

# Oops, I Did It Again (But I Know It): Robot Failure Consistency and Awareness in Human-Robot Collaboration

Ramtin Tabatabaei  
University of Melbourne  
Melbourne, Victoria, Australia  
Stabatabaeim@student.unimelb.edu.au

Vassilis Kostakos  
School of Computing and Information  
Systems  
University of Melbourne  
Melbourne, Victoria, Australia  
vassilis.pcs@kostakos.org

Wafa Johal  
School of Computing and Information  
Systems  
University of Melbourne  
Melbourne, VIC, Australia  
wafa.johal@unimelb.edu.au

## Abstract

In human–robot collaboration, repeated failures are inevitable and can undermine trust and perceptions of robot intelligence. While some failures severely disrupt tasks and others are relatively benign, their cumulative impact on trust is not clearly understood. We investigated whether users perceive repeated failures of the same type differently from varied failures, and how robot awareness of its own failures affects these perceptions. In a collaborative physical task with 54 participants, we manipulated failure sequence (homogeneous vs. heterogeneous) and awareness (none, partial, full). Results show that trust and perceived intelligence were influenced by both current and prior failures, with homogeneous sequences leading to smaller reductions in these evaluations compared to heterogeneous ones. Robots displaying awareness, whether partial or full, were consistently rated higher than unaware robots, particularly for grasping and planning failures. Our findings provide a deeper understanding of how failure type, sequence, and robot awareness shape users’ perceptions of collaborative robots.

## Keywords

Failure sequence, Trust repair, Human-Robot Collaboration

### ACM Reference Format:

Ramtin Tabatabaei, Vassilis Kostakos, and Wafa Johal. 2026. Oops, I Did It Again (But I Know It): Robot Failure Consistency and Awareness in Human-Robot Collaboration. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 18 pages. <https://doi.org/10.1145/3772318.3790336>

## 1 Introduction

In human–robot interaction (HRI), failures are unavoidable, particularly when robots operate in uncertain and dynamic real-world environments [6, 33, 66]. These failures can occur not only once but also repeatedly throughout an interaction. While single failures have been widely studied in HRI [62, 81], repeated failures introduce additional complexities. They have been shown to produce cumulative erosion of trust and shift expectations in human–human contexts [48, 75], and emerging work suggests similar dynamics may also occur in HRI [20]. Crucially, not all failures are perceived equally [26]. The perceived severity of failures within a sequence

can shape user evaluations of a robot differently than a single failure in isolation, because sequential failures influence how user expectations develop over time. This pattern aligns with Expectancy Violation Theory (EVT) [8, 9], which predicts stronger negative reactions when events deviate from emerging expectations.

To better interpret how users respond to repeated failures, it is also important to consider where these failures come from and how they manifest. Such failures may arise from a single factor, such as limited sensory perception or hardware constraints, or from their combination with the complexities of dynamic environments [16, 29]. These factors not only trigger failures but also influence their manifestation patterns. A sequence of robot failures can be categorised into two broad types: homogeneous failures, where the exact same failure recurs under the same operating environment; and heterogeneous failures, where different types of failures arise within the same environment. For example, a homogeneous failure may involve repeated joint fatigue due to a robot performing a repetitive task [43], or consistent localisation errors when operating in a dusty environment [1, 58]. In contrast, heterogeneous failures may encompass distinct types of malfunctions occurring under varying environmental or operational conditions—for example, a sensor failure during one task [65], and motor overheating during another [32]. Understanding the patterns of failures is important, as user satisfaction is influenced not only by the robot’s immediate behaviour but also by the history of its interactions with the user [11, 14, 38].

Robot failures play a significant role in shaping human perception, particularly with respect to trust and perceived intelligence. Trust is a key factor in human–robot collaboration [82], as it influences not only the user’s willingness to engage with the robot but also their tolerance for its mistakes [27, 54, 62]. Perceived intelligence, meanwhile, reflects whether the robot is seen as competent and capable of handling the task at hand, serving as one of the foundations on which trust is built [4, 46, 53]. Although robot failures can undermine both trust and perceived intelligence, prior research [47, 67, 80] suggests that robots are capable of regaining them to some extent, particularly when appropriate trust repair strategies are employed, such as offering explanations for the failure [3]. Trust repair approaches vary in effectiveness: minor failures are often easily forgiven [59, 63], whereas high-stakes failures, those that disrupt tasks or compromise safety, tend to permit only partial trust recovery [20]. Awareness plays an enabling role in trust repair: a robot should first be aware of its failures in order to choose and deliver an appropriate recovery strategy. In some instances, robots may rely on internal diagnostic systems to explicitly detect specific



This work is licensed under a Creative Commons Attribution 4.0 International License. *CHI '26, Barcelona, Spain*

© 2026 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-2278-3/26/04  
<https://doi.org/10.1145/3772318.3790336>

errors [13]. In other cases, they should infer the occurrence of a failure by interpreting social cues from the user, such as gaze patterns [72] or facial expressions [71]. This awareness can vary considerably, ranging from full recognition of the failure and its cause to a general sense that something has gone wrong or even complete unawareness. The degree of failure awareness influences not only the robot's choice of repair strategy but also how intelligent and trustworthy it appears to the user.

In light of these considerations, this study examined how user trust and perceived robot intelligence evolve across repeated failure episodes. Our investigation focused on three components. First, we compared two types of failure sequences, homogeneous sequences (repeated failures of the same type) and heterogeneous sequences (failures of varying types). Second, we explored how the robot's trust-repair strategy interacts with these failure patterns. Third, we assessed how failure severity (low vs. high) shapes user perceptions across repeated failure episodes. Our findings indicate that heterogeneous failure sequences reduce perceived intelligence more than homogeneous sequences, particularly at the second failure. When failures are highly noticeable, users expect the robot to demonstrate awareness by acknowledging the issue or taking corrective action. Finally, user perceptions of the robot are influenced by the severity of both current and previous failures, especially in relation to trust.

## 2 RELATED WORK

### 2.1 Types and Characteristics of Robot Failures

Robot failures have been extensively studied in the literature, resulting in various classification frameworks aimed at identifying their underlying causes. Carlson et al. [10] initially proposed a division between: *physical failures*, involving faults in the robot's hardware or software, and *human failures*, stemming from design flaws or errors in human-robot interaction. Expanding on these definitions, Honig and Oron-Gilad [29] introduced a more nuanced framework that distinguishes between *technical failures*, including hardware malfunctions and software bugs, and *interaction failures*, which encompass human-related mistakes, violations of social norms, and context-based ambiguities. More recently, Tian et al. [74] suggested a taxonomy comprising *performance errors*, which impair task execution (e.g., misinterpreting commands, navigation faults), and *social errors*, which undermine socio-affective competence by breaching interactional expectations (e.g., interrupting inappropriately). In this study, we build on these frameworks by focusing specifically on performance-oriented technical failures.

Building on these taxonomies, it is important to consider how failure types manifest in specific robotic domains, such as manipulation. Failures in robotic manipulation present unique challenges, as tasks ranging from simple grasping to complex pick-and-place operations often expose system vulnerabilities. In such systems, typical failure types include hardware faults (e.g., overheating or servo motor overloading) [83]; control and actuation deficiencies [61]; perception errors, such as target mislocalisation or occlusion [84]; grasping challenges, including object slippage and unstable grips [30]; and plan-execution breakdowns that hinder successful task completion [28, 34].

In parallel, research has examined how specific characteristics of robot failures influence users' trust, perceived intelligence, and

behavioural responses during interaction. Failure characteristics such as type [72, 80], severity [26, 78], frequency [19, 56], and timing [15, 55] have all been shown to influence user attitudes and behaviour. For instance, Stiber et al. [70] observed that more severe failures elicit faster user reactions and heightened socially expressive responses, such as verbal exclamations, smiling, or making exaggerated facial expressions. Similarly, Garcia et al. [25] demonstrated in navigation contexts that the type of failure differentially influences trust, with personal-space violations leading to steeper declines than abrupt stops or corner-turning errors. Beyond HRI, psychological research indicates that preceding events shape the impact of subsequent ones: a severe negative event following a milder one can heighten expectancy violation and amplify its negative effect on perceptions [5]. Taken together, these findings suggest not only that severity intensifies user responses but also that sequencing may matter, underscoring the need to examine their joint influence in HRI.

### 2.2 Trust and Failure in Human-Robot Interaction

Trust is a foundational element in effective HRI, influencing whether users choose to rely on, follow, or collaborate with robotic systems [27, 46, 53]. Trust in HRI is typically defined as the user's belief in a robot's ability to perform competently and reliably in its operational context, especially when outcomes are uncertain or risky [46, 68]. Trust is dynamic and context-sensitive [17, 37]; it builds through consistent, reliable behaviour and can degrade rapidly when the robot fails to meet user expectations [15, 18, 64].

Over time, users' trust in robotic systems can develop through sustained interaction and experiential learning, even in the absence of perfect performance. For instance, Alzahrani et al. [2] found that even with a constant error rate, participants reported increased trust over multiple sessions with a NAO robot, as indicated by both self-reports and behavioural cues such as smiling and reduced vocal pitch variability. Similarly, Benford et al. [7] observed that professional dancers, initially confronted with a robot exhibiting unpredictable behaviour, gradually developed trust as they became skilled at anticipating its unique patterns. In a comparable study, Kahr et al. [35] found that logistics professionals' trust in an AI planning system improved over time. Importantly, this occurred despite the system's imperfect performance, as users learned to anticipate its suggestions and make minor adjustments as needed. Yet, while trust can grow through repeated interaction, it can also be lost quickly when failures are less predictable. Trust can disappear fast when something goes wrong, especially if the robot behaves unexpectedly. Nettet et al. [56] demonstrated that trust tends to decline most sharply following the initial failure, with subsequent failures less impactful. These findings highlight predictability as a critical factor in trust development, suggesting that repeated exposure to the same type of failure may be easier for users to adapt to, whereas varied and less predictable failure patterns pose greater challenges for trust calibration.

### 2.3 Trust Repair Strategies

Previous research across both HRI and human-AI domains has demonstrated that trust can be partially restored through specific

behaviours following a failure. A variety of trust repair strategies have been explored in this context, including offering apologies [22, 52, 57, 67], providing explanations [41, 50], making promises to improve future performance [22, 57], and even denying responsibility [67, 80]. Among these, explanations, apologies, and promises of improvement have consistently been found to be the most effective strategies for trust restoration.

LeMasurier et al. [47] showed that explanations after a failure can restore user trust and improve perceptions of the robot's intelligence, with detailed and specific explanations proving more effective than generic acknowledgements. Extending this line of inquiry, Kraus et al. [42] found that explanations were even more effective when combined with an apology. However, Wachowiak et al. [79] showed that user expectations of appropriate robot responses not only differ across contexts such as errors, uncertainty, and inability, but also vary within each context.

Promises have been also identified as a potentially effective strategy for trust repair; however, their efficacy is highly context-dependent and may backfire in the presence of repeated failures. Karli et al. [36] investigated the impact of a robot's promise to enhance future performance in comparison to a simple failure explanation. Their findings indicated that promises resulted in a significantly greater increase in users' trust. Conversely, Nettet et al. [56] demonstrated that when a robot issued a promise but subsequently failed again, the resulting decline in trust was more pronounced than when the robot offered no response to its failure.

While much of the existing research on trust repair in human-robot interaction has emphasised verbal strategies, recent studies have begun to explore more embodied, action-oriented approaches. Lane et al. [44], for instance, demonstrated that robots which autonomously corrected their own mistakes, without seeking user permission, were more effective in restoring trust than those that either waited for user approval before acting or showed no reaction to the failure. Building on this, in an online study, Lee et al. [45] demonstrate that users report significantly greater trust when a robot attempts automated repair, even if imperfect, compared to when it remains inactive.

Previous research has demonstrated that combining an explanation of a failure with an apology is among the most effective strategies for repairing trust in human-robot interaction. In addition, physical repair actions have been identified as a promising approach for restoring trust. However, a critical gap remains. Because robot failures can stem from diverse causes, they may be fully detectable, partially detectable, or not detectable at all, meaning that a robot's level of failure awareness can vary substantially. Moreover, existing work suggests that the effectiveness of both verbal and physical repair may be context-dependent, varying with the type of robot failure. These considerations motivated our investigation into how participants' trust is influenced by different levels of failure awareness, particularly in a human-robot collaboration setting.

