If It's Safe, Don't Ask: Decreasing Frustration through User Involvement for Risky Robot Behaviors

Jan Leusmann
LMU Munich
Munich, Germany
jan.leusmann@ifi.lmu.de

Sarah Schömbs
University of Melbourne
Melbourne, Australia
sschombs@student.unimelb.edu.au

Maximilian Diedrich
LMU Munich
Munich, Germany
m.diedrich@campus.lmu.de

Florian Müller
TU Darmstadt
Darmstadt, Germany
florian.mueller@tu-darmstadt.de

Abstract

Human–robot collaboration faces a speed–accuracy trade-off (SAT): higher speed lowers latency but increases errors; lower speed improves accuracy but extends waiting time. Both pathways can frustrate users in research and real-world deployments. Despite this importance, the impact of SAT on frustration and how to mitigate it remains underexplored. We conducted a user study ($N = 24$) in which participants collaborated with a robot in an assembly task. We investigate three levels of SAT (conservative, moderate, risky), and examine how uncertainty communication and offering decision autonomy affect user frustration. Our results show that user frustration is highest for risky robot behavior. User involvement decreased frustration for risky behaviors, but increased it for conservative ones, while verbal uncertainty communication had no effects. We further found that perceived transparency, agency, intelligence, and utility of the robot increase with conservative SAT, with user workload decreasing. We propose that user involvement is advisable in higher-risk settings to mitigate user frustration, whereas autonomous operation is preferable in lower-risk scenarios.

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**.

Keywords

human robot interaction, speed-accuracy tradeoff, uncertainty communication, frustration

ACM Reference Format:
Jan Leusmann, Sarah Schömbs, Maximilian Diedrich, and Florian Müller. 2026. If It's Safe, Don't Ask: Decreasing Frustration through User Involvement for Risky Robot Behaviors. In *Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction (HRI '26), March 16–19, 2026, Edinburgh, Scotland, UK*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3757279.3785545

HRI '26, Edinburgh, Scotland, UK

ACM ISBN 979-8-4007-2128-1/2026/03
https://doi.org/10.1145/3757279.3785545

1 Introduction

Collaborative robots operate under a speed–accuracy trade-off: going faster reduces execution time but increases error risk, slowing down avoids mistakes but costs time. Both options can potentially frustrate users, either through execution errors or needless pauses [12, 28]. As robots leave structured factories and become collaborative partners across homes and workspaces, they face uncertain, dynamic contexts. In such settings, robots often lack reliable estimates of their likelihood of success before acting, which yields overly conservative pauses or risky errors. Either outcome can frustrate users and undermine the success of the collaboration.

One strategy to mitigate frustration is to increase transparency by communicating the robot's uncertainty. Uncertainty cues can help users better understand and predict the robot's behavior, potentially calibrating trust and enabling more informed decisions [23, 36, 39]. In HRI, uncertainty has been conveyed through visual signals, hesitation gestures, or explicit verbal statements [13, 14, 22], with mixed effects on user experience. Another approach is to involve the user directly in the decision-making process, giving them oversight and agency over risky robot actions [21]. While user involvement can mitigate negative outcomes and foster accountability, it also introduces interaction overhead and may disrupt task flow, especially if the robot frequently defers to the human [34]. Thus, both uncertainty communication and user decision-making have the potential to reduce frustration by increasing user control and awareness. However, when applied at the wrong time, they may disrupt workflow and increase frustration instead.

We hypothesize that the decision when to involve the user and when to communicate uncertainty without adding unnecessary friction depends on the risk induced by the speed-accuracy tradeoff. When the robot behaves conservatively, with errors unlikely or minor, involving the user may create unnecessary friction and increase frustration by slowing progress. Conversely, when the robot behaves riskily, where errors are more frequent or potentially costly, giving the user the option to intervene may reduce frustration by preventing visible mistakes or damage.

We conducted a user study ($N = 24$) in which participants collaborated with a robot in an assembly task to address this gap by systematically investigating two strategies for handling robotic uncertainty: (1) verbal uncertainty communication as a form of transparency and (2) explicit user decision-making as a mechanism for adjustable autonomy. By combining these strategies with different levels of the speed-accuracy tradeoff (conservative, moderate, and risky), we examined how they interact to shape frustration, workload, perceived agency, intelligence, utility, and transparency. This allowed us to identify when transparency and user control genuinely improve collaboration, and when they may unintentionally increase frustration or inefficiency.

We found that user frustration and related constructs are significantly affected by the robot's SAT behavior, while being moderated by the ability of users to make decisions. User frustration was highest when the robot utilized risky behavior and users had no control over its decisions. For risky behaviors, user decisions significantly decreased frustration; for moderate behaviors, this effect was weaker, and for conservative behaviors, user decisions increased frustration by slowing progress. Uncertainty communication alone had no significant effect on frustration. Perceived transparency, agency, and intelligence were significantly higher in the conservative configuration than in the risk configuration. Further, uncertainty cues enhance perceived transparency in risky configurations, and including users in the decision process yielded higher utility. These findings indicate that simply adding transparency or user control does not always improve user experience. Instead, robots should only involve humans when it matters. This insight helps HRI researchers configure robot behavior in studies to minimize frustration-related confounds and guides designers toward building collaborative robots that are efficient and less frustrating for their human partners.

2 Related Work

In the following, we present prior work on how the speed-accuracy tradeoff might affect user frustration in HRI, introduce uncertainty communication in HRI, and discuss the advantages and disadvantages of including users in robots' decisions.

2.1 Speed-Accuracy Tradeoff & Frustration

Frustration is a key factor influencing the success of human–robot collaboration, as it impacts engagement, learning, and trust [32, 43]. Prior work shows that robot errors are a major driver of frustration: technical failures such as dropping or misplacing objects elicit stronger negative reactions than social norm violations [1, 9]. Users also tend to hold robots to higher standards of reliability than human partners, which amplifies their negative responses to errors [19]. Destructive errors, which undo user progress or create safety risks, are particularly damaging, while repeated minor inaccuracies can be equally frustrating over time [31, 40].

One common source of such errors is the inherent tradeoff between speed and accuracy in robotic systems. Higher movement speed typically reduces precision, increasing the likelihood of mistakes [25]. Conversely, overly slow robots may frustrate users by slowing down task completion or appearing incompetent [2]. Recent work suggests a bell-shaped relationship between speed and user experience: moderate speeds are perceived as both safe and efficient, whereas extreme speeds (too fast or too slow) lead to poorer trust and comfort ratings [2, 45].

The speed–accuracy tradeoff (SAT) is therefore a central design challenge in HRI. Riskier, high-speed robot behavior can accelerate collaboration but also raises the risk of destructive errors, while conservative behavior prioritizes safety and accuracy at the cost of efficiency [24]. These dynamics are critical because frustration not only undermines the immediate interaction but also damages long-term acceptance of the robot as a teammate [40]. Our study builds on this literature by explicitly manipulating SAT to examine how different performance profiles—conservative, moderate, or risky—affect frustration, workload, and user perceptions, as well as how communication strategies can mitigate these effects.

