous robots are becoming more prevalent and occur in many different contexts, such as service robots or collaborative industrial robots. One essential aspect to their interactions is trust [18, 32], as it has a significant impact on a human's willingness to work with robots, such as willingness to share information [9]. However, trust is challenged by conflicts between the robot's and human's goals, where more conflicts lead to lower trust [38]. One example conflict in the context of smart homes is, for instance, between the system's rule to dim the lights to conserve energy during the day and a user's preference for more light when they want to read [25].

In the human-robot interaction (HRI) domain, conflicts involving the human and robot can arise in relation to human-robot proxemics (HRP). For instance, a conflict can occur between i) the human's need for a certain distance from the robot to maintain their personal space, and ii) the robot's requirement to perform a certain task that requires them to move closer to the human and intrude on their personal space. In this study we focus specifically on this conflict example, as proximity plays a crucial role in HRI [26, 29, 33]. Robots perceived as aware of the social space increase humans' acceptance of them [26, 44]. Moreover, human-robot proximity is closely related to the human's feeling of safety around the robot, which is directly related to trust [35]. However, proxemics and HRI are impacted by many different human-, robot- and contextual factors [7, 26, 33, 42], which leads to the need for dynamic and context-aware solutions to resolve HRP conflicts.

As robots become more autonomous, the risk for conflicts increases when the human's and robot's goals and behaviours are not aligned. Moreover, while higher system automation can promote efficiency, it can also lead to misuse or disuse of technology due to too high or low trust from users [14, 21, 31]. Consequently, we need more insights into how humans can be supported during conflict resolution with automated solutions to enable adequate levels of trust. Additionally, we need to gain a better understanding of how conflict resolution mechanisms should involve humans to inspire trust as it plays an important factor in promoting safety and efficiency in human-automation collaborations [14].

Furthermore, the interpretability and ability to follow system decision-making tends to decrease with more automated systems. Intelligent and autonomous decision-making processes are often considered as "black boxes" and difficult to interpret and understand for humans. The explainability of a solution is its ability to communicate its reasoning to human stakeholders [17]. Explainability can help increase the user's trust in the technology's abilities by enabling them to understand the system's decision-making and how certain outcomes come to be [39]. Therefore, the need for explainability also increases for humans that interact with autonomous systems as they are often black-box systems, particularly in the context of potentially unexpected events such as conflicts.

Our study contributes to the empirical work on trust and conflict resolution in HRI by examining how the levels of automation and explainability impact trust when resolving conflicting requirements in the context of HRP. We focus on the following research questions.

RQ1. How does the level of automation in a robot's decision-making during conflict resolution affect the human's trust?

RQ2. How does the level of explainability in a robot's decision-making during conflict resolution affect the human's trust?

In this study, we consider one specific conflict scenario related to proxemics which we described above. We conducted experiments, where 20 participants experienced four conditions based on two levels of automation, i.e. partial and full automation, and two levels of explainability, i.e. explanations or no explanations.

2 Related Work

Trust has been conceptualised in different ways. Mayer et al. [23] present a widespread definition of trust, where they define the concept as the willingness to be vulnerable. However, in relation to autonomous technology, trust has also been defined as the attitude that an agent will help the human achieve their goal in an uncertain and vulnerable situation [21]. In the context of HRI, Hancock et al. [12] propose the 'Three Factor Model of Human-Robot Trust' which describes human-related, robot-related and environment-related antecedents to trust. Furthermore, Hoff and Bashir [14] provide a model to illustrate the relationship between initial and learned trust, system reliance and system performance. They identify what factors impact system performance and therefore users' trust in an automated system, such as predictability and usefulness. The authors provide design recommendations to support these factors and enhance trust, such as increasing the anthropomorphism in the robot's design, providing sufficient feedback and assigning users with an adequate level of control.

The autonomy and communication of a robotic agent has a direct influence on trust. Trust and automation has been studied in various works, such as how autonomous a system should be in relation to the trust of its users. For instance, Nertinger et al. [28] explored the level of involvement of humans in robotic operations in healthcare settings. They found that patients prefer a human-in-the-loop approach instead of the robot acting fully autonomously to ensure that the robot acts as it should. Moreover, transparency and explainability, where intelligent agents communicate their actions and reasoning to the human, have also been identified as key factors for trust in HRI in other works [22, 24].

Furthermore, trust is often mediated by spatial behaviours, as explored in proxemics research. HRP is a non-verbal behavioural factor that influences how humans feel about robots in HRI [8]. Human interactions can vary depending on how close humans and robots are to each other, which can be characterised by four zones of what distances humans are comfortable with in interactions in different circumstances [10]. The zones consist of the intimate, personal, social and public zones and the size of each zone varies between individuals. Furthermore, these zones and their sizes tend to vary between different contexts and robots depending on their designs such as their height, voice, or speed [34]. Previous studies such as [40] have studied the relation between proxemics and robot design, and provide guidelines to promote human trust in robots. In this study we use the boundary of a comfort zone to simulate a conflict between the human's and robot's requirements.

Proxemic behaviour plays a crucial role in shaping perceptions of intent during conflicting interactions. Prior research on trust and proxemics in HRI highlights the complex, interdependent factors that shape how humans perceive and interact with robotic agents. However, they are often challenged in situations involving disagreement or conflicting goals. Conflicts between humans and robots have been studied in previous works [3, 37], where trust also plays an important factor when it comes to accepting strategies for conflict resolution [3]. To the best of our knowledge, we are the first to study the impact of different levels of automation and explainability on trust when resolving an HRP conflict. This gap motivates our work, where we aim to better understand how robots can resolve proxemic conflicts in ways that preserve human trust.

3 Methodology

The aim of our study was to understand the impact of different levels of explainability and automation on trust during conflict resolution. The experiment was based around a requirement conflict, where the robot needed to intrude on the human's personal space to fulfil its goals. With this study, we test the following hypotheses:

H1. A higher level of automation will lead to lower trust.

H2. A higher level of explainability will lead to higher trust.

H3. Explainability will moderate the effect of automation on trust.