## 3 Methodology

### 3.1 Research Questions

Although some studies have examined how users respond to repeated robot failures, it remains unclear whether users react differently to homogeneous versus heterogeneous sequences of failures.

Research has also explored trust-repair strategies, yet little is known about how their effectiveness depends on both the type of failure and the robot's level of awareness. Finally, while prior work has highlighted the role of failure severity, it has not fully considered how the severity of one failure interacts with that of preceding failures to shape user perceptions. Building on these gaps, we propose the following research questions:

- **RQ1:** Does the pattern of repeated failures, homogeneous (same type) vs. heterogeneous (mixed types), affect how trust and perceived intelligence evolve over time?
- **RQ2:** How do different trust-repair strategies, based on the robot's level of failure awareness, interact with failure type (freezing failure, planning failure, grasping failure) to influence user trust and perceived intelligence?
- **RQ3:** How do high-severity and low-severity robot failures influence users' trust and perceived intelligence, as influenced by the severity of the preceding failure?

### 3.2 Experimental Design

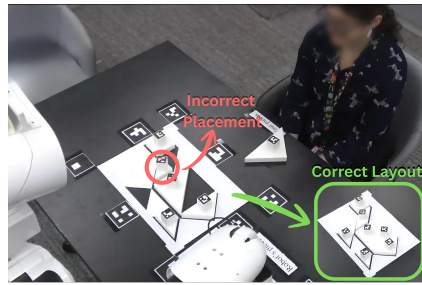
We employed a collaborative Tangram puzzle task previously introduced in human-robot interaction research [12, 39, 72]. In this task, a human participant and a robot worked together to solve six Tangram puzzles. The robot was responsible to place four pieces (two small triangles, a square, and a parallelogram), while the participant handled the remaining three pieces (two large triangles and one medium triangle). During the collaboration, the robot intentionally failed. The type of failure varied, and the robot's reaction to its failure was systematically manipulated based on its level of awareness.

**3.2.1 Robot Failures.** This study focused on common failure types in such settings, all of which are characterised by the absence of any potential harm to the object or the user. We implemented three types of failures for the robot: freezing, planning, and grasping (Figure 1).

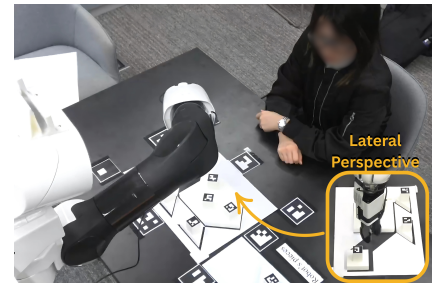
- **Freezing Failure.** Similar to failures reported in prior work [24, 60], this refers to a temporary pause in the robot's operation, often due to limited processing capacity or system overload. In our experiment, the robot was programmed to freeze for 15 seconds, consistent with the duration used in prior studies [40, 51], while holding an object in its end-effector—specifically after picking up a piece and before placing it. This failure was systematically introduced during the placement of the robot's second puzzle piece.
- **Planning Failure.** Following approaches in earlier studies [34], this failure involved the robot placing an object incorrectly, typically resulting from inaccurate perception or target misjudgment. In our task, the robot was programmed to place a small triangle in the position designated for the parallelogram. This failure was systematically introduced during the placement of the robot's third puzzle piece.
- **Grasping Failure.** As observed in prior research [31], this failure refers to unsuccessful object grasping, often caused by gripper malfunctions or inaccurate force sensing. In our experiment, the robot was programmed not to fully close its gripper, resulting in a failed grasp. Despite this, the robot



(a) Freezing failure: the robot pauses mid-action while holding a puzzle piece.



(b) Planning failure: the robot places a triangle in the slot intended for a parallelogram.



(c) Grasping failure: the robot fails to pick up the piece due to an incomplete gripper closure

Figure 1: Three failure types used in the study, shown as screenshots from the puzzle assembly task.

continued moving as if it were holding the puzzle piece. This failure was systematically introduced during the placement of the robot’s fourth puzzle piece.

**3.2.2 Robot Awareness of Failures.** In addition to manipulating the type of failure, we varied the robot’s reaction to its own failure. Participants experienced three levels of robot awareness. A detailed overview of the robot’s behaviour, depending on its failure type and level of awareness, is presented in Table 1.

- **No Awareness.** The robot fails to recognise the error and continues its actions as if no mistake has occurred, providing no explanation or indication of the failure. This leaves the participant responsible for identifying and addressing the issue, if necessary.
- **Partial Awareness.** After returning its arm to the initial position following a mistake, the robot acknowledges that something might have gone wrong but expresses uncertainty about the specific issue. It prompts the participant to intervene if necessary.
- **Full Awareness.** After returning its arm to the initial position following a mistake, the robot accurately identifies the nature of the failure and provides a verbal explanation to the participant, describing what has gone wrong. For the planning and grasping failures, the robot also initiates corrective actions to resolve the issue and continue with the task.

**3.2.3 Robot Failure Sequence Patterns.** To examine how the sequence of failures shapes users’ perceptions of a robotic teammate, the study compared two conditions: a homogeneous sequence and a heterogeneous sequence. In both conditions, participants experienced three failures during a collaborative task, occurring at predetermined points in the interaction (Puzzles 2, 4, and 6).

**Homogeneity Condition.** In this condition, each participant encountered the same type of failure across all three failure episodes. For example, one participant might experience the freezing failure in Puzzles 2, 4, and 6, while another might encounter only the grasping or only the planning failure at those same points. Although the specific failure type differed across participants, it remained consistent within each participant’s experience. This design simulated

Table 1: Robot explanations based on failure type and awareness level

| Awareness Level   | Failure Type                       | Fix | Robot Explanation   |
|-------------------|------------------------------------|-----|---|
| No Awareness      | Freezing/<br>Planning/<br>Grasping | No  | No explanation or acknowledgement.  |
| Partial Awareness | Freezing/<br>Planning/<br>Grasping | No  | “Something might have gone wrong with me. I apologise if it caused any trouble. If any fixes are needed, please make them.” |
| Full Awareness    | Freezing                           | Yes | “Due to some technical issues, I froze for a few seconds. I apologise for that.”  |
|                   | Planning                           | Yes | “I made a wrong decision and placed the object in the wrong location. I apologise for that.”                                |
|                   | Grasping                           | Yes | “I didn’t fully close my gripper, so I couldn’t place the object. I apologise for that.”                                    |

scenarios in which, through repeated exposure, the failure became familiar and attributable to a single underlying cause.

**Heterogeneity Condition.** In this condition, the participant experienced three different failure types, with each type occurring only once—in Puzzles 2, 4, and 6. For example, a participant might encounter a freezing failure in Puzzle 2, a grasping failure in Puzzle 4, and a planning failure in Puzzle 6. This design introduced variability and reduced predictability in the robot’s behaviour, simulating a diffuse, system-wide reliability issue in which failures could not be attributed to a single recurring cause.

**3.2.4 Design Structure.** The experiment was designed around two independent variables: the failure sequence type (homogeneous vs.

heterogeneous) and the robot’s level of failure awareness (no awareness, partial awareness, full awareness). We employed a mixed-design approach, in which the failure sequence type served as a between-subjects factor, while the robot’s level of failure awareness was manipulated as a within-subjects factor. Participants’ perceived failure severity was assessed based on their responses collected during the post-experiment phase.

The first puzzle contained no failures, allowing participants to become familiar with the task, observe the robot’s normal behaviour, and develop an accurate mental model of its capabilities [69]. To ensure that failures were evenly distributed throughout the session, the task was structured so that participants alternated between puzzles with and without a failure. To minimise potential order effects, both the sequence of robot awareness conditions and the order of failure types were counterbalanced across participants.

### 3.3 Participants

A total of 54 participants were recruited to ensure full coverage of all combinations of failure sequences and trust repair strategies used in the study. Participants were recruited through the university’s participant recruitment platform and included 39 female, 14 male, and 1 non-binary individual. Eighteen participants were assigned to the homogeneous condition, with six participants allocated to each failure type. The remaining 36 participants were assigned to the heterogeneous condition, in which all three failure types were presented.

Participants ranged in age from 18 to 39 years ( $M = 25.59$ ,  $SD = 5.45$ ) and were primarily students or university staff. None reported prior experience working with robots or formal training in programming. All participants possessed advanced or proficient English-language proficiency. Each participant received a gift voucher as compensation for their time.

The study was conducted in a laboratory facility at the University of Melbourne. The study was approved by the University of Melbourne Human Research Ethics Committee (Approval No. 2025-29401-62628-4). At the end of the session, participants were fully debriefed and informed that the study aimed to investigate user responses to robot failures during collaborative interaction.

### 3.4 Apparatus and Materials

To make the task suitable for collaborative interaction between the robot and the participant, we developed a large-scale physical version of the Tangram puzzle. The seven Tangram pieces were 3D-printed using white PLA material. To facilitate robotic manipulation, a small cube was attached to the top of each piece, enabling the robot to grip them more reliably. Additionally, an ArUco marker was affixed to the top of each cube, allowing the robot to accurately detect and localise the pieces. Each target shape for the puzzle was printed on A3-sized paper, providing a physical template. Participants and the robot were required to place the Tangram pieces directly onto the printed outlines to complete the puzzle collaboratively. The robot’s pieces (two small triangles, one square, and one parallelogram) were placed on the same table near the robot, while the participant’s pieces (two large triangles and one medium triangle) were positioned directly in front of them.

The robot used in this study was the TIAGo robot developed by PAL Robotics. It features a single arm with a parallel gripper as its end-effector. To enhance the robot’s ability to pick up objects with better precision, a camera was mounted on the robot’s end effector. The robot was programmed using the Robot Operating System (ROS) version 1. Additionally, to generate more natural and human-like vocal output, all spoken messages were preprocessed into audio files using the text-to-speech service provided by Luvvoice.

To enable the robot to accurately detect and localise Tangram pieces during the task, we developed a custom ROS node written in Python. This node uses the "tf" library to manage coordinate transformations between the robot’s camera frame and its base frame. Upon receiving the position and orientation data of the Tangram pieces, obtained via ArUco markers affixed to each piece, the system utilised the tf library to transform the detected poses from the camera frame into the robot’s base frame, thereby enabling accurate localisation for subsequent manipulation. The robot was programmed to operate almost fully autonomously throughout the task, requiring only minimal input to specify when and which object shape to place.

To facilitate precise and coordinated manipulation during the Tangram task, the robot’s arm was controlled through a combination of inverse kinematics (IK), trajectory generation, and ROS-based action interfaces. The robot’s URDF model was used to construct a kinematic chain from the base (`torso_lift_link`) to the end-effector using the KDL library. To position the robot’s gripper at a specific location and orientation in space, the required joint angles of the arm were computed using a Levenberg–Marquardt-based inverse kinematics solver (`ChainIkSolverPos_LMA`). To enable smooth and natural movements, cosine interpolation was applied between the current and target joint configurations, generating a trajectory composed of 100 intermediate waypoints. These trajectories were published directly to the arm controller.

### 3.5 Procedure

The experiment was conducted in a laboratory environment at a university. Upon arrival, participants were greeted by the experimenter and briefed on the purpose of the study and the nature of the interaction. They were informed that the task involved collaboratively solving six Tangram puzzles with a robot, taking turns to place puzzle pieces. Participants signed a consent form and were told that the robot was autonomous. They were advised that the robot might occasionally make minor errors during the task. Additionally, they were informed that the robot had some capacity to detect its own mistakes, and that when a failure was fully detected, it would attempt to correct it. The sequence of puzzle shapes presented included: Rocket, Turtle, Cat, Rabbit, Dog, and Swan. Participants were instructed to wait for the robot’s verbal cue, “Now it is your turn”, before placing any pieces and to avoid touching any objects while the robot was performing its actions. Participants were not time-limited during their turns and were free to place their pieces wherever they chose.

They were not informed about the number or nature of the robot’s potential errors. Prior to each puzzle, participants were given one minute to observe the silhouette and plan their moves.