2.2 Uncertainty Communication and User Decisions in HRI

Robots are imperfect systems acting under varying uncertainty due to external factors such as changing environments, sensor noise, as well as their inherent probabilistic nature. When a robot is uncertain, its actions may be incorrect, leading to inefficiency, safety hazards, or user frustration. In human-AI interaction, uncertainty communication is investigated as a way to improve system transparency and assisted decision-making, often with the overarching objective to calibrate appropriate user trust in these systems [5, 26], with mixed results depending on the type of uncertainty communication[27, 44].

In HRI, communicating uncertainty is particularly critical, given that the consequences of robot failures can range from causing user frustration to severe safety hazards. Despite its importance, work on communicating uncertainty in HRI remains scarce [13, 20, 35]. Schömbs et al. [36] show that the degree of uncertainty communicated significantly influences perceived transparency of the robot during a collaborative assisted-decision-making task. Notably, the uncertainty was communicated through either graphical uncertainty cues or "embodied" through the robot's velocity, informed by hesitation gestures [13], and presented at each assessment during the collaborative task, akin to local model confidence communication. Similarly, LeMasurier et al. [17] investigated proactive versus reactive communication of a robot erring, with a proactive communication supporting intelligence attribution and perceived system trustworthiness compared to a reactive, post-failure communication.

While these insights are important, human-robot collaboration often relies on verbal cues to accomplish a common goal. Therefore, we investigate *verbal uncertainty cues*, in line with prior works investigating uncertainty communication in LLM interaction [14], which shows that uncertainty expressed through a first-person narrative supports decision accuracy and tempers over-reliance. However, the effects of verbal uncertainty cues have been underexplored in physical HRI, where users simultaneously observe a robot's actions.

Uncertainty communication alone does not guarantee better collaboration. Keeping users in the loop during collaborative tasks is crucial for the user to adhere oversight and agency over robotic systems [4, 21]. Whereas this requires the user to often actively take on decision-making, prior research reveals people's desire to hand over the final say to the robot completely, indicating a reduced willingness for active oversight through decision-making [36], which is non-trivial given the fallible nature of robotic systems. This aligns with the concept of *adjustable autonomy*, where control dynamically shifts between human and robot depending on the situation and level of risk [33]. Prior work highlights the value of such deferrals, as when robots "ask for help" instead of making risky decisions alone, users can prevent potentially costly or dangerous mistakes [18, 29].

3 User Study

In collaborative HRI settings, failures can be frustrating or costly. To improve collaboration, we focus on frustration as a key measure and examine how robot performance and communication shape user experience. Based on prior work, we formulated three hypotheses:

- **H1** User frustration increases with riskier robot behaviors, which are faster but inaccurate.
- **H2** Uncertainty communication, as a form of transparency, mitigates user frustration for risky robot behavior, but increases it for conservative behavior.
- **H3** Including users in the robot's decision-making, by giving them the option to take over, mitigates frustration for risky robot behavior, but increases it for conservative behavior.

To test these hypotheses, we designed a collaborative assembly task in which participants and the robot jointly built sample towers. Participants were told they were working in a laboratory setting, handling both non-hazardous and hazardous blocks. The robot was introduced as a helper specifically responsible for handling the hazardous blocks, reflecting real-world scenarios where robots assist humans in high-stakes tasks with varying task responsibilities.

3.1 Study Design

We conducted a within-subjects experiment with a 3 × 2 × 2 factorial design to investigate how the Speed-Accuracy Tradeoff (SAT), Uncertainty Communication, and User Decision influence user frustration, workload, perceived agency, intelligence, utility, and transparency in a collaborative task. This design resulted in 12 experimental conditions. To mitigate order effects, we balanced condition order using a Latin square, blocking User Decision so that participants either experienced all six *decision* conditions first or all six *no-decision* conditions first, counterbalanced across participants.

Speed-Accuracy Tradeoff (SAT): This factor had three levels: *conservative* (slow and accurate), *moderate* (medium speed, slightly inaccurate), and *risky* (fast and inaccurate). In robotic systems, higher movement speed often reduces placement precision. Because speed and accuracy are inherently linked, we treated them as a single combined factor. The robot's behavior was consistent within each SAT level. We used this deterministic behavior over probabilistic behavior to have full control over what participants experience. In *conservative*, the robot always placed blocks accurately. In *moderate*, blocks were placed with a slight misalignment, requiring users to correct them with tongs. In *risky*, the robot failed to execute its task correctly, which caused the tower to collapse completely, forcing participants to rebuild it completely. This consistent behavior allowed us to simulate predictable performance differences and ensured comparable difficulty across participants.

Uncertainty Communication: This factor had two levels: *uncertainty cue* and *no uncertainty cue*. In the *uncertainty cue* condition, before placing a block, the robot verbally announced: *"I am uncertain whether I can accurately place the block on top of the tower."* This design was informed by prior findings that users prefer explicit, active communication of uncertainty over passive cues [14].

User Decision This factor also had two levels: *decision* and *no decision*. In the *decision* condition, the robot asked *"Should I place the block on top of the tower or move it to the safe area?"* The robot then waited for the participant's verbal response. In the *no decision* condition, the robot acted autonomously without asking. We blocked this factor to ensure that participants do not have to constantly switch between interaction styles within a single session, reducing confusion and learning effects.

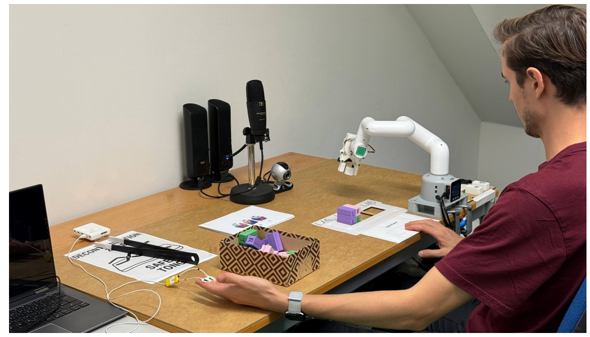

Figure 1: Picture of the Study Setup.

3.2 Task

In each condition, participants collaborated with the robot to build three sample towers, each consisting of five blocks. Participants could freely choose the order in which to build the towers. A printed manual showed the required block color order for the base of each tower. To build a tower, participants first assembled the four-block base by hand, following the manual. The final "hazardous" block could only be handled by the robot. Once a base was complete, participants had to verbally request the robot to place the last block. The robot's behavior at this point varied depending on the experimental condition, as described in subsection 3.1.

Depending on the SAT level and the user's decision, the robot either: accurately placed the block in the designated safe area (see 1), on top of a tower (*conservative*), slightly misaligned (*moderate*), or destroyed the already built tower while attempting to place it (*risky*). In the latter case, the participants had to start building the respective tower from scratch. Participants were required to use safety tongs whenever they needed to handle the "hazardous" block. The task was completed when all three towers had been successfully built and all top blocks were correctly placed. Tolerance markings on the blocks provided a clear visual cue for participants to determine whether the robot's placement was accurate (see 1).

3.3 Apparatus

We implemented the study using a 6-DOF myCobot 280 robotic arm[1] controlled via its Python API. We operated the robot using a Wizard-of-Oz setup and developed a custom GUI that allowed the wizard to trigger pre-recorded movements and verbal responses for each experimental condition.