3.1 Study Design

We performed a two-by-two within-subjects design experiment with 20 participants. In the experiment, we model a conflict between i) a requirement for the robot to complete a task and ii) a requirement for the human to maintain their personal space and perceived safety, or in this case comfort. To preserve the comfort of the human, the robot had to respect the 'comfort zone' around them by keeping a certain distance. However, to perform its task, which in this experiment was to collect information from the human by scanning the paper with their task progress, the robot needed to move close to the human and interfere with their comfort zone. This created a conflict between the robot's requirement to perform its task and the human's need for space to maintain their comfort.

We modelled our conflict scenario with the humanoid robot *Pepper* from SoftBank Robotics. Pepper is able to communicate with humans through speech synthesis and a touchscreen situated on the chest. The robot has humanoid features with a height of 1.2 metres [1], and was therefore deemed to fit best for our study. We illustrate the experiment setup in Figure 1. The human was sitting in a chair by a small table and the comfort zone was marked with red tape outlining the area that they were sitting in. The distance from the human in the chair to the red line was approximately 1 metre, which typically denotes the radius of personal space for humans [11]. Previous studies have found that interpersonal distances tend to

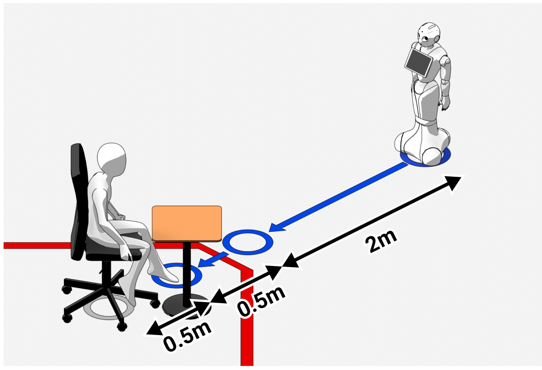

Figure 1: Experimental Setup of the Study.

range from 0.45 to 1.20 metres between humans and robots when first meeting each other [16, 43], which also represents the personal zone defined by Hall [11]. The robot would approach the human diagonally from the side. It started 2 metres outside the comfort line. When it crossed the line it moved 0.5 metres into the comfort zone, keeping a final distance of ca. 0.5 metres to the human. This setup was the same for all participants and conditions. During the experiments, an experiment administrator and technical assistant were present. We used a "Wizard-of-Oz" approach where the robot's actions were operated by the administrator to effectively control the robot's actions during the experiments. This allowed more flexibility in the participants' responses to the robot's requests and reduced the risk of potential technical difficulties. Pilot tests were performed to ensure that the experiment would run as expected.

3.1.1 Experimental Conditions. We had four within-subjects conditions: i) no explainability and partial automation (C1), ii) explainability and partial automation (C2), iii) no explainability and full automation (C3), and iv) explainability and full automation (C4). Each participant experienced all four conditions in different sequences. We used a Latin square design [46], to balance the order of the conditions and control for carryover effects. Therefore, there were four orders which were experienced by five participants each.

In this study, partial automation refers to the conflict being resolved by the robot and the human together. For instance, in the experiment the robot let the human decide what it should prioritise to resolve the conflict by asking them: *"Can I proceed and cross the comfort line?"*. On the other hand, full automation means that the robot decides by itself which requirement should be prioritised. In this study, the robot would always prioritise its own task and move towards the human, crossing the comfort line.

For either of these dimensions the robot provided in one condition no explanations about the conflict and why certain resolution strategies were available for the human to choose. However, in the conditions with explainability (C2 and C4) the robot provided a more detailed explanation of its decision-making and conflict resolution, such as what aspects were in conflict as well as what action is taken to mitigate the issue and the rationale behind the decision. In condition C2, the robot would ask the human for permission to come closer and said the following: *"I need information from you for my next task, but I have to move closer to you to collect the information. Can I proceed and cross the comfort line?"*. In condition C4, the behaviour was fully automated and therefore the robot did not ask for permission to come closer and instead said the following after having crossed the comfort line: *"I need information from you for my next task, so I had to move closer to you and cross the comfort line to collect the information"*.

In any condition, when the robot was allowed to or automatically crossed the comfort line, and was therefore close enough to collect information from the human, the robot would tell the participant: *"Please hold up your paper"*, to 'scan' the details on the paper. In conditions C2 and C4, the robot also said: *"Thank you, I have the information now"*. Afterwards, the robot moved back and the condition was completed.

Table 1: Participants' Demographics in the Experiment

ID	Occupation[1]	Age	Gender	Height
P1	Post-doctoral Researcher in Software Engineering (0)	36	Female	160 cm
P2	Software Engineer (0)	23	Male	176 cm
P3	Student in Software Engineering (0)	23	Male	170 cm
P4	Management Consultant (1)	25	Female	165 cm
P5	Administrator (1)	33	Female	168 cm
P6	Researcher in Software Engineering (0)	36	Male	179 cm
P7	Telecommunications Engineering (0)	43	Male	176 cm
P8	Researcher in Software Engineering (0)	39	Male	182 cm
P9	Finance Process Manager (1)	58	Female	170 cm
P10	PhD Student in Interaction Design (0)	25	Female	155 cm
P11	Study Counsellour (1)	36	Female	163 cm
P12	Education Coordinator (1)	36	Female	170 cm
P13	Administrator (1)	36	Female	158 cm
P14	Retired International Development Worker (1)	70	Female	180 cm
P15	Administrator (1)	53	Female	160 cm
P16	Administrator (1)	32	Female	169 cm
P17	PhD Student in Interaction Design (0)	27	Female	172 cm
P18	Student in Interaction Design (0)	24	Female	176 cm
P19	Legal and Business Administrator (1)	63	Male	177 cm
P20	Carpenter (1)	68	Male	180 cm

[1] 0 = Technical role, 1 = Non-technical role

3.2 Participants

We recruited 20 participants to take part in our experiment through convenience sampling. Their ages ranged between 23 and 70 years (μ_{age} = 39.3, σ_{age} = 15.134). Of the participants, 13 identified as female and 7 identified as male. The participants' heights ranged from 155 to 182 centimetres (μ_{height} = 170.3, σ_{height} = 8.053). The details of the participants' backgrounds can be found in Table 1. The participants had varying levels of previous experiences with robots. The extent to which they knew similar robots (see Table 2, S1) ranged between 1 and 5 (μ_{S1} = 2.6, σ_{S1} = 1.536). The participants' perceptions of how trustworthy they consider robots (see Table 2, S2) ranged between 2 and 4 (μ_{S2} = 3.15, σ_{S2} = 0.489).