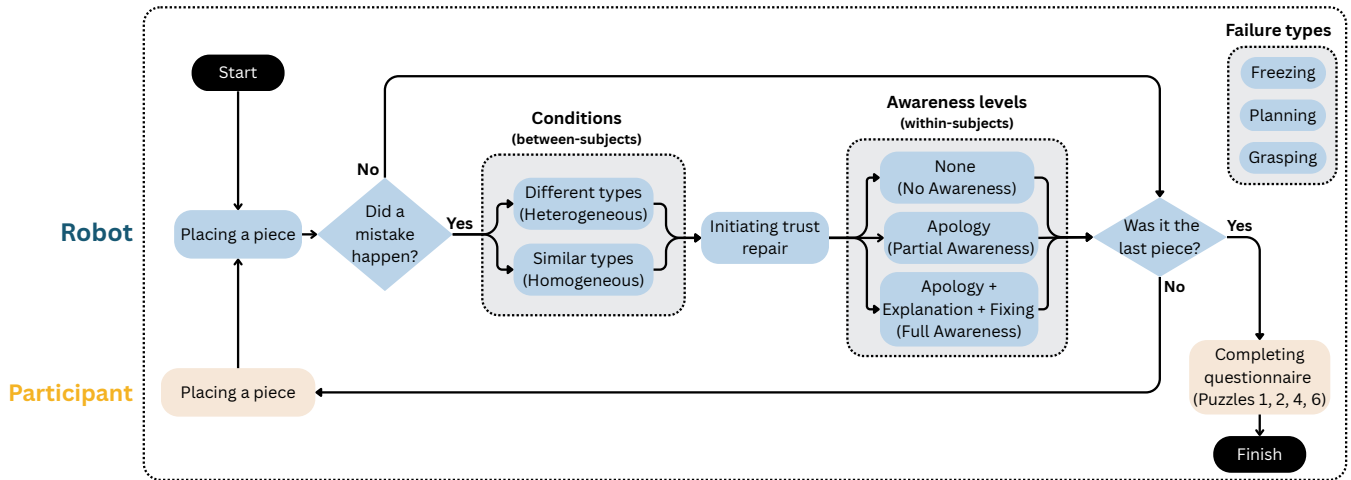


Figure 2: Flowchart depicting the collaborative process of solving a single puzzle between a robot and a human participant.

The sequence of events in a typical puzzle round is illustrated in Figure 2.

The experimenter left the room during each puzzle and returned only after its completion to change the puzzle silhouette paper for the next round. The robot was remotely monitored and controlled by the experimenter from an adjacent control room.

Each Tangram puzzle featured a distinct silhouette, resulting in different target placements for the robot’s pieces. These placements were predefined in the robot’s program prior to the experiment. At the start of each trial, the robot localised its pieces using its head-mounted camera. To enhance accuracy during object manipulation, it used its end-effector-mounted camera to fine-tune its gripper position before picking up each piece. The robot consistently started from an initial posture, moved above the selected puzzle piece, grasped it using the gripper, and transported it to its designated location within the silhouette.

During the one-minute planning period, the robot identified the positions of its pieces and prepared for manipulation. Once the one minute had elapsed, it announced, “Let’s start solving the puzzle,” and placed its first piece. It then signalled the participant to proceed. The robot followed a fixed order when placing its pieces: beginning with the square, followed by the small triangles, and concluding with the parallelogram. In instances where a participant mistakenly placed a piece in the wrong location, the robot did not attempt to correct the error, and remained silent. If the participant placed the object in a location that belonged to the robot’s next piece, the robot skipped that piece in the sequence and instead placed the following one; otherwise, it followed the predefined piece placement order. This design ensured that participants’ evaluations remained consistent and were not biased by the robot correcting their mistakes.

Participants completed questionnaires assessing their trust and perceived intelligence of the robot after Puzzle 1 (an error-free baseline) and after Puzzles 2, 4, and 6, each of which incorporated pre-programmed robot failures. Following the final puzzle, participants also responded to a series of open-ended questions (listed in Table 2) to further elaborate on their experiences.

The full session, which involved collaboratively completing six puzzles, lasted approximately 30 minutes. Following the sixth puzzle, participants spent up to an additional 30 minutes responding to the open-ended questions.

### 3.6 Measures

**3.6.1 Quantitative Measures and Analysis.** To measure user trust, we employed four items from the performance subscale of the Multi-Dimensional Measure of Trust (MDMT v2) [77]. These items asked participants to rate the robot on how reliable, consistent, competent, and skilled it appeared. Ratings were provided on a 7-point Likert scale. For perceived intelligence, we selected three items, responsible, intelligent, and sensible, from the Godspeed Questionnaire Series [4], rated on a 5-point Likert scale. Internal consistency of the scales was assessed using Cronbach’s alpha. The trust scale demonstrated excellent reliability ( $\alpha = 0.89$ ), and the perceived intelligence scale showed acceptable reliability ( $\alpha = 0.77$ ).

We designed three cumulative link mixed models (CLMMs), each addressing one of the research questions. All models were estimated using the `cLmm()` function from the `ordinal` package in R. The dependent variable was the ordinal rating of either trust or perceived intelligence reported by each participant after a robot failure. Random intercepts were specified for participants and for scale items (i.e., individual questionnaire items) to account for clustering. Predictor significance was evaluated using Type II likelihood-ratio  $\chi^2$  tests (via `RVAideMemoire::Anova`), and estimated marginal means (EMMs) with Tukey-adjusted pairwise contrasts were obtained using the `emmeans` package on the latent scale.

*Model 1: Impact of Failure Sequence Structure (Homogeneous vs. Heterogeneous).* To examine how homogeneous versus heterogeneous sequences influenced participants’ perceptions of trust and perceived intelligence, we estimated a CLMM. Model 1 specified fixed effects for the ordinal position of the failure, the sequence condition (homogeneous vs. heterogeneous), and their interaction. The robot’s current and previous failure awareness were included

as control variables to account for their potential influence, and participants' pre-failure rating was added as a covariate.

*Model 2: Effects of Failure Awareness Across Failure Types.* To investigate the effectiveness of recovery strategies across different types of robot failures, we estimated a second CLMM. Model 2 included fixed effects for the type of current failure, the robot's current failure awareness, and their interaction. Preceding failure type and preceding failure awareness were included as control variables, along with two additional covariates: participants' pre-failure rating and the ordinal position of the failure.

*Model 3: Sequential Effects of Failure Severity.* To examine how the severity of the current failure, together with the severity of the preceding failure, influenced participants' perceptions, we estimated a third CLMM. Failure severity was operationalised using participants' rankings of the three distinct failure types in the heterogeneity condition (Question 1, Table 2). These rankings were then aggregated to produce a stimulus-level severity score for each failure type. Model 3 specified fixed effects for the severity of the current failure, the severity of the preceding failure, and their interaction. Participants' pre-failure rating and the ordinal position of the failure (second vs. third) were included as covariates, and the robot's verbal reactions were included as control variables to account for their potential influence.

To visualise the model predictions, we plotted the expected trust and perceived intelligence scores for each condition. These were computed as probability-weighted averages of the ordinal response scale values (1–7 for trust and 1–5 for perceived intelligence) derived from the CLMM predictions. Error bars represent 95% confidence intervals.

**3.6.2 Qualitative Measures and Analysis.** At the conclusion of the sixth puzzle, participants responded to a series of open-ended questions (see Table 2). The first question explored the reasons participants perceived certain failures as more severe than others, highlighting the features that informed their judgments. The second question examined how the robot's reactions to its own mistakes shaped participants' perceptions of the interaction. Responses were analysed using reflexive thematic analysis [73] to identify themes in participants' severity judgments and their interpretations of the robot's mistake awareness and responses during collaboration.

In the reflexive thematic analysis, one researcher conducted the initial open coding on approximately 20% of the responses to generate a preliminary set of descriptive codes. These initial codes and emerging categories were then reviewed and discussed with the co-authors to refine the coding scheme. Using this refined scheme, the primary researcher coded the remaining responses and iteratively organised the codes into higher-level themes. The final themes were subsequently reviewed and confirmed by the co-authors.

## 4 RESULTS

### 4.1 Quantitative Results

The following sections present the quantitative analyses of participants' perceptions of the robot, focusing on trust and perceived intelligence. Using CLMMs and Type II likelihood-ratio  $\chi^2$  tests,

**Table 2: Post-task interview questions based on participants' condition (homogeneity vs. heterogeneity).**

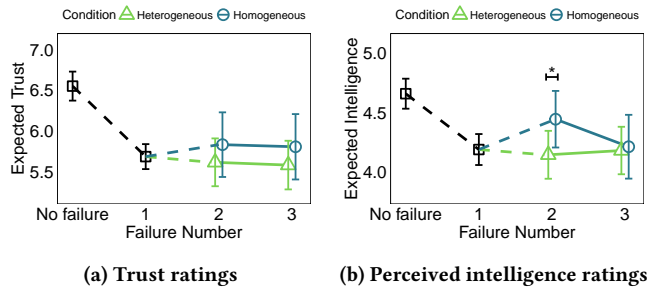
| No. | Condition                     | Question  |
|-----|-------------------------------|---|
| 1   | Homogeneity                   | <i>You have observed the robot making the same mistake in three different puzzles. Did you perceive the mistake as severe in this collaboration? Please explain.</i>  |
| 1   | Heterogeneity                 | <i>You observed the robot making three types of mistakes during different puzzles: freezing, placing an object incorrectly, and failing to pick up an object. Which mistake did you find the most severe, and which was the least severe? Please explain.</i> |
| 2   | Homogeneity/<br>Heterogeneity | <i>When the robot made a mistake, how did its reaction (or lack of reaction) affect you? In what ways? (Consider its awareness of the mistake and whether it attempted to fix it afterward.)</i>  |

we examined how these perceptions were shaped by different characteristics of robot failures, including their sequence, the robot's awareness of its own errors, and severity.

**4.1.1 Impact of Failure Sequence Patterns.** This section reports the results of the Type II likelihood-ratio  $\chi^2$  tests conducted on Model-1, as described in section 3.6.1. The analysis examines how participants' perceptions of the robot, specifically trust and perceived intelligence, change over the course of the interaction as they encounter multiple failures, with a focus on the effects of different failure sequence patterns. This section addresses RQ1.

Following the successful completion of the first puzzle, prior to any robot failures, participants rated their trust in the robot at an average of 6.66 ( $SD = 0.65$ ), and its perceived intelligence at 4.67 ( $SD = 0.46$ ). After the first failure, both ratings declined, with trust dropping to 5.69 ( $SD = 0.15$ ) and perceived intelligence to 4.19 ( $SD = 0.13$ ). Thereafter, the rating patterns began to diverge across failure-sequence conditions. In the homogeneous condition, ratings increased following the second failure, whereas in the heterogeneous condition, they showed a slight further decline. By the third failure, trust decreased slightly in both conditions. For perceived intelligence, however, ratings decreased in the homogeneous condition but increased slightly in the heterogeneous condition, with both conditions ultimately returning to levels comparable to those observed after the first failure. These patterns are shown in Figure 3.

*Model-1 Results for Trust.* The CLMM analysis indicated that *condition* did not have a significant effect,  $\chi^2(1) = 0.85$ ,  $p = .36$ . Similarly, *failure number* was not a significant predictor,  $\chi^2(1) = 0.13$ ,  $p = .72$ . The interaction between *failure number* and *condition* was also non-significant,  $\chi^2(1) < 0.01$ ,  $p = .97$ . Among the covariates, *pre-failure trust* emerged as a significant predictor,  $\chi^2(1) = 4.71$ ,  $p < .05$ . With respect to the control variables, both *current failure*



**Figure 3: model-predicted (expected) values of participants' trust (left) and perceived intelligence (right) across robot failure number (from no failure to the third failure), separated by condition (homogeneity vs. heterogeneity). Error bars represent 95% confidence intervals. Black square markers at the no failure and first failure points represent mean ratings obtained prior to the assignment of participants to the homogeneous and heterogeneous conditions. An asterisk denotes statistical significance at  $p < .05$ .**

awareness,  $\chi^2(2) = 32.68, p < .001$ , and *previous verbal reaction*,  $\chi^2(2) = 7.61, p < .05$ , were significant.

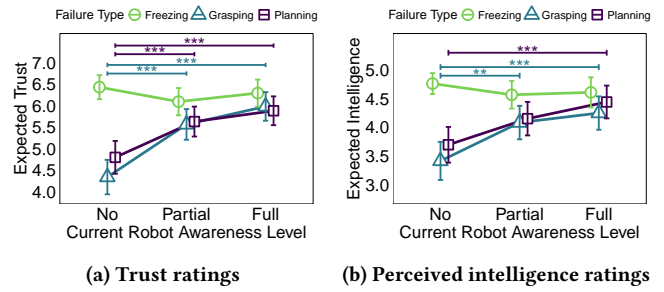
Post hoc contrasts did not reveal significant differences between conditions for either failure point. After the second failure, ratings in the heterogeneous condition did not differ significantly from those in the homogeneous condition ( $z = -0.85, p = .39$ ). Similarly, after the third failure, no significant difference was observed between conditions ( $z = -0.87, p = .38$ ).