We mounted the robot on a table and created a workspace with dedicated zones for tower building and for the safe placement of blocks (see 1). We laser-cut templates to mark the exact positions where participants were instructed to build the towers and 3D-printed colored blocks that participants used for construction. We placed these

[1] https://www.elephantrobotics.com/en/mycobot-en/

colored blocks in a box within easy reach of participants and placed the "hazardous" white blocks, which only the robot was allowed to move, in predefined locations next to the robot. We placed the safety tongs, which participants had to use whenever handling "hazardous" blocks in a contamination area, and asked participants to always put them back in their place when they were done using them. This prevented participants from keeping the tongs in their hands throughout the task and reinforced the notion that hazardous blocks required special handling. We measured skin conductance during the study using the BITalino biomedical toolkit by attaching electrodes to the index and middle fingers of each participant's left hand. To generate the robot's verbal utterances, we used .

3.4 Procedure

We first introduced participants to the study, answered questions, and obtained informed consent. Participants then completed a demographic questionnaire and the Personal Level Negative Attitude module from the General Attitudes Towards Robots Scale (GATORS) [15].

Afterwards, we asked participants to take a seat at the workspace and started the study with three tutorial rounds, one for each Speed-Accuracy Tradeoff level. In contrast to the main study block, in these rounds, the robot successfully placed two of three blocks correctly on top of the towers, regardless of its configuration. We did this to prevent participants from preemptively assuming that the robot will always fail in risky settings, which could encourage them to always opt for manual control. During the tutorial, we addressed any remaining questions before starting the main experiment.

In the main study, participants completed the task for each of the 12 experimental conditions. Before each task, the robot performed a short "calibration movement," moving once between the pick-up and a placement location while verbally announcing: *"Calibration. Please wait ... Calibration done. You can start assembling the first base."* Participants then worked on the task as described in subsection 3.2. The wizard triggered the robot's actions based on participants' requests, with the robot verbally confirming each action before executing it. Depending on the condition, the robot either announced its uncertainty, asked participants for a decision, or acted autonomously. If participants attempted impossible actions or built a tower incorrectly, the wizard played corrective responses such as "I can't do that. Please only request valid actions." or "The base is not correct, I can't place the sample there."

Between each task block, we asked participants to complete a questionnaire. After completing all task blocks, we conducted a short semi-structured interview. Finally, we reimbursed participants, answered any open questions, and thanked them for their participation.

3.5 Measures and Analysis

We collected and analyzed three types of data: questionnaire responses, robot log data, and interview transcripts.

Questionnaires: After each condition, participants completed a set of questionnaires (see supplemental material for the full list). As a manipulation check, we asked participants to rate the robot's speed on a 7-point Likert scale. We captured answers to four self-constructed items to measure user frustration related to both user and robot performance and decisions. Beyond our main outcome frustration, we also collected a compact set of secondary measures to interpret mechanisms and trade-offs around frustration: We used the Utility and Transparency subscales of the Trust Questionnaire [30], NASA-TLX for workload [10], the HRIES agency subscale [38], and the Godspeed Perceived Intelligence subscale [3] to capture the impact on the attitude of towards the robot.

We analyzed outcomes using mixed-effects models in R. We fitted Linear Mixed Effect Models (LMEs) using `lme4` for composite scores (e.g., raw TLX), Cumulative Linked Mixed Models (CLMMs) using `ordinal` for single Likert items, and binomial Generalized Linear Mixed Models (GLMM) using `lme4` for proportions (e.g., safety actions out of three). We included participant-level random intercepts and added random slopes for experimental factors where supported by the data. If diagnostics flagged convergence or singularity, we trimmed the random-effects structure using AIC/BIC-guided comparisons using the `easystats` package. Before testing effects, we checked the respective assumptions of the models using the `performance` package. For omnibus inference, we performed Wald tests via `emmeans::joint_tests`. For significant terms, we estimate Estimated Marginal Means (EMMs) and pairwise contrasts using `emmmeans` with Bonferroni correction.

Log Data: We logged every robot action triggered by the wizard into CSV files. From this, we derived task duration and the number of times participants instructed the robot to use the safe area.

Interviews: After all conditions, we conducted semi-structured interviews (see supplemental material). We transcribed recordings with *atrain*[2] and analyzed them via thematic analysis using Atlas.ti. Two authors independently coded 20% of the interviews to create a shared codebook, then independently coded half of the remaining interviews each, meeting at the end to resolve discrepancies.

3.6 Participants

We recruited 24 participants (13 female, 11 male) via our university's mailing list. On average, participants were 26.38 years old ($SD = 4.43$, $min = 20$, $max = 39$). Participants were from four different nationalities. One participant had a doctoral degree, six had a master's degree, fourteen had a bachelor's degree, and three had a high school degree. Participants' personal level negative attitude towards robots [15] was 2.5 ($SD = 0.99$). 11 participants had never interacted with robots before, 12 participants interacted with robots between 1 and 7 times, and one participant more than 7 times. Among these, participants mainly named robotic arms and social robots. The study took approximately 60 minutes, and we reimbursed participants for their time.

4 Results

In the following, we report our quantitative results from the questionnaire data (see 2 and 1) and the qualitative results.

4.1 Manipulation Checks

Perceived Robot Speed: We verified the manipulation of the robot's speed by asking participants whether they perceived the robot's movement as fast. Perceived speed tracked the SAT, with higher ratings at *risky* than *moderate* and *conservative* (risky–moderate $M = 0.87, p = 0.0003$; risky–conservative $M = 1.61, p < .0001$; moderate–conservative $M = 0.73, p = 0.0050$). Participants in no decision

[2] https://github.com/JuergenFleiss/aTrain

Table 1: CLMM and LME results with Wald χ^2 tests for all dependent variables.

	Speed-Accuracy		Uncertainty		User Decision		SA × UC		SA × UD		UC × UD		SA × UC × UD	
	$\chi^2_{(2)}$	p	$\chi^2_{(1)}$	p	$\chi^2_{(1)}$	p	$\chi^2_{(2)}$	p	$\chi^2_{(2)}$	p	$\chi^2_{(1)}$	p	$\chi^2_{(2)}$	p
Speed Manipulation Check	18.796	<.001	0.116	.733	9.152	.003	2.868	.238	5.182	.075	1.265	.261	0.330	.848
Trust Questionnaire [30]														
Utility Score	24.971	<.001	3.540	.061	1.174	.280	0.380	.684	6.362	.002	0.263	.608	0.068	.934
Transparency Score	8.850	.000	0.279	.598	5.234	.023	4.460	.013	0.496	.610	0.774	.380	0.289	.749
Raw NASA TLX [10]	19.255	<.001	0.627	.430	0.019	.890	1.174	.311	10.158	.000	0.850	.358	0.353	.703
Self-Defined Frustration Questions														
Robot Performance	67.334	<.001	0.170	.680	0.816	.366	1.546	.462	21.314	<.001	0.017	.896	0.160	.923
Own Performance	2.818	.244	0.201	.654	0.176	.675	0.500	.779	1.784	.410	0.945	.331	1.324	.516
Robot Decisions	55.096	<.001	0.309	.578	2.641	.104	2.098	.350	32.692	<.001	0.410	.522	1.556	.459
Own Decisions	0.564	.754	0.027	.869	1.317	.251	0.438	.803	1.634	.442	0.000	.997	0.028	.987
HRIES Perceived Agency [38]	66.317	<.001	0.039	.844	5.272	.023	0.010	.990	5.416	.005	0.389	.534	2.181	.115
Godspeed Perceived Intelligence [3]	108.491	<.001	0.282	.596	21.069	<.001	0.026	.974	12.656	<.001	0.144	.705	0.864	.423