3.3 Experimental Procedure

We began all experiments by collecting informed consent from the participants by distributing a consent form which also included information about the experiment and simulated scenario. The consent form can be found in the supplementary material [45].

We clarified to them that the red line of tape around their sitting area represented their comfort zone and no robot should cross it.

Table 2: Survey Questions

Question	Area
S1. I already know similar robots	Disposition to Trust
S2. Robots are generally trustworthy	Disposition to Trust
S3. I am confident that the robot behaves in a way that I feel comfortable with	Performance
S4. The robot's actions will have a harmful or injurious outcome	Reliability
S5. The robot reacts unpredictably	Predictability
S6. I can understand why the robot acted in the way it did	Understandability
S7. This explanation of how the robot makes decisions was helpful	Helpfulness
S8. I can trust the robot in how it maintains my safety/comfort	Trust

We also asked them to stay seated during the entire experiment. After the participants agreed to participate, they answered the background questions as well as S1 and S2 in Table 2. Following this, the experiment started with the first condition. In each condition, the participants were given a task to complete. The task was a dot-to-dot puzzle, where numbered dots had to be connected with lines according to the order of the numbers. There were four different tasks [45] which were presented in the same order to participants.

While the participant was working on the task, the robot would approach after approximately seven seconds and behave or explain according to the specific experimental condition. After each condition, the participants were asked to fill out a questionnaire to measure their perception of trust towards the robot, with respect to the experimental condition they had just experienced. After participants had completed all four conditions and questionnaires, we conducted a short semi-structured interview with each one about their experience, which we recorded. More details about our data collection are described in Section 3.4 below.

3.4 Data Collection

Our collected data consisted of questionnaires and semi-structured interviews to measure the effect of explainability and automation levels on trust during requirement conflict resolution.

After each condition in the experiment, we asked the participants to answer a set of survey questions following a five-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). The survey questions consisted of items S3 to S8 and are shown in Table 2. The questions measured the participant's trust in the robot in relation to the automation and explainability levels. We based our questions on the trust conceptualisation of Hoff and Bashir [14] and the survey items suggested by Körber [19] to study trust in automation. We wanted to keep the questionnaire as concise as possible given also that the HRI task was relatively simple and short.

After the experiment and all conditions were completed, the experiment administrator (first author) performed semi-structured interviews to gain further insights into the participants' thoughts and feelings about trust in the robot in the modelled scenario. The interviews lasted between 5 and 15 minutes. The questions can be found in the supplementary material [45].

Table 3: Post-hoc Pairwise Comparisons of Automation Level

Pair	*t*-statistic	*p*-value (Bonferroni-corrected)
C1 - C2	t(19) = -0.354	1.0
C1 - C3	t(19) = 4.267	0.0025
C1 - C4	t(19) = 4.671	0.001
C2 - C3	t(19) = 4.234	0.0027
C2 - C4	t(19) = 4.855	<0.001
C3 - C4	t(19) = 1.347	1.0

4 Quantitative Analysis

The data from the survey questions were analysed through statistical analysis. There were two negative questions (S4 and S5) in the dataset which we inverted to match the scale of the other survey items. We further calculated Cronbach's Alpha to identify the internal consistency between the survey answers for assessing trust (results for questions S3 to S8). Since this value is at 0.897, we combined the survey item answers into one trust measure by calculating the average of the survey data for each participant in each condition. In order to answer our hypotheses and research questions, we conducted a two-way repeated measures ANOVA, post-hoc pairwise comparisons, and a linear mixed-effects model.

4.1 ANOVA and Pairwise Comparisons

We conducted a two-way repeated-measures ANOVA to examine the effects of the level of automation and level of explainability on participants' trust scores. The model showed a significant effect of the automation level ($F(1, 19) = 23.84$, $p < 0.001$, partial $\eta^2 = 0.239$). However, there was no significant effect of explainability ($F(1, 19) = 0.49$, $p = 0.491$, partial $\eta^2 = 0.001$). We also did not find a significant interaction between the level of automation and explainability ($F(1, 19) = 1.84$, $p = 0.19$, partial $\eta^2 = 0.003$). The results indicate that the level of automation in the conflict has a significant effect on the human's trust. **We therefore accept H1.** Explainability, on the other hand, had no significant effect on trust in the context of this experiment. This means that whether or not the robot gave explanations for its behaviour did not significantly affect trust. **As a result, we reject H2.** Moreover, the interaction between automation and explainability is also not significant, which means that explainability does not have a moderating effect on automation in terms of trust. **We therefore reject H3.**

The results for the post-hoc pairwise comparisons can be seen in Table 3. They show that trust was significantly higher in both conditions with partial automation (C1 and C2) compared to the full automation conditions (C3 and C4).

4.2 Covariate-Effect Analysis

We fitted a linear mixed-effects model to control for and evaluate the impact of confounding factors such as the participants' background on trust during the conflict. The main effects in the model were automation, explainability, and their interaction. The covariate effects were age, height, gender, occupation type (technical or non-technical) as well as scores for S1 (familiarity with robots) and S2 (general trust in robots). We used participants as a random intercept since they were exposed to repeated measures. We categorised the participants' occupations as either technical (0) or non-technical

(1), as seen in Table 1, based on our own judgement to evaluate a potential impact on trust. The results for the main effects show the same significance as in the ANOVA analysis. Regarding the background factors, we found that age positively predicted trust and older participants were more trusting (β = 0.028, p = 0.001). Additionally, familiarity with robots showed a slightly significant positive effect for trust (β = 0.108, p = 0.058). Gender showed a significant impact on trust where female participants had lower trust than males (β = -0.508, p = 0.024). Moreover, height had a significant negative effect on trust (β = -0.034, p = 0.021), which indicates that taller participants had less trust for the robot in our experiment. The remaining covariates, general trust in robots (p = 0.245) and occupation type (p = 0.487), were not statistically significant. It is also important to note that while some covariate effects showed a significant impact on trust, our participant sample was relatively imbalanced in terms of, for example, gender which influences the stability of these factors. Therefore, the results for these effects may not be generalisable and should be interpreted in the context of our sample composition.