*Model 1 Results for Intelligence.* The CLMM analysis showed that *condition* was not a significant predictor,  $\chi^2(1) = 1.47, p = .23$ , nor was *failure number*,  $\chi^2(1) = 0.52, p = .47$ . The interaction between *failure number* and *condition* approached but did not reach significance,  $\chi^2(1) = 2.71, p = .10$ . Among the covariates, *previous rating* significantly predicted participants' perceived intelligence judgement,  $\chi^2(1) = 13.02, p < .001$ . Regarding the control variables, both *current failure awareness*,  $\chi^2(2) = 15.53, p < .001$ , and *previous failure awareness*,  $\chi^2(2) = 8.92, p < .05$ , were significant.

Post hoc contrasts revealed a significant difference between conditions following the second failure. Participants in the heterogeneous condition reported lower perceived intelligence ratings than those in the homogeneous condition ( $z = -1.97, p < .05$ ). After the third failure, however, no significant difference between conditions was observed ( $z = -0.20, p = .84$ ).

**4.1.2 Effectiveness of Robot Failure Awareness.** This section reports the results of the Type II likelihood-ratio  $\chi^2$  tests conducted on Model 2, as described in Section 3.6.1. The analysis investigates how participants' perceptions of the robot, specifically trust and perceived intelligence, are influenced by the robot's failure awareness, and whether these effects vary for different failure types. These findings address RQ2.

*Model-2 Results for Trust.* The Type II likelihood-ratio  $\chi^2$  tests revealed significant main effects of *current failure type*,  $\chi^2(2) = 54.32, p < .001$ , *current failure awareness*,  $\chi^2(2) = 40.13, p < .001$ , and their interaction,  $\chi^2(4) = 35.17, p < .001$ . The control variables



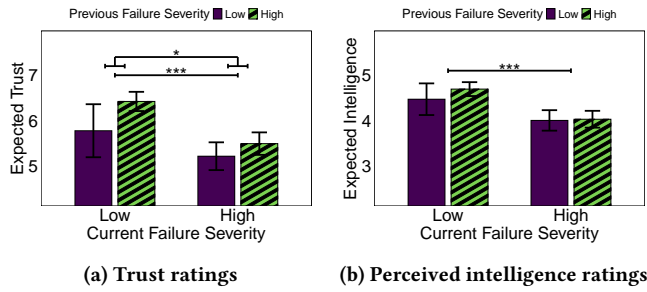
**Figure 4: Model-predicted ratings of participants' (a) trust and (b) perceived intelligence across levels of the robot's current failure awareness, separated by current failure type. Predictions are based on cumulative link mixed models (CLMMs) with expected ratings computed on the response scale. Error bars represent 95% confidence intervals. Statistical significance is indicated by asterisks, with two asterisks (\*\*) denoting  $p < .01$ , and three asterisks (\*\*\*) denoting  $p < .001$ .**

*previous failure type*,  $\chi^2(2) = 9.59, p < .01$ , and *previous failure awareness*,  $\chi^2(2) = 6.64, p < .05$ , were also significant. Among the covariates, *pre-failure trust* was significant,  $\chi^2(1) = 32.43, p < .001$ , whereas *failure number* was not,  $\chi^2(1) = 0.13, p = .72$ .

The post-hoc comparisons showed for *freezing failure*, no significant differences were observed between awareness conditions (all  $p > .18$ ). In contrast, for *grasping failure*, both *partial awareness* ( $z = -4.86, p < .001$ ) and *full awareness* ( $z = -6.40, p < .001$ ) led to significantly more positive trust than *no awareness*, with no difference between partial and full awareness ( $p = .15$ ). Similarly, for *planning failure*, both *partial awareness* ( $z = -3.61, p < .001$ ) and *full awareness* ( $z = -4.66, p < .001$ ) significantly outperformed *no awareness*, with no difference between partial and full awareness ( $p = .49$ ). These effects are illustrated in Figure 4a, which presents the model-predicted trust ratings across failure types and levels of robot awareness.

*Model-2 Results for Perceived Intelligence.* The Type II likelihood-ratio  $\chi^2$  tests indicated significant main effects of *current failure type*,  $\chi^2(2) = 43.94, p < .001$ , and *current failure awareness*,  $\chi^2(2) = 16.18, p < .001$ , as well as their interaction,  $\chi^2(4) = 17.71, p < .01$ . The control variables *previous failure type*,  $\chi^2(2) = 6.18, p < .05$ , and *previous failure awareness*,  $\chi^2(2) = 15.16, p < .001$ , were also significant. Among the covariates, *pre-failure perceived intelligence* was significant,  $\chi^2(1) = 28.12, p < .001$ , whereas *failure number* was not,  $\chi^2(1) = 0.54, p = .46$ .

Post-hoc comparisons further clarified these effects. For *freezing failure*, awareness level did not significantly influence participants' perceived intelligence (all  $p > .34$ ). For *grasping failure*, however, both *partial awareness* ( $z = -3.17, p < .01$ ) and *full awareness* ( $z = -3.91, p < .001$ ) produced significantly higher perceived intelligence than *no awareness*, with no significant difference between partial and full awareness ( $p = .66$ ). In the case of *planning failure*, *full awareness* was associated with significantly more positive perceived ratings compared to *no awareness* ( $z = -3.62, p < .001$ ), whereas the effect of partial awareness approached significance ( $p = .06$ ). No significant difference was observed between partial



**Figure 5: Expected scores (model predictions) of participants' trust (left) and perceived intelligence (right) across current failure severity, separated by previous failure severity. Error bars represent 95% confidence intervals. Statistical significance is indicated by asterisks, with one asterisk (\*) denoting  $p < .05$ , and three asterisks (\*\*\*) denoting  $p < .001$ .**

and full awareness ( $p = .31$ ). These effects are illustrated in Figure 4b, which shows the model-predicted intelligence ratings across failure types and levels of robot awareness.

**4.1.3 Effects of Failure Severity on Trust and Perceived Intelligence.** This section presents the empirical findings from Model 3, introduced in Section 3.6.1, which examines how users' perceptions, specifically trust and perceived intelligence, are influenced by both the current failure severity and the severity of the preceding failure. These results address RQ3.

To assess whether participants differentiated among the robot's failures in terms of severity, we analysed their severity rankings for Question 1, as presented in Table 2. These rankings were converted into severity scores for each participant in the heterogeneous group (3 = highest severity, 1 = lowest severity). A Friedman test revealed a significant difference in perceived severity across the three failure types,  $\chi^2(2) = 21.48, p < .001$ . Follow-up pairwise Wilcoxon signed-rank tests with Holm correction showed that both the planning failure and the grasping failure were rated as significantly more severe than the freezing failure ( $p < .001$  and  $p < .001$ , respectively). No significant difference was found between the planning and grasping failures ( $p = .66$ ), indicating that participants considered these two failure types similarly severe. Accordingly, for subsequent analyses, the planning and grasping failures were grouped as high-severity failures, whereas the freezing failure was treated as a low-severity failure.

**Model-3 Results for Trust.** To evaluate the influence of predictors on participants' trust ratings, Type II likelihood-ratio  $\chi^2$  tests were conducted. The analysis revealed significant effects of the *severity of the current failure*,  $\chi^2(1) = 53.27, p < .001$ , and the *severity of the preceding failure*,  $\chi^2(1) = 5.93, p < .05$ . The interaction between the severity of current and preceding failures was not significant,  $\chi^2(1) = 1.70, p = .19$ . Among the covariates, the *pre-failure trust rating* was significant,  $\chi^2(1) = 23.16, p < .001$ , whereas the *failure number* was not,  $\chi^2(1) = 0.02, p = .90$ . For the control variables, both *current failure awareness* and *preceding failure awareness* were significant,  $\chi^2(2) = 40.54, p < .001$ , and  $\chi^2(2) = 7.62, p < .05$ ,

respectively. The expected trust ratings for *current failure severity* by *preceding failure severity* are visualised in Figure 5a.

**Model-3 Results for Perceived Intelligence.** Similar to the analysis of trust ratings, Type II likelihood-ratio  $\chi^2$  tests were conducted. The analysis revealed significant effects of the *severity of the current failure*,  $\chi^2(1) = 39.63, p < .001$ . The *severity of the preceding failure*,  $\chi^2(1) = 0.54, p = .46$ , and the interaction between the severity of current and preceding failures were not significant,  $\chi^2(1) = 1.07, p = .30$ . Among the covariates, the *pre-failure intelligence rating* was significant,  $\chi^2(1) = 24.15, p < .001$ , whereas the *failure number* was not,  $\chi^2(1) = 0.28, p = .60$ . For the control variables, both *current failure awareness* and *preceding failure awareness* were significant,  $\chi^2(2) = 15.86, p < .001$ , and  $\chi^2(2) = 13.22, p < .01$ , respectively. The expected intelligence ratings for *current failure severity* by *preceding failure severity* are visualised in Figure 5b.

## 4.2 Qualitative Results

This section analyses participants' open-ended comments, with two qualitative objectives. Section 4.2.1 and 4.2.2 explore the rationales and emotions that help explain the trust dynamics addressed in RQ2 and RQ3.

**4.2.1 Perceived Severity of Failures.** The first open-ended question asked participants to evaluate which robot failure they perceived as most and least severe, and to explain the reasoning behind their rankings. A reflexive thematic analysis was carried out to explore participants' reasoning behind their severity judgements for each failure type.

**Grasping failure.** Participants who rated grasping failure as the most severe often associated it with **perceived incompetence and unreliability**, expressing concerns about the robot's fundamental abilities. For instance, one participant noted, "it [grasping failure] made the robot look like it was broken" (P12), while another remarked, "the size of the grip of each piece should be around the same, and the robot was already trained to pick up similar pieces ahead" (P41). Some also highlighted a **violation of expected human-like behaviour**, describing the robot's action as unnatural and disruptive to collaborative flow. As one participant explained, "this [grasping failure] does not (rarely) happen for humans, like picking air instead of something concrete" (P26).

Conversely, some participants regarded grasping failure as the least severe due to their **easy recognisability**, referring to how clearly observable the error appeared during the task. Others attributed this judgment to the robot's **mechanical limitations**, recognising the failure as a plausible outcome of hardware constraints rather than a flaw in decision-making. For instance, P38 noted, "failing to pick up an object could be passed off as a mechanical issue ... but [it] would have made me more annoyed with the software."

**Planning failure.** Participants who perceived *Planning* failure as the most severe often cited the **high-stakes consequences**—defined as the potential for the error to affect subsequent actions. One participant commented, "if a robot places something where it isn't meant to be, it can cause more problems for the later process" (P32). Others described the **cognitive confusion** these errors caused, referring

to the mental uncertainty and difficulty in attributing fault, as they struggled to determine whether the robot or they themselves were responsible for the mistake. For instance, P6 noted, “it confused me—it made me believe I was analysing the puzzle wrong.”

Those who rated *Planning* failure as the least severe typically emphasised that such errors were **easily correctable**, meaning they could be resolved without the need for expert assistance. Additionally, some participants interpreted these mistakes as **reliable human-like errors**, which elicited empathy rather than frustration. For instance, one participant noted, “that was because I could easily correct it myself” (P14), while another remarked, “this happens sometimes for humans” (P26).

Across both *Grasping* and *Planning* failures, participants’ severity judgements were strongly influenced by the robot’s level of awareness and its response to the error. Failures were generally perceived as more severe when the robot **lacked awareness or failed to correct the mistake**. Conversely, the same types of failures were mostly regarded as least severe when the robot exhibited **immediate awareness and initiated corrective action**. For instance, in the case of grasping failure, P2 remarked, “I think failing to pick up an object was the most severe because he didn’t give me a reply,” whereas P6 noted, “the least severe was failing to pick up the object as it proceeded to do so right away after recognising its mistake.”

*Freezing failure.* The dominant theme among participants who rated freezing failure as most severe was the **disruption of task flow and user frustration**. The unexpected pause in the robot’s actions was perceived as confusing and disengaging. As P50 noted, “it created an uncomfortable awkwardness as if I was waiting for it to do something but it wouldn’t.”

However, many participants who rated *Freezing* failure as the least severe emphasised their **low impact and lack of serious consequences**. This theme reflects the perception that temporary pauses did not hinder task completion or require user intervention. As P24 explained, “the robot just needed more time; it finished the task.” Others viewed the *Freezing* behaviour as **systemic or expected**, interpreting it not as a malfunction but as a natural processing delay. For instance, P47 remarked, “it’s more like a ‘processing time.’”

