Learning Through Confusion in Human-Robot Interaction: A Pilot Study with Furhat Robot

Andreas Naoum
anaoum@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

Samin Chowdhury
saminc@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

Kevin Noventa
noventa@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

Hiba Qutbuddin Habib
hibaqh@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

Abstract

Robots are increasingly being deployed in education domains, where facilitating learning and maintaining user engagement are essential. A pivotal factor in this dynamic is the ability of robots to navigate and manage user confusion during interactions. This mental state can have dual effects, it can motivate critical thinking, however, if left unmanaged, it can lead to frustration, and disengagement. This study examined how a robot can intentionally create confusion by providing insufficient information, and investigates how confusion affects learning outcomes and user engagement in HRI. The findings aim to support the development of adaptive robots that enhance learning and engagement by effectively managing confusion.

Keywords

Human-Robot Interaction, Robot Tutoring, Learning, Confusion, Productive, Unproductive

1 Introduction

In the field of Human-Robot Interaction (HRI), understanding how robots can effectively engage users while supporting their learning efforts is critical. Robots are increasingly being deployed in education domains, where facilitating learning and maintaining user engagement are essential. A pivotal factor in this dynamic is the ability of robots to navigate and manage user confusion during interactions.

Confusion is a natural and often inevitable part of the learning process. It arises when users encounter information or tasks that challenge their current understanding or skills. This mental state can have dual effects. On the one hand, productive confusion, when confusion is properly resolved, can motivate critical thinking, exploration of alternative solutions, and deeper cognitive engagement. On the other hand, unproductive confusion, if confusion left unmanaged, can lead to frustration, errors, and disengagement, ultimately hindering both learning and user satisfaction. The fine line between these two outcomes highlights the complexity of managing confusion effectively in learning-focused interactions.

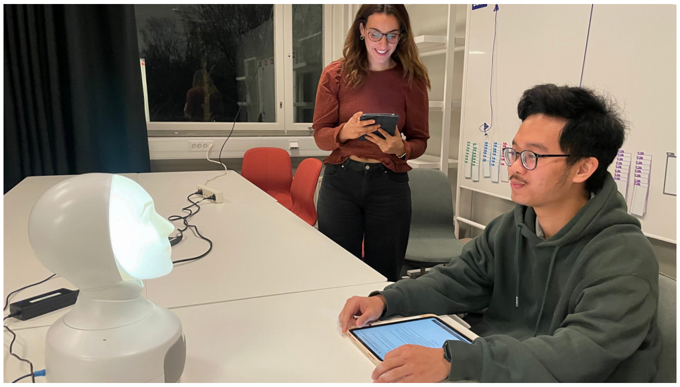

Figure 1: Overview of the experiment setup. The participant sits in a quiet room in front of the Furhat robot, while the observer assesses the interaction nearby. A tablet displaying the multiple-choice questions is placed in front of the participant for reference only (no other interaction). A full video of the experiment is available at [10].

1.1 Contribution

This study investigates how confusion affects learning outcomes and user engagement in HRI. It examines how a robot can intentionally create confusion by providing insufficient information. The research focuses on three scenarios: no confusion, productive confusion, and unproductive confusion, to understand their effects on learning outcomes and user engagement. The findings aim to support the development of adaptive robots that enhance learning and engagement by effectively managing confusion. To conclude, our contributions are as follows:

- We develop and validate strategies for robots to intentionally introduce confusion through insufficient information
- We investigate the effects of different confusion states on learning outcomes and engagement in HRI
- We demonstrate through our experiment that properly managed confusion can enhance learning outcomes and user engagement

, ,
© 2025
2025. ACM ISBN

2 Related Work

Social robots are increasingly being deployed in educational settings, serving as tutors or peer learners, with studies demonstrating their effectiveness in enhancing both cognitive outcomes and affective outcomes [3]. These robots employ a range of strategies, including personalized support, attention-directing behaviors to maintain engagement, and empathetic responses to student emotions [3]. While these strategies have shown promise, recent research suggests that leveraging cognitive states like confusion in learning could offer new opportunities for deeper learning and engagement [4]. Specifically, the deliberate use of confusion as a pedagogical tool represents an unexplored frontier in educational robotics.

One way to integrate confusion into learning is through confusion induction, a phenomenon in which specific scenarios trigger a state of uncertainty. Prior work [6, 7] identifies four primary triggers for confusion: complex information, contradictory information, insufficient information, and feedback inconsistencies. Among these, insufficient information is particularly relevant to our study, as it occurs when information is missing.

2.1 Confusion in Learning

Confusion is an epistemic emotion that emerges when new information conflicts with existing knowledge during cognitive processing. Often referred to as cognitive disequilibrium, this state can lead to goal disruption, inconsistencies, and cognitive dissonance. Despite its challenges, it is considered valuable for learning, as it encourages critical thinking, reassessment of knowledge, and persistence.