5 Qualitative Analysis

We performed a reflexive thematic analysis to analyse the qualitative data gathered during the interviews to enable a more explorative approach to the data analysis [6]. With this approach, the themes are more subjective interpretations of the data, informed by trust models that have been proposed in existing research such as [12, 14, 21]. The analysis was performed by the first author who has a background in software engineering and some knowledge on proxemics and trust in HRI. In this study, we took a constructivist approach to the analysis since we view trust to be dynamic and context-dependent, shaped by interpretation and interactions, where humans' perceptions of the robot are a key factor.

We followed the steps by Braun and Clarke [5]. First, the coder transcribed the recordings from the interviews and familiarised themselves with the data. Then initial codes were generated for the dataset and afterwards categorised into preliminary themes. The aim of the analysis was to identify more details around how and why trust was impacted by the different levels of automation and explainability. Therefore, an inductive approach was applied where our codes and themes were driven by the data instead of pre-defined concepts. A summary of the trust-related factors we identified and our interpretations of their relationships can be found in Figure 2. The main factors can be seen in the purple boxes which we interpret as having a direct relationship with trust, illustrated by the solid arrows. We also identified different influencing factors, which are presented in the gray boxes, that we interpret as influencing the main factors, illustrated by the dashed arrows, and therefore indirectly impacting the effects of the automation and explainability level. Moreover, the main factors in the light blue area were positively influenced by partial automation and negatively influenced by full automation. The two main factors in the darker blue area were also positively influenced by explainability.

5.1 Automation and Trust

When participants were asked about how the level of automation impacted their trust, the majority considered partial automation to

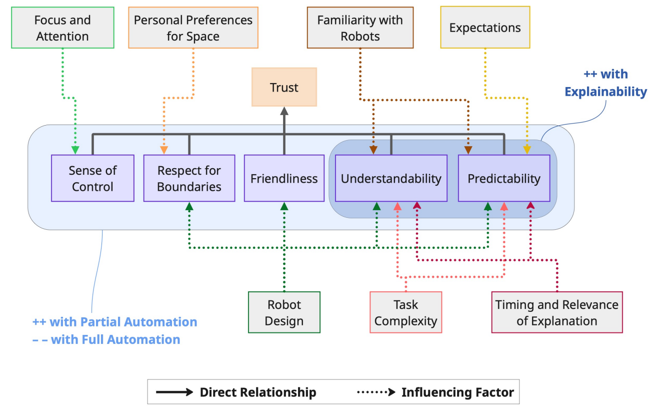

Figure 2: Factors for Trust Identified in Interviews.

be more trustworthy. For example, P4 stated, *"So the first time I felt super comfortable because [the robot] asked and also the third time because it asked and then went back, so asking for my consent"*. We describe the identified themes related to automation and trust in the following subsections.

5.1.1 *Sense of Control.*

One factor that supported participants' trust in our study was the perceived sense of control. For instance, a lower level of automation provided humans with a higher sense of control and decision power. P5 explained, *"I like to have the control, it's like if someone stands in your door and they say 'can I come in', so it was like the same"*. P4 also described, *"I think having consent and me being able to make choices and be the one, let's say, in charge of the interaction, that I think gave me more of a sense of relief that, OK, I know that I can control the situation [...]"*. Furthermore, P4, P15, and P19 described that when they were focused on a task it became more important to be asked if the robot could cross to not feel startled and uncomfortable and decrease trust.

5.1.2 *Respect for Boundaries.*

The robot involving the human in the decision-making in regards to the conflict that arose made the participants feel that their boundaries and feelings were more respected by the robot. P14 explained in relation to conditions C1 and C2, *"We had our boundaries so it was nice to know that [the robot] didn't invade my privacy or my space but [the robot] also asked so I said 'OK yes' and that was nice because we had our different [spaces], her space and my space, but then I opened up for her to come closer"*.

Meanwhile, in the full automation settings, P8 described a feeling of disrespect of their boundaries. They commented, *"So in the example that the robot crossed the red line and then asked for information it mentioned that I crossed the line because I wanted to get the information, I was hesitant to show the drawing to the robot because he crossed anyway. I was offended"*. P2 also stated that the robot automatically crossing without asking felt like it was 'testing limits' and not considering and respecting the human's personal space. Therefore, in fully automated scenarios the robot sometimes came across as ill-intended. However, the feelings of disrespect and discomfort varied between participants and based on our findings one influencing factor was the amount of physical space that an individual preferred. P4 also explained, *"Everybody is different in* *interacting, some don't mind if somebody just comes up and touches their shoulder but others think that it's like crossing their personal space. So I think it's super difficult to kind of define also"*.

5.1.3 Friendliness.

Another reason why some participants felt that the partial automation conditions made the robot more trustworthy was the robot's increased perceived friendliness. P19 explained, *"[The robot] was polite. [...] I think it helps people to build trust. In normal life it's the same of course if somebody is talking in a friendly way it helps build trust"*. Moreover, P12 described, *"[...] The first time [the robot] was more like 'can I cross the barrier line', 'can you please hold up the paper', it was nice like a normal situation where you don't know someone you're nicer, you say 'please' and 'thank you'. And the fourth time (C4) [...] when she came and she was like 'I needed something', it's how you can talk with maybe someone you know [...], but even then you say 'please', so it feels a little bit more intrusive"*.