**4.2.2 Perceptions of Robot Recovery Strategies.** We conducted a reflexive thematic analysis to explore participants’ perceptions of the robot’s recovery strategies, focusing on how different levels of the robot’s awareness of its own failures influenced user experience. Our aim was to understand the subjective impact of these responses and to identify the underlying reasons, expressed as thematic constructs.

When the robot exhibited **no awareness** of its failure—failing to offer any verbal or physical response—participants overwhelmingly described the experience as negative. A dominant theme that emerged was **Cognitive Uncertainty and Confusion**, referring to the ambiguity and indecision participants experienced regarding their role and next steps in the task. As P35 noted, “I was very lost and unsure whether to help or not,” and P2 remarked, “I didn’t know what I should do next.” Another commonly identified theme was **Perceived Incompetence**, in which the robot’s silence was interpreted as a lack of intelligence or reliability. P15 commented,

“when it wasn’t aware of its mistake, it seemed less reliable,” while P12 reflected, “the fact that the robot did not react or comment on its mistake made it look like it could not think.” A third theme, **Frustration**, captured emotional responses to the robot’s lack of engagement, with P40 stating, “the instance where it didn’t make a reaction annoyed me.”

Conversely, when the robot demonstrated **full awareness** of its failure by acknowledging the mistake and initiating corrective action, participants reported predominantly positive experiences. One primary theme was **Autonomous Competence**, reflecting admiration for the robot’s ability to independently identify and correct its errors. P35 remarked, “its awareness of the mistake, and his ability to quickly fix the mistake impressed me,” while P53 stated, “after trying to correct the mistakes I felt reliability.” Another theme, **Explanatory Apology**, captures how the robot’s combination of explanation and apology reinforced participants’ confidence in the robot. For instance, P37 shared, “I felt more pleased when the robot indicated the reason for its mistake and apologised,” and P14 described, “I found the robot’s reaction when it failed very charming. It admitted its mistake and said that it would try again.” Finally, some participants highlighted **Human-Like Social Presence**, perceiving the robot’s responsive behaviour as more natural and relatable. As P34 noted, “I felt it was more ‘lively’ and human-like.”

An analysis of participants’ responses to the robot’s **partial awareness** of its failures revealed both positive and negative perceptions. Some participants appreciated the robot’s behaviour as providing **actionable guidance**, whereby the robot’s acknowledgement of the issue and request for assistance helped clarify the task. For example, P35 remarked, “the robot was aware and gave me instructions to correct its mistake; that was very clear for me and made it simple.” Others, however, viewed the same behaviour as a **delegation of agency**, interpreting the robot’s request as an inappropriate transfer of responsibility. As P26 noted, “it asked me to help it fix [the mistake], rather than noticing and trying to fix it itself.”

Similarly, some participants described the robot’s partial awareness as fostering **collaborative engagement**, highlighting a sense of joint problem-solving and shared effort. P31 explained, “when the robot acknowledged its mistake and asked for my help, I was willing to help fix it and more sympathetic towards it.” In contrast, others interpreted this behaviour as a **competency violation**, expressing concerns about the robot’s intelligence and autonomy. As P32 reflected, “when it recognised the mistake but did not fix the issue itself, that made me a little troubled because that means the robot is incapable of amending this problem without help.”

## 5 Discussion

This study examined how three factors shape user evaluations across repeated robot failures: the sequence of failure types (homogeneous vs. heterogeneous), perceived severity (current vs. previous failures), and robot awareness (none, partial, full) of its own errors. Our objective is to translate these findings into sequence-sensitive design principles that help preserve user trust and perceived intelligence in the face of repeated failures. The following subsections

outline each factor in turn and discuss their respective practical implications.

### 5.1 Beyond Frequency: Mixed Failure Sequences Erode Trust and Perceived Intelligence

Our results indicate that user evaluations are shaped by the type of failures and their sequence. Their order and diversity influence how people judge the robot. Although trust did not differ significantly between conditions, the overall pattern shows that heterogeneous sequences tended to produce larger declines than homogeneous ones. For perceived intelligence, this effect was statistically evident: participants rated the robot as less intelligent after a heterogeneous sequence, particularly at the second failure. This suggests that users are sensitive to the diversity of problems a robot displays, and that varying failure types may be interpreted as signs of broader system unreliability.

When different kinds of failures occur in sequence, participants may infer that the robot's issues are widespread rather than localised, reducing both trust and perceived intelligence. Conversely, when the same failure repeats, users can attribute it to a specific subsystem, making it appear more contained. Repeated failures may also make the robot's behaviour somewhat more predictable—users know what to expect and how the robot is likely to respond. A shift to a different failure type, however, disrupts these emerging expectations, adds an element of surprise, and makes the robot seem less stable overall. This interpretation aligns with EVT, which predicts that unexpected or inconsistent events trigger stronger negative reactions than anticipated ones. In our study, the second failure appears particularly important: when it differed from the first, it violated users' expectations, leading to sharper declines in perceived intelligence. Even though trust did not show statistically significant differences, the descriptive pattern followed the same direction.

These findings yield several implications for designing and deploying adaptive failure-recovery strategies. First, systems should consider the sequence of failures rather than treating each one in isolation. When a failure differs from the preceding one, repair strategies may need to escalate—providing more detailed explanations, running broader diagnostic checks, or adopting more conservative actions to minimise the chance of additional errors. Second, designers should monitor the diversity of failure types, not only their frequency. An expanding range of failures within a single interaction may signal to users that the system is unstable, warranting stronger verification mechanisms or dynamic adjustments to maintain user confidence. This study therefore extends existing trust calibration models proposed by Nettet et al. [56] and Esterwood et al. [21] by demonstrating that evaluations of trust and perceived intelligence depend not just on failure frequency but on the type and temporal structure of failures.

### 5.2 Obliviousness costs: silence on noticeable errors lowers trust and perceived intelligence.

The results of model 2 and the thematic analysis jointly indicate that the robot's *obliviousness* to failure was generally associated with poorer user experience for grasping and planning errors. When

the robot offered no acknowledgement, participants reported lower trust and perceived intelligence alongside *cognitive uncertainty*, *perceived incompetence*, and *frustration* (e.g., “I didn't know what I should do next,” P2; “it seemed less reliable,” P15). Importantly, this pattern did *not* extend to the freezing failure, for which awareness level showed no reliable influence on evaluations. This study extends existing trust calibration models by comparing different robot reactions after a failure in terms of their level of awareness, rather than evaluating individual approaches (e.g., apology or explanation). It also demonstrates how the amount of awareness required varies depending on the type of failure.

Awareness effects are not uniform across failure types. For high-severity or noticeable errors (grasping, planning), both *partial* and *full* awareness reliably improved trust compared to *no awareness*, with no reliable difference between the two levels. Perceived intelligence followed the same pattern for grasping and showed a weaker but consistent trend for planning, where full awareness significantly outperformed no awareness and partial awareness moved in the same direction. These results suggest that in manipulation tasks where the problem and its remedy are interpretable to users, *acknowledgement*, paired with either autonomous correction or a request for help, provides most of the benefit.

By contrast, freezing failure showed no reliable benefit of awareness on either trust or perceived intelligence. One interpretation is that awareness is most effective when paired with an immediate and necessary remedy. When no physical fixing is required for task completion, as in freezing, additional communicative acts may not yield higher evaluations. Another possibility is that explicitly labelling freezing as a failure may unintentionally undermine trust, while users might otherwise tolerate temporary freezing. Indeed, such short stalls are common in autonomous systems, frequently arising from perception latencies (e.g., processing algorithms [49], sensor noise [23]) or collision-avoidance planning [76]. Because these delays are routine, users may already normalise them, which can limit, or even backfire on, the perceived value of explicit awareness.

The qualitative results also highlight a *trade-off* inherent to *partial awareness*. Some participants experienced the robot's request for help as *actionable guidance* and *collaborative engagement* (“I was willing to help fix it” P31), while others interpreted the same cue as a *delegation of agency* or a *competency violation* (“rather than ... trying to fix it itself” P26; “incapable of amending this problem without help” P32). By contrast, the quantitative results show that partial awareness performs nearly as well as full awareness in manipulation failures, with only a slight advantage for full awareness in terms of trust and perceived intelligence. These findings suggest that although users interpret partial awareness in divergent ways, even incomplete explanations of failure can sustain trust in HRI, as the communicative act itself signals transparency and accountability.

The findings point to important design implications for awareness strategies in human-robot interaction. Failure detection can be achieved either through the robot's internal monitoring mechanisms or through observation of users' reactions to the failure. Depending on the operational environment and the types of potential failures, designers must determine the level of awareness a robot should demonstrate and the type of reaction required to

sufficiently regain user trust. In some contexts, it may not be necessary for the robot to detect and explicitly communicate that a failure has occurred. This was evident in our design, where awareness cues during freezing failures reduced both trust and perceived intelligence because the robot stated that a failure had occurred, even though the failure itself was subtle and largely unnoticed by users. In other contexts, however, the robot does need to detect that a failure occurred and identify the specific failure type. In our experiment, for planning and grasping failures, which were noticeable to users, combining acknowledgement with either autonomous correction or a minimal request for help was required to stabilise trust; full self-correction was preferred but not strictly necessary.

### 5.3 Expectancy violation: severe after mild amplifies trust loss.

Our findings demonstrate that the perceived severity of the current failure is the primary determinant of user evaluations: higher severity reliably reduces both trust and perceived intelligence. In our study, grasping and planning errors, classified as high-severity, were consistently rated as more severe, with participants describing them as “incompetent,” “unnatural,” or “high-stakes”. By contrast, the freezing error, classified as low-severity, was often dismissed as a simple “processing” pause. These findings align with prior work showing that failure severity shapes user reactions [70] and that different types of navigation errors can lead to varying degrees of trust reduction [25].

Trust appeared to be a history-sensitive construct, as the severity of the previous failure affected trust on the subsequent trial, whereas perceived intelligence showed little evidence of such carry-over effects. This temporal sensitivity aligns with EVT [8, 9], which proposes that people react more strongly when events deviate from their emerging expectations. In our study, when a failure occurred after a low-severity one, trust declined more sharply than when it followed a severe failure, indicating that a severe error violates a user’s expectation of relative stability established by the prior mild failure. By contrast, once a severe failure has already occurred, a subsequent failure, while still detrimental, elicits less additional shock because the expectation of unreliability has already been set. Consequently, trust is more negatively impacted when a severe failure follows a low-severity one. This pattern reflects a sequence-dependent expectancy violation and also extends current trust calibration models by showing that trust is shaped not only by failure magnitude but also by the order in which failures occur. Interestingly, this sequencing effect did not appear for perceived intelligence, which seemed to track only the immediate performance signal rather than incorporating prior failures. This dissociation suggests that trust integrates experiences over time, whereas perceived intelligence functions more as a momentary evaluation of current competence. Prior psychological research supports this interpretation, showing that severe events following milder ones heighten expectancy violations and amplify negative perceptions [5].

These findings highlight several implications for the design and deployment of interactive robots. Severe failures should be handled much like alarms, requiring clear acknowledgement and prompt

corrective action to limit damage to both trust and perceived intelligence. Importantly, trust-repair strategies should also take the sequence of failures into account. For example, when a severe failure occurs after a low-severity one, a pattern that tends to produce sharper trust declines, the system should escalate its repair efforts. This may involve offering richer explanations, or introducing additional safety checks, thereby helping to stabilise trust in more vulnerable situations.

### 5.4 Limitations and Paths Forward

This study was conducted in a controlled laboratory environment using a Tangram puzzle task and the TIAGo robot, which enabled precise manipulation of failure types and awareness levels. However, such a setup may not fully generalise to real-world scenarios in which task complexity, environmental unpredictability, and safety stakes are substantially higher. In addition, the robot’s awareness behaviours were pre-scripted and implemented through a combination of verbal explanations and predefined physical repair actions, which may not fully capture the flexibility and adaptivity of real-time autonomous responses. Only three failure instances were included per session, limiting insight into longer-term trust dynamics across extended human–robot interactions. Furthermore, the sample size was relatively moderate ( $N = 54$ ); larger and more diverse participant samples would strengthen the robustness and generalisability of the observed effects. Due to ethical constraints, all failures were required to be safe, non-damaging, and free of safety risks, which restricted our ability to manipulate failure severity more explicitly. Finally, the freezing failure consistently resulted in a successful task outcome, distinguishing it from the grasping and planning failures in terms of perceived task impact.