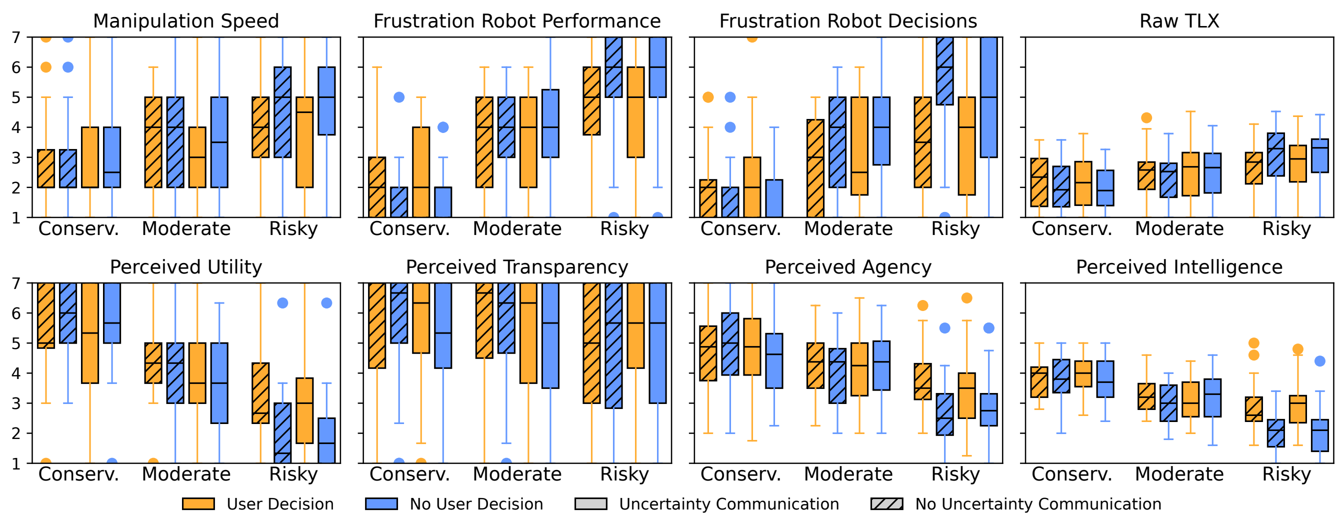

Figure 2: Boxplot of self-defined items for the manipulation check of perceived speed, frustration regarding the robots performance and decisions, user workload (raw TLX) [10] (scaled to 1-7 for the plot), perceived utility and transparency [30] (scaled to 1-7 for the plot), perceived agency [38], and perceived intelligence [3]

reported higher perceived speed than decision ($M = 0.43, p = 0.0007$). No interactions were reliable for the manipulation check.

Task Duration: As expected, both with uncertainty communication ($M = 130$ s, without $M = 121$ s) and with user decision ($M = 134$ s, without $M = 118$ s) conditions took longer on average, as this entailed additional waiting times for the participants. Considering the SAT, tasks took longer for risky ($M = 142$ s) compared to moderate ($M = 121$ s) and conservative ($M = 114$ s). We also found an interaction SAT × User Decision. At risky speed, we found a small non-significant difference with ($M = 146$ s) or without ($M = 139$ s) decision, $p = .070$. At moderate and conservative levels, decisions added a clearer time penalty ($M = 112$ s vs. $M = 131$ s, $M = 105$ s vs. $M = 125$ s, both $p < .001$).

4.2 Quantitative

Frustration (robot's performance): Frustration with the robot's performance increased with riskier SAT (risky 5.20, moderate 3.79, conservative 1.89; all $p < .001$). This effect was moderated by SAT × User Decision. Comparing decision settings within SAT levels, we found that in the risky conditions, no decision yielded higher frustration than decision ($\Delta \approx 0.98$, $p = .004$), at moderate the difference was small and not reliable ($\Delta \approx 0.53$, $p = .195$), and at conservative it slightly reversed but was not reliable ($\Delta \approx -0.58$, $p = .059$). Uncertainty Communication showed no effect. We found a slightly increasing frustration over the study (+0.082, $p = .031$).

Frustration (robot's decisions): Frustration with the robot's decisions increased with riskier SAT (risky 4.34, moderate 3.11, conservative 1.66; all $p < .001$). This effect was again moderated by a SAT × User Decision interaction. Comparing decision settings within SAT levels, we found that in the risky conditions, no decision yielded higher frustration than decision ($\Delta \approx 1.95$, $p < .001$), while at moderate and conservative, the difference between settings was small and not reliable ($p > .05$). Uncertainty Communication showed no effect and we found no other interactions.

Frustration (own performance and decisions): We found no significant effects on the items about the participants' frustration with their own performance and decisions.

Workload (Raw TLX): Workload increased with riskier SAT (risky 6.79, moderate 5.68, conservative 4.52; all $p < .005$). This effect was again moderated by SAT × User Decision and showed a similar trend. Comparing decision settings within SAT levels, we found that in the risky conditions, no decision yielded higher workload than decision ($\Delta \approx 0.88$, $p = .001$), at moderate the difference was small and not reliable ($\Delta \approx 0.53$, $p = .196$), and at conservative it reversed (no decision < decision; $\Delta \approx -0.58$, $p = .019$). Uncertainty Communication showed no effect. We observed a small decrease in perceived workload over the study (-0.045 per block, $p = .042$).

Godspeed Intelligence: Intelligence impressions improved with more conservative SAT (risky < moderate, $p < .001$; risky < conservative, $p < .001$; moderate < conservative, $p < .001$). This pattern was moderated by SAT × User Decision. Comparing decision settings within SAT levels, we found that in the risky conditions, decision yielded higher intelligence ratings than no decision (difference ≈ 0.85, $p < .001$). Moderate and conservative settings did not differ reliably (all $p \geq .330$). Uncertainty Communication again showed no effect.

HRIES agency: Agency impressions increased with more conservative SAT (risky 3.16, moderate 4.18, conservative 4.74; all $p \leq .01$). This pattern was moderated by SAT × User Decision. Comparing decision settings within SAT levels, we found that in the risky conditions, decision yielded higher agency than no decision (difference ≈ 0.79 in no-decision–decision, $p < .001$), while at moderate and conservative, settings did not significantly differ (all $p \geq .850$). Uncertainty Communication showed no effect.