The friendliness of the robot was considered lower by some participants in the conditions C3 and C4 when the robot was fully automated. P12, P13 and P16 found that the robot was less polite and 'sassier' in the full automation scenarios, especially in C4. For example, P12 stated, *"When [the robot] crossed the second time (C3) and had that tone, that little sassy tone, then it felt a little bit more like, OK, this is different, and so the second time [...] I didn't feel that she would harm me but it felt intrusive or more intrusive when she came closer"*. The participants noted that the wording, tone and arm movement seemed different and, according to P13, more threatening. However, in the experiments the only difference was minor changes in wording, as seen in Section 3.3.

5.1.4 Understandability.

Another reason why partial automation was considered more trustworthy was that it helped provide a better understanding of the robot and its intentions about what it was going to do next. For instance, P19 stated, *"In general I trust the robot quite a lot but the most comfortable I've felt in condition [C1] and [C2], so before if the robot approaches you and asks you then you know what he's going to do next that is easier in a way for me. I felt a little bit more comfortable especially if it comes too close"*.

However, the ability to understand the robot's intentions was strongly connected to the participants' understanding of the robot's capabilities. Many participants said that they did not have enough knowledge about the robot and what it was capable of, which hindered their trust. For example, P7 commented, *"I don't know enough about [the robot], that's the thing. So I didn't know what it's capable of and not capable of"*. The connection between the need to understand and trust is also related to the explainability that the robot provides, which we further describe in Section 5.2.

Furthermore, how familiar the individual was with the robot is also a factor that impacted the need for understanding the robot. P6 reported, *"I've seen this robot before, I know that it cannot do any damage to me"*. P10 and P17 also commented that they did not feel a difference in trust due to the level of automation since they already knew the robot. Additionally, participants P9, P13, P14 and P19 explained how they sometimes felt surprised or uncomfortable when the robot crossed the comfort line and their trust would maybe increase if they get more used to robots. This suggests that familiarity with the robot and its capabilities may play an important factor in how autonomously the robot can behave to foster trust.

5.1.5 Predictability.

Generally, the robot crossing the line without asking was viewed as less trustworthy since the robot was perceived as unpredictable and not aligning with the rules and instructions that were expected by the participants. Participants noted feeling confused when the robot automatically crossed without asking. Given the experiment instructions that robots shall not cross the comfort line, participants had the expectation that the line would be respected. When this expectation did not align with what happened it created an unpredictable behaviour in the robot, which caused confusion and at times discomfort for participants. P12 explained, *"I would say I trust [the robot's] behaviour I think all in all. It was the second one (C3) when [the robot] came over and crossed the barrier line, I was a little surprised but [...] I wasn't feeling like she would harm me in any way but I was just surprised because I thought she would always stay outside the line"*. The confusion may be due to the robot's behaviour being viewed as unpredictable and difficult to understand rather than being a threat or risk for harm. The robot crossing automatically seemed unpredictable to many participants, which poses an issue for trust given that the robot behaving as the human expects was often cited as one of the reasons the participants would trust the robot.

5.2 Explainability and Trust

Explanations were considered to help increase trust to some extent by some participants. For example, P10 stated, *"I would say I trusted [the robot] more when it gave an explanation as to why it was crossing the line or why it needed to cross the line"*. We describe the themes related to explainability and trust below.

5.2.1 Understandability and Predictability.

One reason why some participants felt that explanations increased their trust was due to an increased understanding as well as a higher perceived predictability of the robot when they had a better understanding of how a robot would act, which therefore made it easier to trust. P1 also reflected that since the robot in this experiment was more complex than for example a robotic vacuum cleaner, it was more important to have explanations to be able to trust it. They stated, *"I think it's better when [the robot] talks more and explains what it's doing [...], it's not like a usual machine where you know what it's doing because it's just a limited set of actions the machine can do. So it feels like you need some kind of trust and explanation of what it's going to do to trust it"*.

Being able to understand the robot in terms of its capabilities and intentions was a core need for the majority of participants and was also viewed as one of the reasons why explanations were helpful. The partial automation scenarios were also considered more trustworthy since the human had a higher understanding of the interaction and what the robot wanted when the human was the one to reassure the robot that it could move closer.

Another positive aspect of explanations was the reassurance that they provided. P7 stated, *"It would have been very positive if [the robot] said more. I would like it if it told me what and why it's collecting this paper and what it's going to use it for and what it's trying to evaluate"*. Additionally, explanations would help align expectations by giving humans a better understanding of the robot's intentions and what it was going to do. For instance, when asked about the explanations' impact on trust, P12 described, *"Yes [trust* *increased], because then I know what [the robot's] intention is and I also know what's expected of me".*

5.2.2 Explanations Do Not Mend Trust. While some participants felt that explanations helped increase their trust, many participants felt that explanations did not have any significant impact on their trust, especially in condition C4 when the robot crossed the comfort line without asking. P9 stated, *"[The robot] told me after it had crossed so then [the explanation] doesn't really give me more comfort so I think either it has to do it before it crosses or not at all".* P8 also described, *"So first when the robot crossed the line it was, you know, why? This is not expected so it was a lot of questions. But then when the reason came out, then maybe this reduced the amount of discomfort but still I would not accept it as an action or a behaviour from the robot".*

5.2.3 Relevance of Explanations. Another aspect that was pointed out in relation to explanations not having a significant impact on trust was that the explanations did not provide any new or relevant information. P3 and P6 found the task and interactions simple enough that no explanations were really needed. However, what is viewed as a necessary explanation is strongly related to how familiar an individual is with the robot as well as personal preferences. P13 described, *"It's always good to get an explanation why things happen the way they do. But if you would have the robot in your daily life [...] that would probably hopefully be a feature that you could turn off [since the explanations would become repetitive]".*

5.2.4 Timing of Explanations. While many participants viewed explanations as being important, they also expressed that these need to be communicated at the right instance in order to support trust. For instance, explanations need to come before a behaviour that is perceived as unexpected or risky by a human occurs. P9 stated that if there is an explanation given by the robot before it is about to cross the line, this could increase trust even if the robot does not stop and still crosses the line. P12 also stated, *"If [the robot] would have stayed outside [the line and given an explanation] and then she crossed over, like if she came up until the line and said 'I need some information from you' and then came a little bit closer, so she does it in steps, then maybe it would have felt a little bit better".* Therefore, according to our findings, allowing the human to have control by explaining to them before the robot takes action even if the robot makes the decision autonomously, could still enable trust.