Research by D'Mello et al. demonstrated that confusion, when managed effectively, can foster deeper cognitive engagement and lead to better learning outcomes [4, 5]. Similarly, Arguel and Lane investigated the role of confusion in online learning, showing how deliberate induction and resolution of confusion helped learners grapple with complex concepts [2]. These findings highlight that confusion is not inherently detrimental; its impact depends on whether it is resolved productively or allowed to become unproductive.

In large-scale online learning settings, Yang, Kraut, and Rose explored how confusion manifests itself in Massive Open Online Courses, uncovering its influence on student performance and engagement [11]. This work emphasized the importance of addressing confusion effectively to enhance user experience in digital learning environments. However, much of this research has focused on human-computer interactions, leaving the role of confusion in HRI less thoroughly examined.

2.2 Confusion in HRI

Li et al. explored how confusion can be identified and managed in human-avatar dialogues, focusing on detecting different levels of confusion [6]. Building on this, Li et al. extended this research to HRI environments, investigating methods to invoke and detect task-oriented confusion in robot-mediated interactions, distinguishing between productive and unproductive states [7]. While these studies provide a foundation for understanding confusion in interactive systems, they do not specifically address how confusion affects learning outcomes and engagement in physical HRI contexts.

Table 1: Experimental Conditions for the Experiment

Confusion Type	Level of Information	Cognitive Demand
Unproductive Confusion (UC)	Insufficient information that obstructs the learning process.	High
Productive Confusion (PC)	Partial insufficient information that allows reasoning and hypothesis generation.	Medium
No Confusion (NC)	Fully sufficient information to answer the question.	Low

2.3 Gaps in Previous Work

Despite growing interest in confusion as a learning mechanism, its role in robot tutoring has yet to be fully explored. Most existing studies on confusion induction have been conducted in digital learning environments, leaving a gap in understanding how confusion can be strategically used in HRI-based learning. This study aims to address these gaps by investigating how social robots can induce and resolve confusion in a way that enhances learning and engagement.

3 Methodology

3.1 Research Questions

This study aims to explore how different states of confusion affect learning outcomes and user engagement in HRI. Specifically, we examine three conditions: productive confusion, where confusion is induced and subsequently resolved; unproductive confusion, where confusion persists without resolution; and no confusion, where no confusion is introduced. The key focus is on understanding how the robot's communication strategies, which may intentionally induce these states of confusion, influence the user's experience and learning.

Research Question *How do different states of confusion impact learning and user engagement in HRI?*

3.2 Hypotheses

This study hypothesizes that productive confusion will lead to higher engagement and improved learning outcomes compared to conditions where no confusion or unproductive confusion is induced. In contrast, unproductive confusion is expected to result in lower engagement, frustration, and poorer learning outcomes. Additionally, while participants experiencing no confusion may achieve good learning outcomes, their level of engagement is anticipated to be lower than that of those exposed to productive confusion.

3.3 Experimental Design

The experimental design of this study draws inspiration from previous research on eliciting confusion in HRI. In those studies [6, 7], task-oriented dialogue experiments were conducted across various task types, including logical problems, math questions, and word problems. A controlled methodology was employed in some instances to induce confusion, while others maintained clarity.

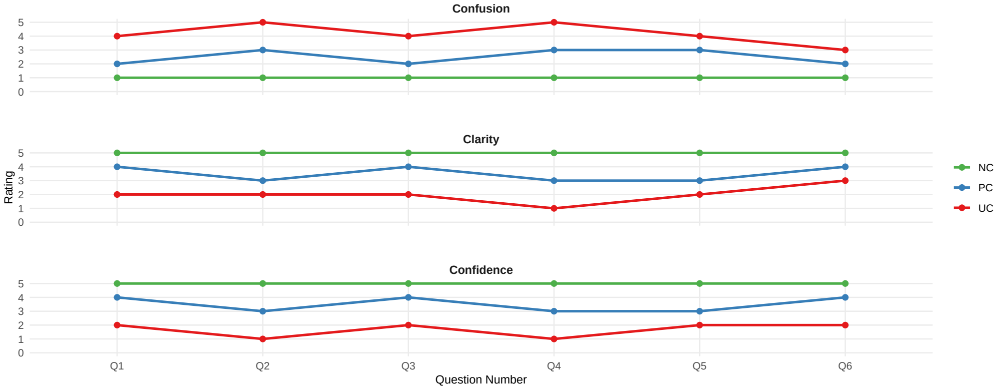

Figure 2: Automated Assessment of Hints. The figure shows the confusion, clarity, and confidence levels as rated by Claude 3.5 Sonnet. It includes assessments for the second round (Q1–Q3) and the third round (Q4–Q6).