Transparency: Transparency ratings increased with more conservative SAT (risky < moderate, $p = .028$; risky < conservative, $p < .001$). This effect was moderated by SAT × Uncertainty Communication. Comparing the two Uncertainty Communication levels within SAT, we found that in the risky conditions, uncertainty cue > no uncertainty cue ($\Delta \approx 0.146$, $p = .034$), while at moderate and conservative the difference was not reliable (both $p \geq .129$). Looking across Uncertainty Communication levels, the SAT gradient was pronounced under no uncertainty cue (risky < moderate, $p = .001$; risky < conservative, $p < .001$) and largely attenuated under uncertainty cue (all $p \geq .790$).

Utility: Utility increased with more conservative SAT (risky 2.34, moderate 2.49, conservative 2.69; risky < moderate, $p = .005$; risky < conservative, $p < .001$; moderate < conservative, $p < .001$). This effect was moderated by SAT × User Decision. Comparing decision settings across SAT levels, we found that in the risky conditions, decision yielded higher utility than no decision ($\Delta \approx 0.215$, $p = .002$), at moderate settings did not differ ($p = .835$), and at conservative no decision yielded slightly higher utility than decision ($\Delta \approx 0.139$, $p = .049$). Uncertainty Communication showed no effect. We found no higher-order interactions.

Safety-action choices in User Decision Conditions: We analyzed the number of safety-action choices in the User Decision conditions. Self-performed actions were more frequent at higher SAT levels (risky 0.447, moderate 0.287, conservative 0.133). Pairwise post-hoc odds ratios confirmed risky > moderate (OR ≈ 2.01, $p = .041$), risky > conservative (OR ≈ 5.28, $p < .001$), and moderate > conservative (OR ≈ 2.63, $p = .008$). Choices became more self-performed over time ($p = .005$). Uncertainty Communication showed no effect; we found no interaction effects.

Moderating effect of safety-action choices on frustration: We analyzed whether the share of safety actions that participants performed themselves moderates frustration about the robot's performance and decisions. For *frustration with the robot's performance*, the self-performed share showed a negative slope at risky (trend $= -3.690$, $p < .001$), was near zero at moderate (trend $= -0.320$, $p = .676$), and was positive at conservative (trend $= +2.460$, $p = .003$). For *frustration with the robot's decisions*, we found the same pattern (risky -2.962, $p = .006$; moderate $+0.115$, $p = .878$; conservative $+2.409$, $p = .004$).

Further, we compared no user decision conditions to decision conditions with no safety actions by the participant. In all of these conditions, the robot executed the task entirely. In no decision conditions, participants were not offered a choice, whereas in user decision conditions, they were offered a choice but elected not to perform any safety actions. We found that *frustration about the robot's performance* did not differ reliably (risky: $\Delta = 0.137$, $p = .711$; moderate: $\Delta = -0.527$, $p = .158$; conservative: $\Delta = 0.162$, $p = .495$). For *frustration about the robot's decisions*, differences were likewise negligible at risky and conservative (risky: $\Delta = -0.653$, $p = .203$; conservative: $\Delta = 0.063$, $p = .769$), but at moderate speed frustration was lower under decision than no decision ($\Delta = -0.962$, $p = .015$).

4.3 Qualitative

We identified four themes. *Uncertainty Communication* and *User Decision* capture participants' thoughts on these two independent variables, *Config Preference & Perceptions* outline user preferences on how the robot should be configured and how these configurations affected their perception, and *Frustration* enhances our understanding of the quantitative findings related to frustration.

Uncertainty Communication: Participants had different opinions and strategies on how they dealt with the robot communicating uncertainty, reflected in six categories. Half of the participants found uncertainty communication to be helpful, as it prepared them for what's going to happen (P3, P13). Some also felt the robot appeared more intelligent when aware of its own uncertainty. Participants reacted differently to uncertainty. Some became more cautious and asked the robot to place blocks in the safe place whenever possible, e.g., P17: *"When it says that it is not sure, then I will ask it to place it in a safe place."* Others used these situations to test the robot, which sometimes fostered curiosity rather than avoidance. The effect on trust was mixed: for some, uncertainty did not erode trust (P13, P14, P19), while for others, it did (P9, P15). However, the majority of participants disliked the mismatch between communicated uncertainty and actual performance, emphasizing that the stated uncertainty should reflect observed uncertainty. Such mismatches could even increase frustration, as P19 stated *"I got annoyed when it could actually perform the task [after stating that it is uncertain]."*

User Decision: We identified three categories related to participants' involvement in decisions. In general, the robot's low accuracy induced caution and made participants take over more often when given the choice. However, many first observed how the robot performed on the initial tower before deciding whether to intervene for subsequent towers. As P22 stated: *"I always tried to see if in the first [tower] it would [place the block accurately], then I decided according to that result for the second and third [tower]."*

Config Preference & Perceptions: Participants expressed preferences regarding speed, accuracy, and their own involvement in decision-making (11 categories). Many appreciated being given agency in the User Decision condition, describing it as giving them a sense of safety and control. P4 stated: *"It gives more control and it also gives you time to correct anything or just completely stop it if something goes wrong, so I think it's the best."* Conversely, a smaller group preferred the robot to take over fully without requiring their input, e.g., P12: *"Don't ask my instruction, just put it on the base."* This perspective was often linked to a desire for efficiency or minimal effort. Most participants preferred the moderate speed configuration, with only some preferring the fast speed. Others found the conservative speed as too slow. However, participants consistently prioritized accuracy over speed. P5 summarized: *"I don't really care about the speed as long as it's precise."* Similarly, P10 explained: *"If they are accurate enough, then the faster process is preferred. But if they are making mistakes when they are fast, then I would prefer the slow ones."* Some participants described the conservative, slower robot as feeling safe and reliable (P6, P8, P14). Participants noted that uncertainty communication and decision-making worked best in combination, allowing them to adjust to the robot's limitations while avoiding errors. P23 said, *"It is nice that if it knows it is not going to be able to do the job, then it makes me [make] a decision."* Similarly, P32 emphasized that having this alternative provided flexibility: *"When it can't do something on its own, then it [should] ask me"*

Interestingly, some participants anthropomorphized the robot, attributing distinct personalities to the different configurations, calling the risky configuration "foolish" (P2), "sassy" (P19), or "behaving like a child" (P18). P19 stated *Honestly, like in the beginning, I thought, oh, yeah, this is like a really cool tool. And then I thought, oh, you are sassy. It has personality for sure."* The conservative configuration was seen as more careful and responsible (P6, P8, P14, P15, P25). P6 reflected: *"When it was relatively slow and the accuracy was higher, it gave me the impression that the robot is more responsible and more useful."* Furthermore, some participants explicitly stated that the conservative configuration reduced cognitive load, while the risky one increased it.