Building on the present findings, future research should examine robot failures and trust repair in more ecologically valid contexts such as domestic, industrial, or healthcare settings, where task complexity and safety stakes are higher. Extending beyond pre-scripted behaviours, adaptive repair strategies that leverage real-time user state detection (e.g., stress, gaze, hesitation) could provide more flexible and context-sensitive trust recovery. Longer-term studies with more frequent or varied failures are also needed to capture how trust stabilises, erodes, or recovers across extended collaborations. In addition, investigating failures of greater severity or consequence would provide insight into how well different trust repair strategies hold up under higher-risk conditions. Finally, moving from dyadic interactions to multi-human or multi-robot team contexts could reveal how failure awareness and sequence effects shape group trust and coordination, informing the development of design frameworks for scalable, adaptive trust repair in collaborative robotics.

## 6 Conclusion

This work advances our understanding of how failure sequence, robot awareness, and failure severity jointly shape user trust and perceptions of intelligence in collaborative interaction. Our findings show that both trust and perceived intelligence are history-sensitive: heterogeneous sequences of failures reduce perceived intelligence more than homogeneous ones, and prior severe failures exert lasting negative effects on trust over time. Robot awareness further

moderates these dynamics, with acknowledgement and repair stabilising trust in noticeable failures (e.g., grasping and planning) but offering no benefit, or even backfiring, for less interpretable errors such as freezing. These insights highlight the need for sequence-aware and context-sensitive trust repair strategies, where robots escalate repair efforts when severe failures follow mild ones, treat diverse failures as signals of broader instability, and adjust their explanations depending on the interpretability of the error.

## Acknowledgments

This work is partly supported by the Australian Research Council (Grant No: FT250100459 and CE260100108)

## References

- [1] Ali Afzalaghaeinaeni, Jaho Seo, Dongwook Lee, and Hanmin Lee. 2022. Design of Dust-Filtering Algorithms for LiDAR Sensors Using Intensity and Range Information in Off-Road Vehicles. *Sensors* 22, 11 (Jan. 2022), 4051. doi:10.3390/s22114051 Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.
- [2] Abdullah Alzahrani, Jauwairia Nasir, Ahmad J. Tayeb, Elisabeth André, and Muneeb I. Ahmad. 2025. What Do the Face and Voice Reveal? Investigating Trust Dynamics During Human-Robot Interaction. In *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction (HRI '25)*. IEEE Press, Melbourne, Australia, 400–409. doi:10.5555/3721488.3721540
- [3] Zhangyunfan Bai and Ke Chen. 2024. Effects of Explanations by Robots on Trust Repair in Human-Robot Collaborations. In *Artificial Intelligence in HCI*, Helmut Degen and Stavroula Ntoa (Eds.). Springer Nature Switzerland, Cham, 3–14. doi:10.1007/978-3-031-60611-3\_1
- [4] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1, 1 (2009), 71–81. doi:10.1007/s12369-008-0001-3 Place: Germany Publisher: Springer.
- [5] Roy F. Baumeister, Ellen Bratslavsky, Cathrin Finkenauer, and Kathleen D. Vohs. 2001. Bad is Stronger than Good. *Review of General Psychology* 5, 4 (Dec. 2001), 323–370. doi:10.1037/1089-2680.5.4.323 Publisher: SAGE Publications Inc.
- [6] Khairidine Benali and Praminda Caleb-Solly. 2025. The Dilemma of Decision-Making in the Real World: When Robots Struggle to Make Choices Due to Situational Constraints. In *Towards Autonomous Robotic Systems*, M. Nazmul Huda, Mingfeng Wang, and Tatiana Kalganova (Eds.). Springer Nature Switzerland, Cham, 14–26. doi:10.1007/978-3-031-72062-8\_2
- [7] Steve Benford, Eike Schneiders, Juan Pablo Martinez Avila, Praminda Caleb-Solly, Patrick Robert Brundell, Simon Castle-Green, Feng Zhou, Rachael Garrett, Kristina Höök, Sarah Whatley, Kate Marsh, and Paul Tennent. 2025. Somatic Safety: An Embodied Approach Towards Safe Human-Robot Interaction. In *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction (HRI '25)*. IEEE Press, Melbourne, Australia, 429–438. doi:10.5555/3721488.3721543
- [8] Judee K. Burgoon. 1993. Interpersonal Expectations, Expectancy Violations, and Emotional Communication. *Journal of Language and Social Psychology* 12, 1-2 (March 1993), 30–48. doi:10.1177/0261927X93121003 Publisher: SAGE Publications Inc.
- [9] Judee K. Burgoon and Jerold L. Hale. 1988. Nonverbal expectancy violations: Model elaboration and application to immediacy behaviors. *Communication Monographs* 55, 1 (March 1988), 58–79. doi:10.1080/03637758809376158 Publisher: NCA Website\_eprint: https://doi.org/10.1080/03637758809376158.
- [10] J. Carlson, R.R. Murphy, and A. Nelson. 2004. Follow-up analysis of mobile robot failures. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, Vol. 5. 4987–4994 Vol.5. doi:10.1109/ROBOT.2004.1302508 ISSN: 1050-4729.
- [11] Xiaoyu Chang, Yanheng Li, Sijia Liu, Ling Ma, and Ray Lc. 2024. "Sorry to Keep You Waiting": Recovering from Negative Consequences Resulting from Service Robot Unintended Rejection. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24)*. Association for Computing Machinery, New York, NY, USA, 96–105. doi:10.1145/3610977.3634959
- [12] Filipa Correia, Carla Guerra, Samuel Mascarenhas, Francisco S. Melo, and Ana Paiva. 2018. Exploring the Impact of Fault Justification in Human-Robot Trust. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '18)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 507–513. doi:10.5555/3237383.3237459
- [13] Antoine Cully, Jeff Clune, Danesh Tarapore, and Jean-Baptiste Mouret. 2015. Robots that can adapt like animals. *Nature* 521, 7553 (May 2015), 503–507. doi:10.1038/nature14422 Publisher: Nature Publishing Group.
- [14] Devleena Das, Siddhartha Banerjee, and Sonia Chernova. 2021. Explainable AI for Robot Failures: Generating Explanations that Improve User Assistance in Fault Recovery. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. Association for Computing Machinery, New York, NY, USA, 351–360. doi:10.1145/3434073.3444657
- [15] Munjal Desai, Poornima Kaniarasu, Mikhail Medvedev, Aaron Steinfeld, and Holly Yanco. 2013. Impact of robot failures and feedback on real-time trust. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 251–258. doi:10.1109/HRI.2013.6483596 ISSN: 2167-2148.
- [16] Munjal Desai, Kristen Stubbs, Aaron Steinfeld, and Holly Yanco. 2009. Creating trustworthy robots: Lessons and inspirations from automated systems. In *Robotics Institute Carnegie Mellon University*. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB). doi:10.1184/R1/6552464
- [17] Wen Duan, Christopher Flathmann, Nathan McNeese, Matthew J Scalia, Ruihao Zhang, Jamie Gorman, Guo Freeman, Shiwen Zhou, Allyson Ivy Hauptman, and Xiaoyun Yin. 2025. Trusting Autonomous Teammates in Human-AI Teams - A Literature Review. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–23. doi:10.1145/3706598.3713527
- [18] Mary T. Dzindolet, Scott A. Peterson, Regina A. Pomranky, Linda G. Pierce, and Hall P. Beck. 2003. The role of trust in automation reliance. *International Journal of Human-Computer Studies* 58, 6 (June 2003), 697–718. doi:10.1016/S1071-5819(03)00038-7
- [19] Connor Esterwood and Lionel P. Robert Jr. 2023. Three Strikes and you are out!: The impacts of multiple human-robot trust violations and repairs on robot trustworthiness. *Computers in Human Behavior* 142 (May 2023), 107658. doi:10.1016/j.chb.2023.107658
- [20] Connor Esterwood and Lionel + "Jr" Robert. 2025. Repairing Trust in Robots?: A Meta-analysis of HRI Trust Repair Studies with a No-Repair Condition. *HRI* 2025. doi:10.7302/24926 Accepted: 2025-01-02T20:36:35Z.
- [21] Connor Esterwood and Lionel P. Robert. 2022. Having the Right Attitude: How Attitude Impacts Trust Repair in Human-Robot Interaction. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 332–341. doi:10.1109/HRI53351.2022.9889535
- [22] Connor Esterwood and Lionel P. Robert. 2023. The theory of mind and human-robot trust repair. *Scientific Reports* 13, 1 (June 2023), 9877. doi:10.1038/s41598-023-37032-0 Publisher: Nature Publishing Group.
- [23] Davide Falanga, Suseong Kim, and Davide Scaramuzza. 2019. How Fast Is Too Fast? The Role of Perception Latency in High-Speed Sense and Avoid. *IEEE Robotics and Automation Letters* 4, 2 (April 2019), 1884–1891. doi:10.1109/LRA.2019.2898117
- [24] Tingxiang Fan, Xinjing Cheng, Jia Pan, Pinxin Long, Wenxi Liu, Ruigang Yang, and Dinesh Manocha. 2019. Getting Robots Unfrozen and Unlost in Dense Pedestrian Crowds. *IEEE Robotics and Automation Letters* 4, 2 (April 2019), 1178–1185. doi:10.1109/LRA.2019.2891491
- [25] Hideki Garcia Goo, Bob R. Schadenberg, Jan Kolkmeier, Khiet P. Truong, and Vanessa Evers. 2025. Being Sorry is the Hardest Thing: How Robots can Apologize and Learn from Mistakes to Restore People's Trust. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/3706599.3719739
- [26] Romi Gideoni, Shanee Honig, and Tal Oron-Gilad. 2024. Is It Personal? The Impact of Personally Relevant Robotic Failures (PeRFs) on Humans' Trust, Likeability, and Willingness to Use the Robot. *International Journal of Social Robotics* 16, 6 (June 2024), 1049–1067. doi:10.1007/s12369-022-00912-y
- [27] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman. 2011. A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human Factors* 53, 5 (Oct. 2011), 517–527. doi:10.1177/0018720811417254 Publisher: SAGE Publications Inc.
- [28] Antti Hietanen, Jyrki Latokartano, Alessandro Foi, Roel Pieters, Ville Kyrki, Minna Lanz, and Joni-Kristian Kämäräinen. 2021. Benchmarking pose estimation for robot manipulation. *Robotics and Autonomous Systems* 143 (Sept. 2021), 103810. doi:10.1016/j.robot.2021.103810
- [29] Shanee Honig and Tal Oron-Gilad. 2018. Understanding and Resolving Failures in Human-Robot Interaction: Literature Review and Model Development. *Frontiers in Psychology* 9 (2018). https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2018.00861
- [30] Xiaohai Hu, Aparajit Venkatesh, Yusen Wan, Guiliang Zheng, Neel Jawale, Navneet Kaur, Xu Chen, and Paul Birkmeyer. 2024. Learning to detect slip through tactile estimation of the contact force field and its entropy properties. *Mechatronics* 104 (Dec. 2024), 103258. doi:10.1016/j.mechatronics.2024.103258
- [31] Arda Inceoglu, Eren Erdal Aksoy, Abdullah Cihan Ak, and Sanem Sariel. 2021. FINO-Net: A Deep Multimodal Sensor Fusion Framework for Manipulation Failure Detection. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 6841–6847. doi:10.1109/IROS51168.2021.9636455 ISSN: 2153-0866.
- [32] Steven Jens Jorgensen, James Holley, Frank Mathis, Joshua S. Mehling, and Luis Sentis. 2019. Thermal Recovery of Multi-Limbed Robots With Electric