Following the established patterns, we developed our experimental study to explore the intentional elicitation and management of confusion as a pedagogical strategy. The experiment employed a between-subjects design where each participant was randomly assigned to one of three confusion conditions (UC, PC, or NC). Each session lasted approximately 10 minutes and was conducted in a quiet room environment, as shown in Figure 1. We utilized three confusion classes, aligned with the previously defined levels of non-confusion, productive confusion, and unproductive confusion. The level of information and cognitive demand are presented in Table 1. For clarity in notation, the confusion classes are labeled as follows: UC refers to unproductive confusion, PC denotes productive confusion, and NC corresponds to stimuli that are fundamentally non-confusing.

To appeal to a general audience and ensure the study's relevance, we conceptualized the robot as a tutor delivering lessons on basic Artificial Intelligence (AI) concepts. The topics included general information about AI, supervised versus unsupervised learning, and reinforcement learning, providing a foundation for understanding key ideas in AI. The study employed the Furhat robot as the platform for investigating our research question. The robot's behavior was standardized across conditions through predetermined scripts and automated response patterns. The robot was in full autonomous mode, with its modalities utilized for implementing a full sense-plan-act process. Specifically, the robot's expressive face was used for showing excitement in the beginning of the experiment, validating correct answers, while also showing disappointment when answers were wrong.

To facilitate learning, we developed introductory information and designed multiple-choice questions that required participants to engage with the material actively. The questions were structured to enable participants to request additional hints if needed, rather than having the robot automatically provide them. These hints were tailored to each of the three experimental conditions, ensuring that the level of support matched the intended confusion condition.

The experiment consisted of three sequential rounds: (1) General information about AI, (2) supervised vs unsupervised learning, and (3) reinforcement learning. Each round began with an introduction to the topic, followed by three multiple-choice questions. For each question, participants had only one opportunity to answer. After each response, the robot provided immediate feedback by indicating whether the answer was correct or incorrect. If participants felt unsure about a question, they could request a hint from the robot, which was varied based on the experimental condition.

3.4 Development of Questions, Hints, and Manipulation Check

Our study implemented a systematic approach to develop and validate questions and hints across three classes of confusion. The development process consisted of three phases: iterative development, automated assessment, and human validation.

In the first phase, we iteratively developed questions and hints by gathering feedback from individuals during the answering process. Our goal was to create a progression of information clarity, starting from minimal information (UC), adding partial information (PC), and complete information (NC). For transparency and reproducibility, we documented all materials, including introductory content, questions, and hints, in pdf documents available in our GitHub repository [9].

In the second phase, we employed the Large Language Model (LLM) Claude 3.5 Sonnet [1] to obtain objective ratings of confusion, clarity, and confidence levels for each question-hint combination. To ensure unbiased assessment, hints were randomly labeled (as Hint 1, 2, or 3) during the rating process. The results of this automated assessment are presented in Figure 2. The prompt template and the results can be found in our GitHub repository [9].

Finally, we conducted a manipulation check with six individuals (two for each condition) to validate the intended confusion induction. The results aligned with our designed progression: participants experienced higher levels of confusion in the UC condition, moderate challenge in the PC condition, and no confusion in the NC condition. This validation confirmed the effectiveness of our information manipulation strategy and strengthens the reliability of our subsequent analyses of learning outcomes and engagement.

3.5 Procedure

The experiment followed a structured procedure to ensure consistency across all participants. The procedure was divided into key phases:

Consent Form and Pre-Experiment Measures: Participants were welcomed and given a brief introduction to what was expected of them during the experiment. As a first step, participants completed the consent form, where they accept the participation terms and gave us permission to use their data in our research, and completed a pre-experiment questionnaire designed to provide a background profile of the participants. The questions covered demographic aspects such as age, gender, and education, as well as their current knowledge of AI and their learning methods.

Introduction: Upon starting the experiment, participants were greeted by the robot as they began their interaction. The robot reintroduced the study, explained the purpose of the session, and informed participants that they would be learning about basic AI concepts while interacting with it. Participants were also reminded that they could ask for hints at any time if they encountered difficulties during the questions.

Familiarization Round To ensure participants were comfortable with the robot and the interaction process, the first round served as a familiarization phase. During this phase, all participants received the same non-confusing (NC) condition treatment, with the robot providing clear and straightforward hints when requested. This standardized approach helped participants become comfortable with the question-answer format, the process of requesting hints, and the overall interaction pattern.

Main Experimental Rounds Following the familiarization round, participants proceeded to two main experimental rounds covering supervised/unsupervised learning and reinforcement learning respectively. Each participant was randomly assigned to one of three conditions that remained constant throughout these rounds: No Confusion (NC), Productive Confusion (PC), and Unproductive Confusion (UC). Each round consisted of three multiple-choice questions. After each answer, regardless of condition, the robot provided immediate feedback about the correctness of the response. The robot maintained consistent behavior within each condition throughout these rounds, varying only in its hint-giving approach.