Frustration: We captured participants' thoughts on frustration across six categories. Frustration peaked when the robot damaged completed structures, as this required rebuilding and increased workload. P18 stated, *"I think it's really frustrating when it destroys your work. It's not that frustrating when it just drops it beside the thing, like it's okay. But I feel like the most annoying part is definitely when it destroys stuff you've already built."* Frustration was also linked to unpredictability and mismatched expectations. P7 noted, *"It was a bit stressful for me to not know if the robot's work would be successful or not."*Participants were especially frustrated when the robot asked for input but then failed, as P11 explained: *"I didn't like that he asked me, and then when I said you could put it on the base, expecting that he was going to do it in the right way, he didn't do it in the right way."* Frequent and repetitive verbal interactions were tiring. Some participants disliked both the robot's repeated utterances and having to repeatedly speak commands, suggesting simpler interaction methods like buttons. Conversely, some found that uncertainty communication and user control reduced frustration. P9 felt *"less frustration because he told me he might mess it up."* Similarly, P13 said, *"When you had the decision to place it on the safe place, it wasn't frustrating anymore, because I knew how it was going to happen."*

5 Discussion

In the following, we discuss our findings regarding causes for frustration, the cost of involving users, the role of actional transparency, and lastly, suggest mitigation strategies.

5.1 Higher Frustration with Risky Robots

Our quantitative results show that user frustration regarding the robot's performance and decisions is highest for the risky configurations, with significant differences for both the conservative and moderate settings. Furthermore, in our qualitative results, we found that accuracy was perceived as more important than speed. The magnitude of the error affected the results significantly. With this, we found strong support for H1. Users' decisions moderated these effects, as both our quantitative and qualitative results show that under risky configurations, providing users with the option to take over decreased frustration, while this decision-related overhead increased frustration under conservative robot configurations, supporting H3. Further analysis of frustration, depending on how often participants actually took over for the robot, supported this trend strongly. The more often the users took over in risky configurations, the lower the frustration, with the reverse trend for the conservative configuration, and no effect for the moderate configuration. However, we could not find significant effects of uncertainty communication on user frustration for the different SAT configurations; thus, we reject H2. For our supporting measurements (perceived agency, perceived intelligence, workload, utility), we confirmed the same trends across all constructs: Perception was worst for risky configurations, and best for conservative configurations, moderated by users' decisions. This indicates that user perception of robots is linked with frustration, and users perceive conservative robots as more intelligent, attribute higher agency, and utility. Interestingly, perceived utility, as an opposing construct to frustration, was highest with conservative behavior, indicating that conservative behavior not only decreases frustration but even fosters perceived usefulness.

In summary, we found that riskier robot configurations lead to higher user frustration, thus suggesting that accurate robot behavior should be prioritized over speed.

5.2 If It's Safe, Don't Ask

Involving users in the robot's decision-making comes with an inherent cost [8, 11, 28]. Our findings highlight that user decision adds overhead that can increase frustration when errors are unlikely, but reduce frustration when errors are costly. Both the communication of uncertainty and user involvement increase task duration even in the destructive settings. Users took longer due to the negotiation with the robot than fixing its errors. However, fixing errors is not always cheap and feasible. Indeed, our results show that user involvement can decrease frustration in risky configurations and also increase perceived utility, indicating that when robot uncertainty or the cost of error is high, involving users in the process is beneficial. Our qualitative results support this as users *liked* having the option to make a decision, aligning with prior findings that humans value having a sense of control over automated systems [16]. When robots are actually uncertain, they can ask users without the downside of frustration or sacrificing utility.

However, our results also indicate that while decreasing frustration in risky configurations, user involvement in conservative settings increases frustration. When the robot is performing reliably, involving the user introduces unnecessary friction without improving team performance. These findings align with the theory of *adjustable autonomy*, which proposes that control should transfer between human and robot only when the expected value of human input exceeds the associated cost of delay or attention [33]. Prior work has shown that humans are willing to accept robot autonomy when it benefits the team [6, 34] and that offloading routine decisions to robots reduces workload [7]. Our qualitative data echo this: participants appreciated the option to intervene when accuracy was low, but described repeated prompts at higher accuracy levels as tedious and disruptive.

In summary, robots should not simply default to asking the user when uncertainty arises, and instead only involve the user when necessary. For this, the robots need to be aware of their own uncertainty [20, 37]. Our qualitative results suggest that a combination of uncertainty communication with user involvement leads to the best user understanding and perception of perceived collaboration.

5.3 Non-Actionable Transparency Does Not Help

In our study, verbal uncertainty communication did not reduce frustration, workload, perceived intelligence, perceived agency, and utility across SAT configurations. Prior work suggests that transparency is most effective when it is *actionable*, that is, when users can respond to the information provided [23, 41, 42]. In contrast, non-actionable transparency risks adding cognitive load without benefits [16]. In our case, uncertainty cues were most useful when paired with user decisions, suggesting that the option to act matters more than the cue itself.

We did observe an interaction effect between SAT and uncertainty communication. In risky conditions, uncertainty cues significantly increased perceived transparency and utility, likely because users could connect the statement to visible errors. In conservative conditions, however, the robot always succeeded despite claiming uncertainty. Participants noticed this mismatch and reported disliking it. Such miscalibration may explain why transparency alone had little effect on frustration. Aligning uncertainty communication with actual performance appears crucial.

We based our uncertainty communication on prior findings [14], which showed that first-person expressions of uncertainty improved calibration in AI chatbots. While some participants reported similar effects,many instead used the cues to prove the robot's behavior. We hypothesize that this difference may stem from the different embodiment levels of the studies; In text-based settings, verbal cues align naturally with interaction, whereas in physical HRI they may not. Two possible explanations arise: (1) verbal uncertainty cues have little impact on frustration and perception in collaborative physical HRI, or (2) the modality of uncertainty cues must match the robot's embodiment. Prior work supports the latter, showing that embodied signals like hesitant movement more effectively convey uncertainty [36].

In our study, verbal uncertainty communication did not decrease user frustration. Future work should further investigate this and match the modality of uncertainty cues to the task at hand.

5.4 Frustration Mitigation Strategies

In general, we found that conservative robot behavior leads to the least user frustration; thus, we recommend that whenever possible, conservative behavior with high accuracy should be preferred. However, this is not always possible, as in some scenarios, the robot's inherent uncertainty [20, 37] will be too high to act conservatively, or a higher speed is needed. If risky behavior is necessary, we found two strategies to mitigate user frustration: (1) Leveraging uncertainty cues to increase transparency, and (2) involving users in the decision-making process to increase utility; with both transparency and utility contributing to perceived trustworthiness. [30].

5.5 Limitations & Future Work

We manipulated the SAT as a combined factor rather than varying speed and accuracy independently to keep the number of conditions manageable, but this limits our ability to determine whether frustration was driven more by speed or by inaccuracy. However, our qualitative findings suggest that accuracy played the larger role in shaping participants' experiences. We intentionally programmed the robot to be deterministic; thus, always inaccurate in the moderate and risky conditions. While this does not fully reflect probabilistic real-world robotic performance, it allowed us to systematically control for the frequency of failures, especially in the user decision condition, where participants' choices could otherwise have led to highly variable task outcomes. We observed a general increase in frustration ratings over time. Although we counterbalanced condition order to reduce these effects, this trend suggests that cumulative factors such as fatigue or learning may have influenced participants' perceptions. Participants perceived our generic uncertainty cues as repetitive. Future research could investigate more nuanced uncertainty communication per trial reflecting AI confidence, or holistic a priori uncertainty communication reflecting AI accuracy as a performance indicator [27]. Our task simulated a high-risk collaborative scenario in a controlled laboratory setting. Future work should examine whether these findings generalize to different types of tasks and varying levels of real-world stakes, where the cost of errors and the value of human intervention may differ significantly.