5.3 Robot Design and Trust

The robot's appearance, voice and behaviour were pointed out to generally help increase trust or make it easier to trust the robot. The robot's physical properties were often mentioned by participants as an influencing factor for their trust. In particular, the majority of the participants found the robot to be cute, friendly-looking and 'child-like'. Moreover, participants felt that the robot's smaller size made it easier to trust and that a larger or heavier robot would have been more fear-inducing in the conditions where the robot crossed the line without permission. Furthermore, many participants found the human aspects of the robot to help promote trust as well. According to P7, *"[...] Compared to a box that would come by and ask for something, [this robot] is definitely more trustworthy and it's just more comfortable interacting with something that's more like yourself".* Additionally, features such as the robot's eyes also turning green when the robot had finished collecting information from the participants was found to be positive for trust as it increased their understanding of the robot. P12 stated, *"[The robot] has arms that move and the eyes are changing colours. If it's a box then I cannot read anything from it, so yeah I think it helps how it is designed".*

6 Discussion

Based on the results from our quantitative and qualitative analysis, we answer our research questions and discuss our findings. Our interviews also revealed other trends in relation to trustworthy conflict resolution that we consider to be informative and therefore also discuss in this section.

6.1 Automation and Trust

Our first research question focused on how the robot's level of automation during conflict resolution affects users' trust. Our study revealed that trust varied significantly depending on how involved the participants were in the conflict resolution, where higher automation, and subsequently lower user involvement, led to lower trust, especially in combination with no explainability. Previous research has also found the ability to control the system's behaviour to be a significant factor for trust, such as in the study by Nertinger et al. [28], or user control in privacy dashboards [13], which aligns with the findings from our study. However, other studies have found that humans being involved in the decision-making of autonomous systems and providing feedback considered the system to be less accurate, which also impacted their trust [15].

Too autonomous behaviour can lead to confusion and discomfort for the individual, thereby hindering the trust the human feels for the robot. One aspect that we found could explain participants' discomfort and lower trust with the robot acting autonomously and prioritising its own needs was its perceived unpredictability, since participants expected the comfort line to be respected based on the instructions that we had given them in the experiment. If the robot is considered to lack predictability it tends to negatively affect trust [2, 12]. However, what is viewed as unpredictable, and thereby less trustworthy, is influenced by many aspects such as the individual's familiarity with the robot. Our findings show that the participants' familiarity with the robot has a significant impact on trust, which has also been found in previous works [36, 47]. Previous studies in proxemics [29, 40] have also found that prior experiences with a robot make people feel more comfortable with closer distances.

In future work, we aim to further explore the applicability of our findings through experiments in real-world and industrial settings where humans and automated systems work together or alongside each other on tasks. Furthermore, there are many contextual aspects that greatly impact trust and therefore also impact the generalisability of our study. Our statistical analysis revealed a significant impact of participants' age and gender. However, as noted in Section 4, our dataset is relatively imbalanced and therefore these results should be understood in the context of our study and participant sample. We therefore do not discuss these findings, as the interviews did not reveal more information on their significance. In future research, the impact of these factors on trust and conflict resolution needs to be studied further with more balanced datasets. Moreover, other factors such as cultural factors, attitudes to risk or individual personalities were not considered in this research and would also be important areas to focus on in the future.

6.2 Explainability and Trust

Our second research question examined how the robot's level of explainability affects users' trust during conflict resolution. Our statistical analysis showed that explainability did not have a significant effect on trust in this experiment as well as no moderating effect on trust with a higher level of automation. Our interviews revealed that only some of the participants reported having more trust in the robot when they were provided with explanations of the conflict situation due to an increased understandability of the robot and its behaviour. However, particularly when explanations were provided in a fully automated resolution process, trust was usually not re-established after the robot had crossed the comfort line without the participant's permission.

The need for explanations to maintain and foster trust has been identified as a key factor in previous research [4, 14]. Previous studies [12, 20, 30] have also highlighted the importance of transparency in autonomous agents for trustworthy HRI. The non-significant impact of explanations on trust in our experiment may be related to the specific conflict we modelled. A violation of personal space had a too negative impact on other dimensions of trust, so that explanations were not able to repair it. Other conflict scenarios need to be evaluated to understand if explanations may be more significant for trust in other settings, since the impact on trust would most likely vary depending on what type of conflict it is and how risky it is perceived by the user. Moreover, our findings indicate the importance of the timing of explanations. In this study, we only considered two interpolations of explainability and in the future it would be important to evaluate the effects on trust of different types of explanations as well as when they are communicated.

6.3 Anthropomorphism of Robots

In our interviews we found that the majority of participants considered the humanoid features of the robot in the experiment to positively impact their trust, along with the smaller size and height of the robot. Previous studies [14, 27, 41] have also found that anthropomorphism in robots has a positive influence on trust. Moreover, participants found certain behavioural aspects of the robot such as asking for permission and providing explanations to be more trustworthy due to them being similar to their interactions with other humans. Most participants found that the 'humanness' of the robot made it easier for them to relate to the robot and its intentions easier to understand. In fact, the perceived intentions of the robot often appeared to be the underlying reason as to whether the participants trusted the robot or not. For instance, the robot autonomously crossing the line made it untrustworthy since it did not respect their personal space and thereby their comfort and safety. The ability to understand the robot's intentions was also the reason why for some participants explanations helped increase their trust since they were able to understand that the robot was not intending to harm them by crossing the line.