Post-Experiment Measures: At the end of the experiment, participants completed a brief questionnaire assessing their subjective experience.

Debriefing The experiment concluded with a debriefing session. During this phase, participants were informed about the true purpose of the study and received detailed explanations about the different confusion conditions and their pedagogical rationale. Special attention was given to explaining each participant's assigned condition, which was particularly important for those in the UC condition to understand why they received limited support from the robot and why they may felt frustrated. The session ended with expressing gratitude for their participation and contribution to the research.

The entire procedure, from welcome to debriefing, was designed to last approximately 15-20 minutes per participant.

3.6 Measures

We decided to collect a variety of data to assess both the participants' learning outcomes and their engagement throughout the interaction. The measures included:

Self-Report Data: Participants were asked to complete a set of questionnaires after the interaction session to provide insight into their subjective experience. This included measuring their perceived level of confusion and frustration, and their engagement with the robot. Additionally, participants rated their believed learning outcomes and understanding of AI concepts after the learning session. The questionnaires used a 5-point Likert scale to ensure consistent measurement across all participants.

Observer Ratings of Confusion and Engagement

We acted as observers to assess participants' confusion levels and engagement during the experiment. The observation protocol consisted of two components: real-time confusion ratings and post-session engagement evaluation. For confusion assessment, we rated participants' responses to each question on a 5-point scale. Following the interaction session, we completed the Engagement Observation Scale [8], which evaluates eight distinct dimensions of participant engagement: level of enjoyment, level of frustration, conceptual understanding, attention maintenance, nervousness, distraction levels, need for encouragement and overall commitment.

Observer assessments were implemented to provide objective behavioral measures that complemented the self-reported data.

Performance Data: Performance outcomes were measured based on the accuracy of participants' responses to the multiple-choice questions. The correct answers were recorded, and the number of correct responses was used to assess participants' learning outcomes.

4 Results

4.1 Participant Demographics

For this study, 15 participants were recruited and assigned to three conditions: non-confusion (n = 5), productive confusion (n = 5), and unproductive confusion (n = 5). Participant demographics, including gender, education level, and prior experience with robots, are summarized in the attached Table 2. Most participants were students at KTH, while others represented diverse professional or academic backgrounds. Participants ranged in age from 21 to 37 years (M = 26, SD = 5.4), and came from various regions across the world, including Europe and Asia.

Table 2: Demographic Summary of Participants

Category	Count
Total Participants	15
Female Participants	6
Male Participants	9
Completed Higher Education	5
Currently Pursuing Higher Education	7
High School Only	1
Interacted with AI Robot	6
Not Interacted with AI Robot	9

4.2 Participant Performance

The experiment was designed to record participant performance throughout its duration. Data was collected on the number of correct answers, the number of hints used, and the success rate for each participant. These results, aggregated from all participants, are presented in Table 3 and.

Table 3: Mean results of participant performance by confusion type.

Group	M (Correct Answers)	M (Success Rate)	M (Hints Used)
NC	5.6	0.932	3.2
PC	5.4	0.900	2.6
UC	5.0	0.832	2.4

The data shows that participants in the NC and PC groups tend to achieve more correct answers and a higher success rate compared to those in the UC group. Participants in the NC group, however, tend to use more hints than the other two groups, while participants in the UC group use the fewest hints. This behavior may be explained by the possibility that hints in the UC group introduce additional confusion, leading participants to attempt answering without relying on hints.

4.3 Results of Observer Ratings

Observers were tasked with monitoring participants and rating their levels of frustration and enjoyment during the experiment. This was carried out using a questionnaire, and the collected data was later compiled and presented in Figure 3 and Table 4.

The hypotheses for these observations were formulated as follows:

- **Hypothesis**:
  - $\mathbf{H_0}$: There is no difference in frustration levels between the groups.
  - $\mathbf{H_1}$: Participants in the UC group exhibit higher frustration levels compared to those in the PC and NC groups.

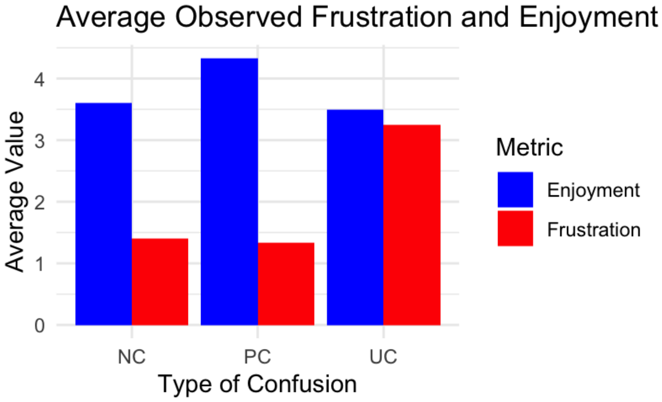

Figure 3: Enjoyment and Frustration across Confusion Types, based on observer rating

Figure 4 illustrates that the observed enjoyment levels are highest in the PC condition, moderate in the NC condition, and lowest in the UC condition. In contrast, frustration levels are observed to be highest in the UC condition, while remaining relatively low and similar in the PC and NC conditions.