6 Conclusion

We investigated how robot performance, uncertainty communication, and user decision-making shape frustration and user experience in collaborative HRI. In a user study ($N = 24$), we found that risky, inaccurate robot behavior was the main driver of frustration. Giving users control reduced frustration, but only when failures were likely; in low-risk settings, it slowed progress and increased frustration. Uncertainty communication alone had little effect unless paired with actionable choices. We propose that robots should selectively involve users, deferring to them when uncertain, while acting autonomously in routine tasks. This provides guidance for researchers to reduce frustration-related confounds in studies and for designers to build efficient, trustworthy collaborative robots.

References

[1] Moaed A Abd, Iker Gonzalez, Mehrdad Nojoumian, and Erik D Engeberg. 2017. Trust, satisfaction and frustration measurements during human-robot interaction. In *Proceedings of the 30th Florida Conference on Recent Advances in RoboticsMay*. 11–12.

[2] Ainhoa Apraiz Iriarte, Ganix Lasa, Maitane Mazmela, Nestor Arana-Arexolaleiba, Antonio Serrano Muñoz, Íñigo Elguea-Aguinaco, and Amaia Etxabe. 2025. Evaluating the Effect of Speed and Acceleration on Human Factors during an Assembly Task in Human–Robot Interaction (HRI). *International Journal of Social Robotics* 17 (02 2025), 211–256. doi:10.1007/s12369-025-01224-7

[3] 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 (Jan. 2009), 71–81. doi:10.1007/s12369-008-0001-3

[4] Jenay M Beer, Arthur D Fisk, and Wendy A Rogers. 2014. Toward a Framework for Levels of Robot Autonomy in Human-Robot Interaction. *Journal of Human-Robot Interaction* 3, 2 (June 2014), 74. doi:10.5898/JHRI.3.2.Beer

[5] Umang Bhatt, Javier Antorán, Yunfeng Zhang, Q. Vera Liao, Prasanna Sattigeri, Riccardo Fogliato, Gabrielle Melançon, Ranganath Krishnan, Jason Stanley, Omesh Tickoo, Lama Nachman, Rumi Chunara, Madhulika Srikumar, Adrian Weller, and Alice Xiang. 2021. Uncertainty as a Form of Transparency: Measuring, Communicating, and Using Uncertainty. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Virtual Event USA, 401–413. doi:10.1145/3461702.3462571

[6] Filippo Cantucci, Rino Falcone, and Cristiano Castelfranchi. 2022. Human-Robot Interaction through Adjustable Social Autonomy. *Intelligenza Artificiale* 16, 1 (July 2022), 69–79. doi:10.3233/IA-210124

[7] J.W. Crandall and M.A. Goodrich. 2001. Experiments in adjustable autonomy. In *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat.No.01CH37236)*, Vol. 3. 1624–1629 vol.3. doi:10.1109/ICSMC.2001.973517

[8] J.W. Crandall and M.A. Goodrich. 2002. Characterizing efficiency of human robot interaction: a case study of shared-control teleoperation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vol. 2. 1290–1295 vol.2. doi:10.1109/IRDS.2002.1043932

[9] Manuel Giuliani, Nicole Mirnig, Gerald Stollnberger, Susanne Stadler, Roland Buchner, and Manfred Tscheligi. 2015. Systematic Analysis of Video Data from Different Human–Robot Interaction Studies: A Categorization of Social Signals during Error Situations. 6 (2015). doi:10.3389/fpsyg.2015.00931

[10] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. arXiv:https://doi.org/10.1177/154193120605000909 doi:10.1177/154193120605000909

[11] Guy Hoffman. 2019. Evaluating Fluency in Human–Robot Collaboration. *IEEE Transactions on Human-Machine Systems* 49, 3 (2019), 209–218. doi:10.1109/THMS.2019.2904558

[12] Shanee Honig and Tal Oron-Gilad. 2018. Understanding and Resolving Failures in Human-Robot Interaction: Literature Review and Model Development. *Frontiers in Psychology* 9 (June 2018), 861. doi:10.3389/fpsyg.2018.00861

[13] Julian Hough and David Schlangen. 2017. It's Not What You Do, It's How You Do It: Grounding Uncertainty for a Simple Robot. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Vienna Austria, 274–282. doi:10.1145/2909824.3020214

[14] Sunnie S. Y. Kim, Q. Vera Liao, Mihaela Vorvoreanu, Stephanie Ballard, and Jennifer Wortman Vaughan. 2024. "I'm Not Sure, But...": Examining the Impact of Large Language Models' Uncertainty Expression on User Reliance and Trust. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 822–835. doi:10.1145/3630106.3658941

[15] Mika Koverola, Anton Kunnari, Jukka Sundvall, and Michael Laakasuo. 2022. General Attitudes Towards Robots Scale (GAToRS): A New Instrument for Social Surveys. *International Journal of Social Robotics* 14, 7 (Sept. 2022), 1559–1581. doi:10.1007/s12369-022-00880-3

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

[17] 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*. ACM, Boulder CO USA, 413–422. doi:10.1145/3610977.3634963

[18] Stephan J Lemmer, Anhong Guo, and Jason J Corso. 2023. Human-Centered Deferred Inference: Measuring User Interactions and Setting Deferral Criteria for Human-AI Teams. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 681–694. doi:10.1145/3581641.3584092

[19] Anders Lenskjold, Janus Uhd Nybing, Charlotte Trampedach, Astrid Galsgaard, Mathias Willadsen Brejnebøl, Henriette Raaschou, Martin Høyer Rose, and Mikael Boesen. 2023. Should Artificial Intelligence Have Lower Acceptable Error Rates than Humans? 5, 1 (2023), 20220053. doi:10.1259/bjro.20220053

[20] Jan Leusmann, Chao Wang, Michael Gienger, Albrecht Schmidt, and Sven Mayer. 2023. Understanding the Uncertainty Loop of Human-Robot Interaction. doi:10.48550/arXiv.2303.07889 arXiv:2303.07889 [cs].