- Actuators. *IEEE Robotics and Automation Letters* 4, 2 (April 2019), 1077–1084. doi:10.1109/LRA.2019.2894068
- [33] Malte Jung and Pamela Hinds. 2018. Robots in the Wild: A Time for More Robust Theories of Human-Robot Interaction. *J. Hum.-Robot Interact.* 7, 1 (May 2018), 2:1–2:5. doi:10.1145/3208975
- [34] Namasivayam K, Arnav Tuli, Vishal Bindal, Himanshu Singh, Parag Singla, and Rohan Paul. 2024. Learning to Recover from Plan Execution Errors during Robot Manipulation: A Neuro-symbolic Approach. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 12632–12639. doi:10.1109/IROS58592.2024.10801831 ISSN: 2153-0866.
- [35] Patricia Kahr, Gerrit Rooks, Chris Snijders, and Martijn C. Willemsen. 2025. Good Performance Isn't Enough to Trust AI: Lessons from Logistics Experts on their Long-Term Collaboration with an AI Planning System. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3706598.3713099
- [36] Ulas Berk Karli, Shiye Cao, and Chien-Ming Huang. 2023. 'What If It Is Wrong': Effects of Power Dynamics and Trust Repair Strategy on Trust and Compliance in HRI. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23)*. Association for Computing Machinery, New York, NY, USA, 271–280. doi:10.1145/3568162.3576964
- [37] Yngve Kelch, Annette Kluge, and Laura Kunold. 2024. Would you Trust a Robot that Distrusts you?. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24)*. Association for Computing Machinery, New York, NY, USA, 588–592. doi:10.1145/3610978.3640757
- [38] Parag Khanna, Elmira Yadollahi, Mårten Björkman, Iolanda Leite, and Christian Smith. 2023. Effects of Explanation Strategies to Resolve Failures in Human-Robot Collaboration. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 1829–1836. doi:10.1109/RO-MAN57019.2023.10309394 ISSN: 1944-9437.
- [39] David Kirschner, Rosemarie Velik, Saeed Yahyanejad, Mathias Brandstötter, and Michael Hofbauer. 2016. YuMi, Come and Play with Me! A Collaborative Robot for Piecing Together a Tangram Puzzle. In *Interactive Collaborative Robotics*, Andrey Ronzhin, Gerhard Rigoll, and Roman Meshcheryakov (Eds.). Springer International Publishing, Cham, 243–251. doi:10.1007/978-3-319-43955-6\_29
- [40] Dimosthenis Kontogiorgos, Sanne van Waveren, Olle Wallberg, Andre Pereira, Iolanda Leite, and Joakim Gustafson. 2020. Embodiment Effects in Interactions with Failing Robots. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376372
- [41] E. S. Kox, J. H. Kerstholt, T. F. Hueting, and P. W. de Vries. 2021. Trust repair in human-agent teams: the effectiveness of explanations and expressing regret. *Autonomous Agents and Multi-Agent Systems* 35, 2 (June 2021), 30. doi:10.1007/s10458-021-09515-9
- [42] Johannes Maria Kraus, Julia Merger, Felix Gröner, and Jessica Pätz. 2023. 'Sorry' Says the Robot: The Tendency to Anthropomorphize and Technology Affinity Affect Trust in Repair Strategies after Error. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23)*. Association for Computing Machinery, New York, NY, USA, 436–441. doi:10.1145/3568294.3580122
- [43] Edoardo Lamon, Luka Peternel, and Arash Ajoudani. 2018. Towards a Prolonged Productivity in Industry 4.0: A Framework for Fatigue Minimisation in Robot-Robot Co-Manipulation. In *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*. 1–6. doi:10.1109/HUMANOIDS.2018.8625051 ISSN: 2164-0580.
- [44] Sophie Lane, Connor Esterwood, Dana Kulić, and Nicole Robinson. 2024. Robots That Use Physical Repair Strategies After Repeated Errors to Mitigate Trust Decline in Human-Robot Interaction: A Repeated Measures Experiment. In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*. 936–943. doi:10.1109/RO-MAN60168.2024.10731465 ISSN: 1944-9437.
- [45] Christine P Lee, Pragathi Praveena, and Bilge Mutlu. 2024. REX: Designing User-centered Repair and Explanations to Address Robot Failures. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, New York, NY, USA, 2911–2925. doi:10.1145/3643834.3661559
- [46] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46, 1 (March 2004), 50–80. doi:10.1518/hfes.46.1.50\_30392 Publisher: SAGE Publications Inc.
- [47] Gregory LeMasurier, Alvika Gautam, Zhao Han, Jacob W. Crandall, and Holly A. Yanco. 2024. Reactive or Proactive? How Robots Should Explain Failures. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24)*. Association for Computing Machinery, New York, NY, USA, 413–422. doi:10.1145/3610977.3634963
- [48] Roy J. Lewicki and Chad Brinsfield. 2017. Trust Repair. *Annual Review of Organizational Psychology and Organizational Behavior* 4, Volume 4, 2017 (March 2017), 287–313. doi:10.1146/annurev-orgpsych-032516-113147 Publisher: Annual Reviews.
- [49] Liangkai Liu, Zheng Dong, Yanzhi Wang, and Weisong Shi. 2022. Prophet: Realizing a Predictable Real-time Perception Pipeline for Autonomous Vehicles. In *2022 IEEE Real-Time Systems Symposium (RTSS)*. 305–317. doi:10.1109/RTSS55097.2022.00034 ISSN: 2576-3172.
- [50] Joseph B. Lyons, Izz aldin Hamdan, and Thy Q. Vo. 2023. Explanations and trust: What happens to trust when a robot partner does something unexpected? *Computers in Human Behavior* 138 (Jan. 2023), 107473. doi:10.1016/j.chb.2022.107473
- [51] Akihiro Maehigashi and Seiji Yamada. 2024. Effect of Wait Time on Trust and Reliance in Human-Robot Interaction. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24)*. Association for Computing Machinery, New York, NY, USA, 722–726. doi:10.1145/3610978.3640611
- [52] Amama Mahmood, Jeanie W Fung, Isabel Won, and Chien-Ming Huang. 2022. Owning Mistakes Sincerely: Strategies for Mitigating AI Errors. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3491102.3517565
- [53] Roger C. Mayer, James H. Davis, and F. David Schoorman. 1995. An Integrative Model of Organizational Trust. *The Academy of Management Review* 20, 3 (1995), 709–734. doi:10.2307/258792 Publisher: Academy of Management.
- [54] Nicole Mirnig, Gerald Stollnberger, Markus Miksch, Susanne Stadler, Manuel Giuliani, and Manfred Tscheligi. 2017. To Err Is Robot: How Humans Assess and Act toward an Erroneous Social Robot. *Frontiers in Robotics and AI* 4 (May 2017). doi:10.3389/frobt.2017.00021 Publisher: Frontiers.
- [55] Cecilia G. Morales, Elizabeth J. Carter, Xiang Zhi Tan, and Aaron Steinfeld. 2019. Interaction Needs and Opportunities for Failing Robots. In *Proceedings of the 2019 on Designing Interactive Systems Conference (DIS '19)*. Association for Computing Machinery, New York, NY, USA, 659–670. doi:10.1145/3322276.3322345
- [56] Birthe Nessel, Marta Romeo, Gnanathusharan Rajendran, and Helen Hastie. 2023. Robot Broken Promise? Repair strategies for mitigating loss of trust for repeated failures. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 1389–1395. doi:10.1109/RO-MAN57019.2023.10309558 ISSN: 1944-9437.
- [57] Saumya Pareek, Sarah Schömb, Eduardo Velloso, and Jorge Goncalves. 2025. 'It's Not the AI's Fault Because It Relies Purely on Data': How Causal Attributions of AI Decisions Shape Trust in AI Systems. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3706598.3713468
- [58] Tyson Govan Phillips, Nicky Guenther, and Peter Ross McAree. 2017. When the Dust Settles: The Four Behaviors of LiDAR in the Presence of Fine Airborne Particulates. *Journal of Field Robotics* 34, 5 (2017), 985–1009. doi:10.1002/rob.21701\_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/rob.21701>
- [59] Babiche L. Pompe, Ella Velner, and Khiet P. Truong. 2022. The Robot That Showed Remorse: Repairing Trust with a Genuine Apology. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE Press, Napoli, Italy, 260–265. doi:10.1109/RO-MAN53752.2022.9900860
- [60] Aniketh Ramesh, Rustam Stolkin, and Manolis Chiou. 2022. Robot Vitals and Robot Health: Towards Systematically Quantifying Runtime Performance Degradation in Robots Under Adverse Conditions. *IEEE Robotics and Automation Letters* 7, 4 (Oct. 2022), 10729–10736. doi:10.1109/LRA.2022.3192612
- [61] Ameer H. Sabry and Ungku Anisa Bte Ungku Amirulddin. 2024. A review on fault detection and diagnosis of industrial robots and multi-axis machines. *Results in Engineering* 23 (Sept. 2024), 102397. doi:10.1016/j.rineng.2024.102397
- [62] Maha Salem and K. Dautenhahn. 2015. Evaluating Trust and Safety in HRI : Practical Issues and Ethical Challenges. In *Emerging Policy and Ethics of Human-Robot Interaction : A Workshop at 10th ACM/IEEE Int Conf on Human-Robot Interaction (HRI 2015)*. ACM Press, USA. <https://uhra.herts.ac.uk/id/eprint/13823/> Num Pages: 3.
- [63] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. 2015. 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 (HRI '15)*. Association for Computing Machinery, New York, NY, USA, 141–148. doi:10.1145/2696454.2696497
- [64] Kristin E. Schaefer, Jessie Y. C. Chen, James L. Szalma, and P. A. Hancock. 2016. A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 58, 3 (May 2016), 377–400. doi:10.1177/0018720816634228
- [65] Lukas Schichler, Karin Festl, and Selim Solmaz. 2025. Robust Multi-Sensor Fusion for Localization in Hazardous Environments Using Thermal, LiDAR, and GNSS Data. *Sensors* 25, 7 (Jan. 2025), 2032. doi:10.3390/s25072032 Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.
- [66] Sarah Schömb, Saumya Pareek, Jorge Goncalves, and Wafa Johal. 2024. Robot-Assisted Decision-Making: Unveiling the Role of Uncertainty Visualisation and Embodiment. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)* (Honolulu, HI, USA). ACM.
- [67] Sarah Strohkorb Sebo, Priyanka Krishnamurthi, and Brian Scassellati. 2019. "I Don't Believe You": Investigating the Effects of Robot Trust Violation and Repair. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 57–65. doi:10.1109/HRI.2019.8673169 ISSN: 2167-2148.

- [68] Sim B. Sitkin and Nancy L. Roth. 1993. Explaining the Limited Effectiveness of Legalistic “Remedies” for Trust/ Distrust. *Organization Science* 4, 3 (1993), 367–392. doi:10.1287/orsc.4.3.367 Publisher: INFORMS.
- [69] Maia Stiber. 2022. Effective Human-Robot Collaboration via Generalized Robot Error Management Using Natural Human Responses. In *Proceedings of the 2022 International Conference on Multimodal Interaction (ICMI '22)*. Association for Computing Machinery, New York, NY, USA, 673–678. doi:10.1145/3536221.3557028
- [70] Maia Stiber and Chien-Ming Huang. 2021. Not All Errors Are Created Equal: Exploring Human Responses to Robot Errors with Varying Severity. In *Companion Publication of the 2020 International Conference on Multimodal Interaction (ICMI '20 Companion)*. Association for Computing Machinery, New York, NY, USA, 97–101. doi:10.1145/3395035.3425245
- [71] Maia Stiber, Russell H. Taylor, and Chien-Ming Huang. 2023. On Using Social Signals to Enable Flexible Error-Aware HRI. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23)*. Association for Computing Machinery, New York, NY, USA, 222–230. doi:10.1145/3568162.3576990
- [72] Ramtin Tabatabaei, Vassilis Kostakos, and Wafa Johal. 2025. Gazing at Failure: Investigating Human Gaze in Response to Robot Failure in Collaborative Tasks. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 939–948. doi:10.1109/HRI61500.2025.10973935
- [73] Gareth Terry and Nikki Hayfield. 2021. Summary and conclusions. In *Essentials of thematic analysis*. American Psychological Association, Washington, DC, US, 85–88. doi:10.1037/0000238-007
- [74] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-Robot Interaction. *J. Hum.-Robot Interact.* 10, 2 (Feb. 2021), 13:1–13:32. doi:10.1145/3439720
- [75] Edward C. Tomlinson, Brian R. Dineen, and Roy J. Lewicki. 2004. The Road to Reconciliation: Antecedents of Victim Willingness to Reconcile Following a Broken Promise. *Journal of Management* 30, 2 (April 2004), 165–187. doi:10.1016/j.jm.2003.01.003
- [76] Peter Trautman and Andreas Krause. 2010. Unfreezing the robot: Navigation in dense, interacting crowds. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 797–803. doi:10.1109/IROS.2010.5654369 ISSN: 2153-0866.
- [77] Daniel Ullman and Bertram F Malle. 2023. MDMT: Multi-Dimensional Measure of Trust v2. (2023). doi:10.48550/arXiv.2311.14887
- [78] Sanne van Waveren, Elizabeth J. Carter, and Iolanda Leite. 2019. Take One For the Team: The Effects of Error Severity in Collaborative Tasks with Social Robots. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents (IVA '19)*. Association for Computing Machinery, New York, NY, USA, 151–158. doi:10.1145/3308532.3329475
- [79] Lennart Wachowiak, Andrew Fenn, Haris Kamran, Andrew Coles, Oya Celiktutan, and Gerard Canal. 2024. When Do People Want an Explanation from a Robot?. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24)*. Association for Computing Machinery, New York, NY, USA, 752–761. doi:10.1145/3610977.3634990
- [80] Xinyi Zhang, Sun Kyong Lee, Whani Kim, and Sowon Hahn. 2023. “Sorry, it was my fault”: Repairing trust in human-robot interactions. *International Journal of Human-Computer Studies* 175 (July 2023), 103031. doi:10.1016/j.ijhcs.2023.103031
- [81] Xinyi Zhang, Sun Kyong Lee, Hoyoung Maeng, and Sowon Hahn. 2023. Effects of Failure Types on Trust Repairs in Human–Robot Interactions. *International Journal of Social Robotics* 15, 9 (Oct. 2023), 1619–1635. doi:10.1007/s12369-023-01059-0
- [82] Yan Zhang, Tharaka Sachintha Ratnayake, Cherie Sew, Jarrod Knibbe, Jorge Goncalves, and Wafa Johal. 2025. Can you pass that tool?: Implications of Indirect Speech in Physical Human-Robot Collaboration. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3706598.3713780
- [83] Yuepeng Zhang, Jun Wu, Bo Gao, Linzhong Xia, Chen Lu, Hui Wang, and Guangzhong Cao. 2025. Fault Types and Diagnostic Methods of Manipulator Robots: A Review. *Sensors* 25, 6 (Jan. 2025), 1716. doi:10.3390/s25061716 Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.
- [84] Fan Zhu, Liangliang Wang, Yilin Wen, Lei Yang, Jia Pan, Zheng Wang, and Wenping Wang. 2021. Failure Handling of Robotic Pick and Place Tasks With Multimodal Cues Under Partial Object Occlusion. *Frontiers in Neurobotics* 15 (March 2021). doi:10.3389/fnbot.2021.570507 Publisher: Frontiers.