The analysis of observer ratings revealed significant differences in participants' observed frustration levels across the confusion types. Table 5 summarizes the p-values obtained from Tukey's post-hoc test. Significant differences were found between **Unproductive Confusion (UC)** and both **No Confusion (NC)** ($p = 0.011$) and **Productive Confusion (PC)** ($p = 0.006$). However, no significant difference was observed between **NC** and **PC** ($p = 0.989$).

Based on these results, the null hypothesis ($\mathbf{H_0}$)—that there is no difference in frustration levels between the groups—can be partially rejected. Specifically:

- Frustration levels in the **UC** group were significantly higher than in the **NC** and **PC** groups, supporting the alternative hypothesis ($\mathbf{H_1}$).
- However, frustration levels between the **NC** and **PC** groups did not differ significantly, meaning the null hypothesis cannot be rejected for this pairwise comparison.

Table 4: Tukey's post-hoc test p-values for frustration levels across confusion types.

Comparison	p-value
PC - NC	0.989
UC - NC	0.011
UC - PC	0.006

4.4 Results of Self-Report Data

After completing the interaction with the robot, participants answered a post-experiment questionnaire. The purpose of this questionnaire was to gain a general overview of the participants' experiences, focusing on their emotional responses, self-assessed improvement, and perceived knowledge development.

The following three hypotheses were formulated:

Hypothesis 1: Emotional Experience

- $\mathbf{H_0}$: There is no difference in the emotional experience (e.g., enjoyment or frustration) across the confusion states (NC, PC, UC).
- $\mathbf{H_1}$: Participants in the PC condition report higher levels of positive emotional experiences (e.g., enjoyment) compared to those in the NC and UC conditions.

Hypothesis 2: Perceived Knowledge Development

- $\mathbf{H_0}$: There is no difference in perceived knowledge development across the confusion states.
- $\mathbf{H_1}$: Participants in the PC condition perceive higher knowledge development compared to those in the NC and UC conditions.

Hypothesis 3: Engagement and Frustration

- $\mathbf{H_0}$: There is no difference in engagement levels across the confusion states.
- $\mathbf{H_1}$: Engagement levels are highest in the PC condition, moderate in the NC condition, and lowest in the UC condition. Additionally, frustration levels are highest in the UC condition compared to PC and NC.

The post-experiment questionnaire data analysis evaluated differences in emotional experiences, perceived knowledge development, and engagement across confusion types (NC, PC, UC). Table 5 summarizes the p-values obtained from ANOVA tests. The results show no significant differences in frustration ($p = 1.000$), enjoyment ($p = 0.825$), or knowledge improvement ($p = 0.687$) between the groups. Consequently, the null hypotheses for all three formulated hypotheses cannot be rejected, indicating no statistically significant differences across the confusion types.

Table 5: P-values for Frustration, Enjoyment, and Knowledge Improvement Across Confusion Types

Variable	p-value
Frustration	1.000
Enjoyment	0.825
Knowledge Improvement	0.687

Note. p-value represents the statistical significance from the ANOVA test for each variable. A p-value below 0.05 indicates a significant difference between groups.

5 Discussion

5.1 Findings

5.1.1 *Findings on Observer Ratings*

Confusion Levels

The average confusion levels across the three rounds, as illustrated in the boxplot (Figure 4), show distinct trends among the three confusion classes (*NC*, *PC*, and *UC*). Participants in the *NC* (No Confusion) group consistently reported the lowest confusion levels across all rounds (*Median* = 1 in Rounds 1 and 3, *Median* = 2 in Round 2), with minimal variability (*SD* < 1.30). This indicates that tasks were clear and posed minimal cognitive challenges for these participants.

Participants in the *PC* (Productive Confusion) group experienced the highest confusion levels in Round 1 (*Median* = 4, *SD* = 1.52), which decreased significantly in subsequent rounds (*Median* = 2 in Rounds 2 and 3, *SD* < 1.00). This pattern suggests that productive confusion initially challenged participants but became less prominent as they adapted and gained understanding.

In contrast, participants in the *UC* (Unproductive Confusion) group reported consistently high confusion levels across all rounds (*Median* = 3), with the greatest variability observed in Round 3 (*SD* = 2.00). These findings, supported by descriptive statistics summarized in Table 6, highlight the balance between clarity and challenge in fostering engagement and learning. Productive confusion appears beneficial, whereas unproductive confusion hampers participant outcomes.