[21] Leila Methnani, Andrea Aler Tubella, Virginia Dignum, and Andreas Theodorou. 2021. Let Me Take Over: Variable Autonomy for Meaningful Human Control. *Frontiers in Artificial Intelligence* 4 (Sept. 2021), 737072. doi:10.3389/frai.2021.737072

[22] 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), 21. doi:10.3389/frobt.2017.00021

[23] Birthe Nesset, David A. Robb, José Lopes, and Helen Hastie. 2021. Transparency in HRI: Trust and Decision Making in the Face of Robot Errors. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder CO USA, 2021-03-08). ACM, 313–317. doi:10.1145/3434074.3447183

[24] Carl Oechsner, Jan Leusmann, Robin Welsch, Andreas Martin Butz, and Sven Mayer. 2025. Influence of Perceived Danger on Proxemics in Human-Robot Object Handovers. In *Proceedings of the Mensch Und Computer 2025 (MuC '25)*. Association for Computing Machinery, New York, NY, USA, 111–120. doi:10.1145/3743049.3743064

[25] Claudia Pérez-D'Arpino, Rebecca P. Khurshid, and Julie A. Shah. 2024. Experimental Assessment of Human–Robot Teaming for Multi-step Remote Manipulation with Expert Operators. *J. Hum.-Robot Interact.* 13, 3, Article 36 (Aug. 2024), 26 pages. doi:10.1145/3618258

[26] Snehal Prabhudesai, Leyao Yang, Sumit Asthana, Xun Huan, Q. Vera Liao, and Nikola Banovic. 2023. Understanding Uncertainty: How Lay Decision-makers Perceive and Interpret Uncertainty in Human-AI Decision Making. In *Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23)*. Association for Computing Machinery, New York, NY, USA, 379–396. doi:10.1145/3581641.3584033

[27] Amy Rechkemmer and Ming Yin. 2022. When Confidence Meets Accuracy: Exploring the Effects of Multiple Performance Indicators on Trust in Machine Learning Models. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–14. doi:10.1145/3491102.3501967

[28] Merle M. Reimann, Florian A. Kunneman, Catharine Oertel, and Koen V. Hindriks. 2024. A Survey on Dialogue Management in Human-robot Interaction. *J. Hum.-Robot Interact.* 13, 2, Article 22 (June 2024), 22 pages. doi:10.1145/3648605

[29] Allen Z. Ren, Anushri Dixit, Alexandra Bodrova, Sumeet Singh, Stephen Tu, Noah Brown, Peng Xu, Leila Takayama, Fei Xia, Jake Varley, Zhenjia Xu, Dorsa Sadigh, Andy Zeng, and Anirudha Majumdar. 2023. Robots That Ask For Help: Uncertainty Alignment for Large Language Model Planners. arXiv:2307.01928 [cs.RO] https://arxiv.org/abs/2307.01928

[30] Eileen Roesler, Tobias Rieger, and Dietrich Manzey. 2022. Trust towards Human vs. Automated Agents: Using a Multidimensional Trust Questionnaire to Assess The Role of Performance, Utility, Purpose, and Transparency. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 66 (10 2022), 2047–2051. doi:10.1177/1071181322661065

[31] Alessandra Rossi, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L. Walters. 2017. How the Timing and Magnitude of Robot Errors Influence Peoples' Trust of Robots in an Emergency Scenario. In *Social Robotics*, Abderrahmane Kheddar, Eiichi Yoshida, Shuzhi Sam Ge, Kenji Suzuki, John-John Cabibihan, Friederike Eyssel, and Hongsheng He (Eds.). Springer International Publishing, Cham, 42–52.

[32] Alexander Rossner, Marie Ambach, and Sven Pagel. 2025. Navigating Misinformation: Understanding User Frustration in AI-Driven Chatbot Interactions. In *Artificial Intelligence in HCI*, Helmut Degen and Stavroula Ntoa (Eds.). Vol. 15820. Springer Nature Switzerland, 263–281. doi:10.1007/978-3-031-93415-5_16

[33] Paul Scerri, David Pynadath, and Milind Tambe. 2001. Adjustable autonomy in real-world multi-agent environments. In *Proceedings of the Fifth International Conference on Autonomous Agents* (Montreal, Quebec, Canada) *(AGENTS '01)*. Association for Computing Machinery, New York, NY, USA, 300–307. doi:10.1145/375735.376314

[34] Paul Schermerhorn and Matthias Scheutz. 2009. Dynamic robot autonomy: investigating the effects of robot decision-making in a human-robot team task. In *Proceedings of the 2009 International Conference on Multimodal Interfaces* (Cambridge, Massachusetts, USA) *(ICMI-MLMI '09)*. Association for Computing Machinery, New York, NY, USA, 63–70. doi:10.1145/1647314.1647328

[35] Sarah Schömbs, Jorge Goncalves, and Wafa Johal. 2025. "I can feel the risks by looking at the robot face": Communicating Risk through a Physical Agent. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. Association for Computing Machinery, New York, NY, USA, 236–252. doi:10.1145/3715336.3735759

[36] Sarah Schömbs, Saumya Pareek, Jorge Goncalves, and Wafa Johal. 2024. Robot-Assisted Decision-Making: Unveiling the Role of Uncertainty Visualisation and Embodiment. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3613904.3642911

[37] Abbas Sidaoui, Naseem Daher, and Daniel Asmar. 2022. Human-Robot Interaction via a Joint-Initiative Supervised Autonomy (JISA) Framework. *Journal of Intelligent & Robotic Systems* 104, 3 (March 2022), 51. doi:10.1007/s10846-022-01592-6

[38] Nicolas Spatola, Barbara Kühnlenz, and Gordon Cheng. 2021. Perception and Evaluation in Human–Robot Interaction: The Human–Robot Interaction Evaluation Scale (HRIES)—A Multicomponent Approach of Anthropomorphism. *International Journal of Social Robotics* 13, 7 (Nov. 2021), 1517–1539. doi:10.1007/s12369-020-00667-4
[39] Jonathan Vitale, Meg Tonkin, Sarita Herse, Suman Ojha, Jesse Clark, Mary-Anne Williams, Xun Wang, and William Judge. 2018. Be More Transparent and Users Will like You: A Robot Privacy and User Experience Design Experiment. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago, IL, USA, 2018) *(Hri '18)*. Association for Computing Machinery, New York, NY, USA, 379–387. doi:10.1145/3171221.3171269
[40] Alexandra Weidemann and Nele Rußwinkel. 2021. The Role of Frustration in Human–Robot Interaction – What Is Needed for a Successful Collaboration? 12 (2021), 640186. doi:10.3389/fpsyg.2021.640186
[41] Robert H. Wortham, Andreas Theodorou, and Joanna J. Bryson. 2017. Robot Transparency: Improving Understanding of Intelligent Behaviour for Designers and Users. In *Towards Autonomous Robotic Systems*, Yang Gao, Saber Fallah, Yaochu Jin, and Constantina Lekakou (Eds.). Vol. 10454. Springer International Publishing, Cham, 274–289. doi:10.1007/978-3-319-64107-2_22
[42] Julia L. Wright, Jessie Y. C. Chen, and Shan G. Lakhmani. 2020. Agent Transparency and Reliability in Human–Robot Interaction: The Influence on User Confidence and Perceived Reliability. 50, 3 (2020), 254–263. doi:10.1109/THMS.2019.2925717
[43] Euijung Yang and Michael C. Dorneich. 2018. Evaluating Human–Automation Etiquette Strategies to Mitigate User Frustration and Improve Learning in Affect-Aware Tutoring. 8, 6 (2018), 895. doi:10.3390/app8060895
[44] Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. 2019. Understanding the Effect of Accuracy on Trust in Machine Learning Models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–12. doi:10.1145/3290605.3300509
[45] Qian Zhang and Lora Cavuoto. 2025. Identifying the impact of robot speed and task time on human-robot collaboration through facial feature analysis. *International Journal of Industrial Ergonomics* 105 (2025), 103691. doi:10.1016/j.ergon.2024.103691

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