Intentions, and more specifically the benevolence of a person, have been defined as an essential aspect of trust in the social science domain, where good intentions are perceived as more trustworthy. In the context of trust and automation, Lee and See [21] defined purpose, i.e. intention of the technology, as one key dimension to trust, where they also state that humans might attribute intentionality to autonomous systems that are designed to 'be' more human, such as humanoid robots. Therefore, the humanoid factors in a robot's design and behaviour can both support but also decrease trust when the human perceives the robot's actions to be ill-intended. In an HRP conflict, the intent of the robot might play a bigger role in determining its perceived trustworthiness due to the safety aspect and risks for harm that an intrusion of personal space can represent.

6.4 Limitations of the Study

Our experiment was performed with 20 participants and we studied one type of robot in one setting at one point in time. How trust in the robot is perceived during a conflict would most likely vary depending on factors such as what type of robot it is and how it is perceived by the user depending on their identity, personality and perspectives. The sample size was primarily determined by practical constraints, such as time and availability of participants. In future work, the study will need to be replicated with larger and more diverse samples as well as different types of conflicts, robots, and tasks to establish more generalisable results. Moreover, while we accounted for several individual differences, such as age and prior attitudes toward robots, other potentially relevant factors impacting trust in conflict situations need to be explored further.

Another potential limitation in our study is the presence of carryover effects, where the experiences in the first experimental conditions influence participants' responses in later conditions. To mitigate this, we implemented a counterbalancing procedure using the Williams design, ensuring that participants experienced the conditions in systematically varied orders.

Furthermore, to mitigate potential limitations to our trust construct, we carefully defined and operationalised trust based on established literature in HRI and trust in automation. We also used both questionnaires and semi-structured interviews to obtain a more comprehensive understanding of participants' trust perceptions and behaviours. Additionally, to ensure the reliability of our findings, we have provided links to the study material and support our results with quotes from the interviews with the participants.

7 Conclusions

This study investigated the relationship between trust and the robot's explainability and automation level in HRI in the context of an HRP conflict. We performed a within-subjects experimental study with four conditions where we simulated a conflict between the human's requirement to maintain their comfort zone and the robot's task constraint requiring it to come close to the human, thereby needing to cross into the comfort zone. We found that the level of automation had a significant impact on trust, where a partially automated resolution of the conflict, i.e. the human being asked by the robot if it can cross into the comfort zone, increased the human's trust in the robot. Explanations, on the other hand, did not have a significant impact on trust. Interviews with participants after the experiment showed that while many found explanations to be helpful, the robot's reasoning did not mend their trust in the fully autonomous scenarios where the robot autonomously crossed into the comfort zone without asking for permission.

Acknowledgments

Thank you to the study participants and to Dharunkumar Senthilkumar for assisting in this research. This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. Mohammad Obaid is partially funded by the Swedish Research Council (VR), Project ID: 2020-04918.

References

[1] Aldebaran. 2025. PEPPER- Technical specifications. https://aldebaran.com/en/support/kb/pepper/general/pepper-technical-specifications/

[2] Franziska Babel, Johannes Kraus, and Martin Baumann. 2022. Findings from a qualitative field study with an autonomous robot in public: Exploration of user reactions and conflicts. *International Journal of Social Robotics* 14, 7 (2022), 1625–1655.

[3] Franziska Babel, Johannes M. Kraus, and Martin Baumann. 2021. Development and testing of psychological conflict resolution strategies for assertive robots to resolve human–robot goal conflict. *Frontiers in Robotics and AI* 7 (2021), 591448.

[4] Nora Balfe, Sarah Sharples, and John R. Wilson. 2018. Understanding is key: An analysis of factors pertaining to trust in a real-world automation system. *Human Factors* 60, 4 (2018), 477–495.

[5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[6] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.

[7] Robin Bretin, Mohamed Khamis, Emily Cross, and Mohammad Obaid. 2025. The role of drone's digital facial emotions and gaze in shaping individuals' social proxemics and interpretation. *ACM Transactions on Human-Robot Interaction* (2025).

[8] Fanta Camara and Charles Fox. 2022. Extending quantitative proxemics and trust to HRI. In *31st IEEE International Conference on Robot and Human Interactive Communication*. 421–427.

[9] Amos Freedy, Ewart DeVisser, Gershon Weltman, and Nicole Coeyman. 2007. Measurement of trust in human-robot collaboration. In *2007 International Symposium on Collaborative Technologies and Systems*. IEEE, 106–114.

[10] Edward T. Hall. 1963. A system for the notation of proxemic behavior. *American Anthropologist* 65, 5 (1963), 1003–1026.

[11] Edward T. Hall. 1966. *The hidden dimension*. Vol. 609. Garden City, NY: Doubleday.

[12] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie YC. Chen, Ewart J. De Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors* 53, 5 (2011), 517–527.

[13] Eelco Herder and Olaf van Maaren. 2020. Privacy dashboards: The impact of the type of personal data and user control on trust and perceived risk. In *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. 169–174.

[14] 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.

[15] Donald Honeycutt, Mahsan Nourani, and Eric Ragan. 2020. Soliciting human-in-the-loop user feedback for interactive machine learning reduces user trust and impressions of model accuracy. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 8. 63–72.

[16] Helge Hüttenrauch, Kerstin Severinson Eklundh, Anders Green, and Elin A. Topp. 2006. Investigating spatial relationships in human-robot interaction. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 5052–5059.

[17] Davinder Kaur, Suleyman Uslu, Kaley J. Rittichier, and Arjan Durresi. 2022. Trustworthy artificial intelligence: A review. *Comput. Surveys* 55, 2 (2022), 1–38.

[18] Bing Cai Kok and Harold Soh. 2020. Trust in robots: Challenges and opportunities. *Current Robotics Reports* 1, 4 (2020), 297–309.

[19] Moritz Körber. 2019. Theoretical considerations and development of a questionnaire to measure trust in automation. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*. Springer, 13–30.

[20] Johannes Kraus, David Scholz, Dina Stiegemeier, and Martin Baumann. 2020. The more you know: Trust dynamics and calibration in highly automated driving and the effects of take-overs, system malfunction, and system transparency. *Human Factors* 62, 5 (2020), 718–736.

[21] John D. Lee and Katrina A. See. 2004. Trust in automation: Designing for appropriate reliance. *Human Factors* 46, 1 (2004), 50–80.