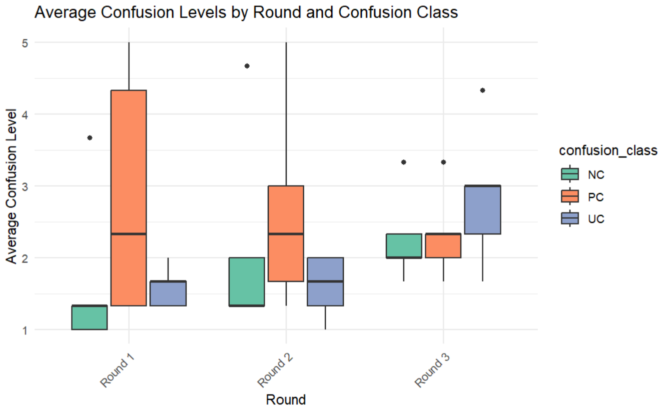

Figure 4: Average Confusion Levels.

Table 6: Descriptive Statistics for Average Confusion Levels Across Rounds by Confusion Class

Confusion	Round	Median	SD	n
NC	Round 1	1	1.30	5
	Round 2	2	1.10	5
	Round 3	1	1.00	5
PC	Round 1	4	1.52	5
	Round 2	2	1.00	5
	Round 3	2	0.90	5
UC	Round 1	3	1.50	5
	Round 2	3	1.80	5
	Round 3	3	2.00	5

Frustration Levels

The observed frustration levels align with the alternative hypothesis ($\mathbf{H_1}$), indicating significant differences among the three confusion classes (*NC*, *PC*, and *UC*). As summarized in Table 7, participants in the *UC* (Unproductive Confusion) group reported the highest frustration levels (*M* = 2.8, *SD* = 1.30, *Median* = 3). In contrast, the *NC* (No Confusion) group exhibited the lowest frustration levels (*M* = 1.2, *SD* = 0.45, *Median* = 1), reflecting minimal cognitive challenges during the task. The *PC* (Productive Confusion) group reported moderate frustration levels (*M* = 1.4, *SD* = 0.89, *Median* = 1), suggesting that while the tasks were challenging, they were engaging and did not escalate to unproductive confusion.

These findings highlight the detrimental effects of unproductive confusion, as seen in the *UC* group, compared to the balance of challenge and engagement experienced by the *PC* group. The results emphasize the importance of designing tasks that promote productive engagement while minimizing unproductive frustration.

Table 7: Descriptive Statistics for Frustration Levels by Confusion Class

Confusion	Mean	Median	SD	n
NC	1.2	1	0.45	5
PC	1.4	1	0.89	5
UC	2.8	3	1.30	5

5.1.2 *Findings on Self-Report Data*

Overall Experience

From self-reported data, we evaluated the overall experience scores in the three confusion classes: *NC* (No Confusion), *PC* (Productive Confusion), and *UC* (Unproductive Confusion). The median scores for overall experience were as follows: *NC* and *PC* both had a median score of **4**, while *UC* had a lower median score of **3**. This indicates that participants in the *UC* group reported a less favorable experience compared to the other two groups. Additionally, the mean scores align with this trend, with *NC* having the highest mean (**4.4**), followed by *PC* (**3.8**), and *UC* (**3.4**). The standard deviation for *UC* (**1.14**) was the largest, suggesting higher variability in experiences within this group.

Below is the box plot illustrating the distribution of overall experience scores across confusion classes (Figure 5).

These findings suggest that confusion negatively influences participants' overall experience, especially when it remains unresolved. Ensuring clear instructions and providing additional support during challenging sections could improve engagement and satisfaction, particularly for participants who might fall into the *UC* category. The descriptive statistics for overall experience scores across confusion classes are summarized in Table 8.

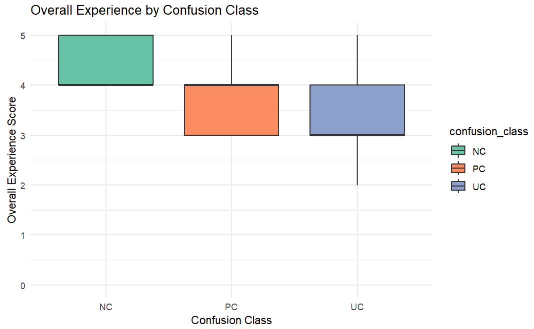

Figure 5: Overall Experience.