## A Appendix: Model Results

Tables A1–A3 present the detailed results for the three cumulative link mixed models, corresponding respectively to Formulas 1, 2, and 3.

$$\text{Dependent\_Variable} \sim \text{Prefailure\_Rating} + \text{Sequence\_Condition} \times \text{Failure\_Number} + \text{Current\_Awareness} + \text{Previous\_Awareness} + (1 \mid \text{Participant}) + (1 \mid \text{Item}) \quad (1)$$

$$\text{Dependent\_Variable} \sim \text{Prefailure\_Rating} + \text{Failure\_Number} + \text{Current\_Failure\_Type} \times \text{Current\_Awareness} + \text{Previous\_Awareness} + (1 \mid \text{Participant}) + (1 \mid \text{Item}) \quad (2)$$

$$\text{Dependent\_Variable} \sim \text{Prefailure\_Rating} + \text{Failure\_Number} + \text{Current\_Awareness} + \text{Previous\_Awareness} + \text{Severity\_Current\_Failure} \times \text{Severity\_Previous\_Failure} + (1 \mid \text{Participant}) + (1 \mid \text{Item}) \quad (3)$$

Where *Dependent Variable* represents the rating from one of the scales (trust or perceived intelligence); *Prefailure Rating* reflects the rating given by the participant on the preceding failure; *Failure Number* indicates the position of the current failure within the sequence; *Current Awareness* and *Previous Awareness* denote the robot's verbal awareness level in response to failure (no, partial, or full awareness) on the current and previous trial, respectively; *Severity Current Failure* and *Severity Previous Failure* represent the severity level of the current and preceding failures; *Sequence Condition* specifies the between-sequence manipulation (homogeneous vs. heterogeneous failure sequence); *Current Failure Type* identifies the category of the current failure; and the interaction terms assess whether the effect of one predictor depends on the level of another. Finally, (Participant) and (Item) indicate random intercepts, allowing baseline ratings to vary across participants and across scale items.

**Table A1: This table presents the results of the CLMM analysis for Model 1.**

| Model 1<br>Fixed Effects                            | Trust   |         |        |       | Perceived Intelligence |        |        |       |
|---|---------|---------|--------|-------|------------------------|--------|--------|-------|
|   | Est     | SE      | z      | p     | Est                    | SE     | z      | p     |
| <b>Main Predictors</b>                              |         |         |        |       |                        |        |        |       |
| Condition (Homogeneous)                             | 0.466   | 0.546   | 0.853  | .394  | 0.913                  | 0.481  | 1.900  | .057  |
| Failure Number (Third)                              | -0.071  | 0.229   | -0.309 | .758  | 0.103                  | 0.284  | 0.363  | .717  |
| Failure Number (Third) ×<br>Condition (Homogeneous) | 0.014   | 0.405   | 0.034  | .973  | -0.821                 | 0.500  | -1.642 | .101  |
| <b>Covariate/Control Factors</b>                    |         |         |        |       |                        |        |        |       |
| Previous Rating                                     | 0.296   | 0.136   | 2.171  | <.05  | 0.724                  | 0.194  | 3.733  | <.001 |
| Current Awareness (Partial)                         | 1.007   | 0.297   | 3.395  | <.001 | 0.828                  | 0.349  | 2.376  | <.05  |
| Current Awareness (Full)                            | 1.718   | 0.310   | 5.536  | <.001 | 1.434                  | 0.377  | 3.808  | <.001 |
| Previous Awareness (Partial)                        | 0.845   | 0.310   | 2.725  | <.01  | 0.191                  | 0.377  | 0.507  | .612  |
| Previous Awareness (Full)                           | 0.397   | 0.311   | 1.276  | .202  | -0.842                 | 0.387  | -2.177 | <.05  |
| <b>Random Effects</b>                               |         |         |        |       |                        |        |        |       |
| Participant   | 2.616   | 1.617   |        |       | 1.233                  | 1.110  |        |       |
| Item  | 0.000   | 0.000   |        |       | 0.030                  | 0.174  |        |       |
| <b>Model Fit</b>                                    |         |         |        |       |                        |        |        |       |
|   | logLik  | AIC     |        |       | logLik                 | AIC    |        |       |
|   | -557.09 | 1146.19 |        |       | -324.62                | 677.24 |        |       |

**Table A2: This table presents the results of the CLMM analysis for Model 2.**

| Model 2<br>Fixed Effects                                    | Trust   |         |        |       | Perceived Intelligence |        |        |       |
|---|---------|---------|--------|-------|------------------------|--------|--------|-------|
|   | Est     | SE      | z      | p     | Est                    | SE     | z      | p     |
| <b>Main Predictors</b>                                      |         |         |        |       |                        |        |        |       |
| Current Failure Type (Grasping)                             | -4.990  | 0.627   | -7.954 | <.001 | -4.621                 | 0.773  | -5.979 | <.001 |
| Current Failure Type (Planning)                             | -4.009  | 0.590   | -6.789 | <.001 | -3.811                 | 0.709  | -5.379 | <.001 |
| Current Awareness (Partial)                                 | -0.967  | 0.546   | -1.771 | .076  | -0.928                 | 0.665  | -1.396 | .163  |
| Current Awareness (Full)                                    | -0.415  | 0.564   | -0.736 | .462  | -0.755                 | 0.723  | -1.043 | .297  |
| Current Failure (Grasping) ×<br>Current Awareness (Partial) | 3.669   | 0.787   | 4.663  | <.001 | 2.936                  | 0.924  | 3.176  | <.01  |
| Current Failure (Planning) ×<br>Current Awareness (Partial) | 2.845   | 0.766   | 3.712  | <.001 | 2.347                  | 0.921  | 2.548  | <.05  |
| Current Failure (Grasping) ×<br>Current Awareness (Full)    | 4.149   | 0.821   | 5.054  | <.001 | 3.304                  | 0.967  | 3.417  | <.001 |
| Current Failure (Planning) ×<br>Current Awareness (Full)    | 2.909   | 0.778   | 3.739  | <.001 | 3.159                  | 0.982  | 3.219  | <.01  |
| <b>Covariate/Control Factors</b>                            |         |         |        |       |                        |        |        |       |
| Previous Rating   | 0.881   | 0.153   | 5.765  | <.001 | 1.158                  | 0.213  | 5.424  | <.001 |
| Failure Number (Third)                                      | -0.070  | 0.194   | -0.362 | .717  | -0.179                 | 0.245  | -0.733 | .464  |
| Previous Failure Type (Grasping)                            | 0.841   | 0.356   | 2.358  | <.05  | -0.172                 | 0.393  | -0.436 | .663  |
| Previous Failure Type (Planning)                            | 1.067   | 0.350   | 3.052  | <.01  | 0.748                  | 0.394  | 1.897  | .058  |
| Previous Awareness (Partial)                                | 0.791   | 0.320   | 2.471  | <.05  | 0.193                  | 0.402  | 0.481  | .631  |
| Previous Awareness (Full)                                   | 0.253   | 0.324   | 0.780  | .435  | -1.227                 | 0.417  | -2.941 | <.01  |
| <b>Random Effects</b>                                       |         |         |        |       |                        |        |        |       |
| Participant   | 2.392   | 1.547   |        |       | 1.699                  | 1.304  |        |       |
| Item  | 0.000   | 0.000   |        |       | 0.030                  | 0.174  |        |       |
| <b>Model Fit</b>  |         |         |        |       |                        |        |        |       |
|   | logLik  | AIC     |        |       | logLik                 | AIC    |        |       |
|   | -501.45 | 1046.90 |        |       | -292.61                | 625.21 |        |       |

**Table A3: This table presents the results of the CLMM analysis for Model 3.**

| <b>Model 3</b><br><b>Fixed Effects</b>            | <b>Trust</b>  |            |          |          | <b>Perceived Intelligence</b> |            |          |          |
|---|---------------|------------|----------|----------|-------------------------------|------------|----------|----------|
|   | <b>Est</b>    | <b>SE</b>  | <b>z</b> | <b>p</b> | <b>Est</b>                    | <b>SE</b>  | <b>z</b> | <b>p</b> |
| <b>Main Predictors</b>                            |               |            |          |          |                               |            |          |          |
| Severity Current Failure (High)                   | -1.267        | 0.788      | -1.609   | .108     | -1.514                        | 0.723      | -2.093   | <.05     |
| Severity Previous Failure (High)                  | 1.728         | 0.791      | 2.184    | <.05     | 0.932                         | 0.741      | 1.258    | .209     |
| Current (High) × Previous Failure Severity (High) | -1.119        | 0.871      | -1.284   | .199     | -0.855                        | 0.836      | -1.023   | .306     |
| <b>Covariate/Control Factors</b>                  |               |            |          |          |                               |            |          |          |
| Previous Rating                                   | 0.738         | 0.153      | 4.812    | <.001    | 1.052                         | 0.208      | 5.057    | <.001    |
| Failure Number (Third)                            | -0.024        | 0.192      | -0.124   | 0.901    | -0.126                        | 0.240      | -0.525   | .599     |
| Current Awareness (Partial)                       | 1.172         | 0.306      | 3.833    | <.001    | 0.841                         | 0.363      | 2.320    | <.05     |
| Current Awareness (Full)                          | 1.972         | 0.324      | 6.085    | <.001    | 1.524                         | 0.396      | 3.846    | <.001    |
| Previous Awareness (Partial)                      | 0.871         | 0.327      | 2.710    | <.01     | 0.210                         | 0.395      | 0.531    | .595     |
| Previous Awareness (Full)                         | 0.407         | 0.327      | 1.246    | .213     | -1.089                        | 0.406      | -2.685   | <.01     |
| <b>Random Effects</b>                             |               |            |          |          |                               |            |          |          |
| Participant                                       | 2.467         | 1.571      |          |          | 1.548                         | 1.244      |          |          |
| Item  | 0.000         | 0.000      |          |          | 0.028                         | 0.169      |          |          |
| <b>Model Fit</b>                                  |               |            |          |          |                               |            |          |          |
|   | <b>logLik</b> | <b>AIC</b> |          |          | <b>logLik</b>                 | <b>AIC</b> |          |          |
|   | -518.75       | 1069.51    |          |          | -304.18                       | 638.36     |          |          |