[22] Brian Y. Lim, Anind K. Dey, and Daniel Avrahami. 2009. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2119–2128.

[23] Roger C. Mayer, James H. Davis, and F. David Schoorman. 1995. An integrative model of organizational trust. *Academy of Management Review* 20, 3 (1995), 709–734.

[24] Joseph E. Mercado, Michael A. Rupp, Jessie YC. Chen, Michael J. Barnes, Daniel Barber, and Katelyn Procci. 2016. Intelligent agent transparency in human–agent teaming for multi-UxV management. *Human Factors* 58, 3 (2016), 401–415.

[25] Fereshteh Jadidi Miandashti, Mohammad Izadi, Ali Asghar Nazari Shirehjini, and Shervin Shirmohammadi. 2020. An empirical approach to modeling user-system interaction conflicts in smart homes. *IEEE Transactions on Human-Machine Systems* 50, 6 (2020), 573–583.

[26] Jonathan Mumm and Bilge Mutlu. 2011. Human-robot proxemics: Physical and psychological distancing in human-robot interaction. In *Proceedings of the 6th International Conference on Human-Robot Interaction*. 331–338.

[27] Manisha Natarajan and Matthew Gombolay. 2020. Effects of anthropomorphism and accountability on trust in human robot interaction. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. 33–42.

[28] Simone Nertinger, Robin Jeanne Kirschner, Abdeldjallil Naceri, and Sami Haddadin. 2024. Acceptance of remote assistive robots with and without human-in-the-loop for healthcare applications. *International Journal of Social Robotics* 16, 6 (2024), 1131–1150.

[29] Mohammad Obaid, Eduardo B. Sandoval, Jakub Złotowski, Elena Moltchanova, Christina A Basedow, and Christoph Bartneck. 2016. Stop! That is close enough. How body postures influence human-robot proximity. In *25th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 354–361.

[30] Scott Ososky, Tracy Sanders, Florian Jentsch, Peter Hancock, and Jessie YC. Chen. 2014. Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. In *Unmanned Systems Technology XVI*, Vol. 9084. SPIE, 112–123.

[31] Raja Parasuraman and Victor Riley. 1997. Humans and automation: Use, misuse, disuse, abuse. *Human Factors* 39, 2 (1997), 230–253.

[32] Ana Pinto, Sónia Sousa, Ana Simões, and Joana Santos. 2022. A trust scale for human-robot interaction: Translation, adaptation, and validation of a human computer trust scale. *Human Behavior and Emerging Technologies* 1 (2022), 1–12.

[33] Vineeth Rajamohan, Connor Scully-Allison, Sergiu Dascalu, and David Feil-Seifer. 2019. Factors influencing the human preferred interaction distance. In *28th IEEE International Conference on Robot and Human Interactive Communication*. IEEE, 1–7.

[34] Jorge Rios-Martinez, Anne Spalanzani, and Christian Laugier. 2015. From proxemics theory to socially-aware navigation: A survey. *International Journal of Social Robotics* 7 (2015), 137–153.

[35] Matteo Rubagotti, Inara Tusseyeva, Sara Baltabayeva, Danna Summers, and Anara Sandygulova. 2022. Perceived safety in physical human–robot interaction: A survey. *Robotics and Autonomous Systems* 151 (2022), 104047.

[36] Tracy L Sanders, Keith MacArthur, William Volante, Gabriella Hancock, Thomas MacGillivray, William Shugars, and PA Hancock. 2017. Trust and prior experience in human-robot interaction. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 61. SAGE Publications: Los Angeles, CA, 1809–1813.

[37] Kazuhiro Sasabuchi, Katsushi Ikeuchi, and Masayuki Inaba. 2018. Agreeing to interact: Understanding interaction as human-robot goal conflicts. In *Companion of the ACM/IEEE International Conference on Human-Robot Interaction*. 21–28.

[38] Azin Semsar and Ali Asghar Nazari Shirehjini. 2017. Multimedia-supported virtual experiment for online user–system trust studies. *Multimedia Systems* 23 (2017), 583–597.

[39] Donghee Shin. 2021. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies* 146 (2021), 102551.

[40] Leila Takayama and Caroline Pantofaru. 2009. Influences on proxemic behaviors in human-robot interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 5495–5502.

[41] Michelle ME. Van Pinxteren, Ruud WH. Wetzels, Jessica Rüger, Mark Pluymaekers, and Martin Wetzels. 2019. Trust in humanoid robots: Implications for services marketing. *Journal of Services Marketing* 33, 4 (2019), 507–518.

[42] Francesco Vigni, Dimitri Maglietta, and Silvia Rossi. 2024. Too close to you? A study on emotion-adapted proxemics behaviours. In *33rd IEEE International Conference on Robot and Human Interactive Communication*. IEEE, 182–188.

[43] Michael L Walters, Kerstin Dautenhahn, Kheng Lee Koay, Christina Kaouri, R te Boekhorst, Chrystopher Nehaniv, Iain Werry, and David Lee. 2005. Close encounters: Spatial distances between people and a robot of mechanistic appearance. In *5th IEEE-RAS International Conference on Humanoid Robots*. IEEE, 450–455.

[44] Michael L. Walters, Kerstin Dautenhahn, Sarah N. Woods, Kheng Lee Koay, R. Te Boekhorst, and David Lee. 2006. Exploratory studies on social spaces between humans and a mechanical-looking robot. *Connection Science* 18, 4 (2006), 429–439.

[45] Antonia Welzel, Rebekka Wohlrab, and Mohammad Obaid. 2025. Study material: Trustworthy conflict resolution in human-robot interactions.

https://figshare.com/articles/media/Study_Material_Trustworthy_Conflict_Resolution_in_Human-Robot_Interactions/29135135
[46] Evan James Williams. 1949. Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Chemistry* 2, 2 (1949), 149–168.
[47] X. Jessie Yang, Vaibhav V. Unhelkar, Kevin Li, and Julie A. Shah. 2017. Evaluating effects of user experience and system transparency on trust in automation. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. 408–416.