Table 8: Descriptive Statistics for Overall Experience by Confusion Class

Confusion	Median	Mean	SD	Count
NC	4	4.4	0.55	5
PC	4	3.8	0.84	5
UC	3	3.4	1.14	5

Learning Metrics

The learning metrics reveal distinct trends across the three confusion classes (*NC*, *PC*, and *UC*). For the *absorption* metric, both *NC* and *PC* groups had a median score of 3, indicating moderate engagement, while the *UC* group reported a lower median of 2 with the highest variability (*SD* = 1.52). Similarly, for the *enjoyable* metric, all groups reported a median score of 3; however, the *UC* and *NC* groups exhibited greater variability (*SD* = 1.30) compared to the *PC* group (*SD* = 0.83).

The learning metric demonstrated a consistent median score of 4 across all groups, though the *NC* group showed greater variability (*SD* = 1.52) than the others. Finally, for *knowledge_improved*, the *NC* group reported the highest median score of 3, while both *PC* and *UC* groups had lower medians of 2. The *UC* group showed the greatest inconsistency (*SD* = 1.52), highlighting the detrimental effect of unproductive confusion on learning outcomes. These findings suggest that unproductive confusion negatively impacts engagement and perceived knowledge improvement, whereas clear understanding fosters a more rewarding experience. The descriptive statistics for the self-reported metrics are presented in Table 9.

Table 9: Learning Metrics by Confusion Class

Confusion	Metric	Median	SD	n
NC	Absorbed	3	1.30	5
PC	Absorbed	3	1.10	5
UC	Absorbed	2	1.52	5
NC	Challenging Fun	4	1.52	5
PC	Challenging Fun	4	0.89	5
UC	Challenging Fun	4	0.84	5
NC	Enjoyable	3	1.30	5
PC	Enjoyable	3	0.83	5
UC	Enjoyable	3	1.30	5
NC	Knowledge Improved	3	1.52	5
PC	Knowledge Improved	2	1.30	5
UC	Knowledge Improved	2	1.52	5

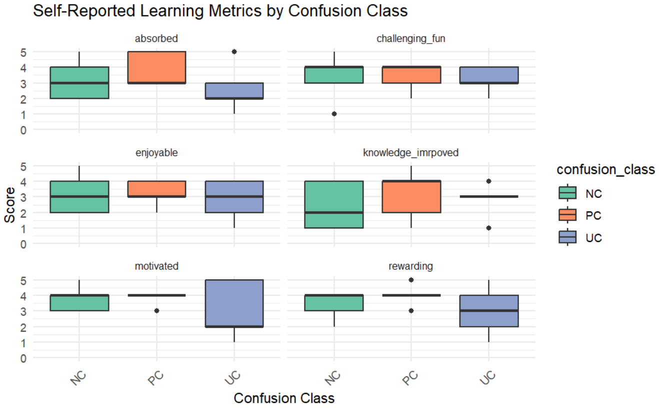

Figure 6: Learning Metrics.

Engagement

The engagement levels varied significantly across the three confusion classes (*NC*, *PC*, and *UC*). Participants in the *PC* (Productive Confusion) group reported the highest median engagement score (*Median* = 5, *M* = 4.6, *SD* = 0.55), indicating that a moderate level of challenge may enhance engagement. The *NC* (No Confusion) group followed with a median engagement score of 4 (*M* = 4.0, *SD* = 1.00), suggesting that clarity fosters consistent engagement. In contrast, the *UC* (Unproductive Confusion) group reported the lowest engagement levels (*Median* = 3, *M* = 3.2, *SD* = 1.30), with the highest variability, indicating that unproductive confusion significantly hinders engagement. These findings highlight the importance of maintaining an optimal balance between challenge and clarity to promote participant engagement. The descriptive statistics for engagement levels by confusion class are presented in Table 10.

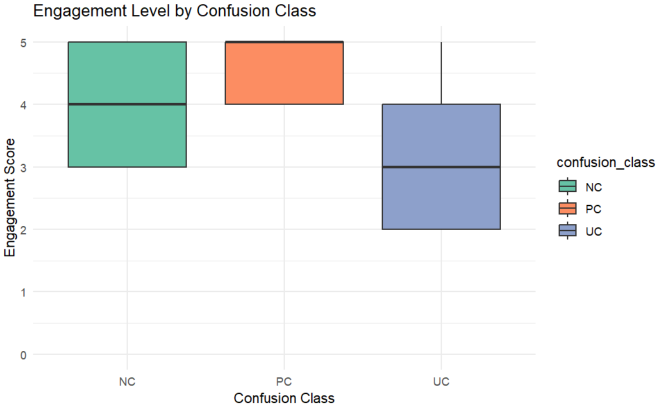

Figure 7: Engagement Level by Confusion Class.

Table 10: Descriptive Statistics for Engagement Levels by Confusion Class

Confusion	Median	Mean	SD	Count
NC	4	4.0	1.00	5
PC	5	4.6	0.55	5
UC	3	3.2	1.30	5

5.2 Limitation

While our study provides valuable insights into confusion in learning within HRI, several limitations should be acknowledged:

Absence of Theoretical Framework: Due to the developing nature of research on confusion in HRI, we could not follow an established theoretical framework for implementing and validating confusion states. This necessitated developing our own experimental approach, which, while grounded in learning theories, requires further validation through replication studies and theoretical development.

Limited Confusion Stimuli and Manipulation: Our experiment focused solely on insufficient information as a confusion stimulus in the robot's explanations. Other potential patterns and types of confusion-inducing stimuli, such as complex problem-solving scenarios, contradictory information, and feedback inconsistency, were not explored. This narrow focus on a single type of confusion stimulus may limit the generalizability of our findings.

Confusion Stimuli and Information Length: The length of hint information varied across confusion conditions, which may have influenced participant interactions. Specifically, UC hints were notably brief, PC hints provided a moderate amount of information, while NC hints were detailed. This variation in hint length might have affected participant engagement differently, with short hints potentially causing frustration in UC condition, and lengthy hints possibly leading to boredom in NC condition. For future studies, standardizing the word count across all conditions while maintaining the intended confusion levels could help isolate the effect of confusion from the potential confounding effect of hint length.

Methodological Challenges in Confusion Assessment: The subjective nature of confusion presented significant measurement challenges. Participants may not always be aware of their confusion states [6], making self-reported data potentially unreliable. While multimodal data collection could provide more objective measures, the behavioral markers of confusion are not yet well-established in the literature. Given our time constraints, we opted to focus on observational data collection through observers' annotation. Future studies could benefit from multiple annotators, standardized confusion markers, and more comprehensive data collection methods to identify confusion states more accurately.

Homogeneous Participant Demographics: The participant pool consisted exclusively of university students from KTH, predominantly from technical backgrounds. This homogeneous sample limits the generalizability of our findings to broader populations with different educational backgrounds, learning styles, and cultural contexts.

5.3 General Discussion

This study highlighted the crucial role of confusion in shaping user engagement and learning outcomes in HRI. The findings strongly support the conclusion that unproductive confusion, as observed in the UC group, leads to higher frustration and reduced enjoyment. In contrast, productive confusion in the PC group fosters engagement and promotes higher enjoyment levels while keeping frustration low. The no confusion group, NC, with minimal cognitive challenges, demonstrates the lowest levels of frustration and moderate enjoyment. These results emphasize the importance of designing tasks that balance challenge and clarity to maximize engagement and minimize negative emotional responses.

Our findings revealed that productive confusion contributed to a more positive emotional experience, higher perceived knowledge development, and greater engagement compared to unproductive confusion. The productive confusion group reported increased enjoyment and engagement, which supports the hypothesis that participants in the PC group report greater levels of positive emotional experience. However, the hypothesis suggesting that participants in the PC group would perceive higher knowledge development was not fully supported. Lastly, engagement levels were highest in the PC group, moderate in the NC group, and lowest in the UC group, with frustration levels also notably higher in the UC condition, emphasizing the importance of managing confusion to maintain engagement and minimize frustration in learning environments.

Our results underscore that confusion management in HRI may have important implications for both research and practice in educational robotics. First, they demonstrate the potential for enhancing learning and engagement in HRI through strategic confusion management. Second, our findings suggest that existing HRI learning strategies could benefit from considering confusion as a key variable in their design. Understanding and monitoring confusion levels could help explain variations in participant engagement and emotional responses, potentially leading to more effective educational interventions. This insight is particularly valuable in understanding why participants might become disengaged or experience negative feelings in certain learning scenarios, offering a new perspective for improving human-robot educational interactions.

6 Conclusions

In this paper, we presented an experimental design to investigate the potential benefits of confusion in learning in HRI. We detailed our experimental design, manipulation check, data collection, and analyzed and discussed the findings. This study shows potential for the development of social robots that enhance learning and engagement by effectively managing confusion.

However, our study also revealed important limitations and areas for future research. Further investigation is needed to develop more refined strategies for implementing productive confusion in HRI settings. Additionally, expanding the participant pool beyond university students would provide more generalizable insights into how different populations respond to confusion-based learning strategies with social robots.

Overall, these findings contribute to our understanding of how confusion can be used as a pedagogical tool in HRI, while also highlighting the complexity of implementing such approaches effectively. The results provide a foundation for developing more sophisticated social robots that can enhance learning through strategic management of confusion.

References

[1] Anthropic. 2024. Claude 3.5 Sonnet. Anthropic AI Model. Available at https://www.anthropic.com.

[2] Amaël Arguel and Rod Lane. 2015. Fostering deep understanding in geography by inducing and managing confusion: an online learning approach. In *Globally connected, digitally enabled. Proceedings ascilite 2015 in Perth*, T. Reiners, B.R. von Konsky, D. Gibson, V. Chang, L. Irving, and K. Clarke (Eds.). Australasian Society for Computers in Learning in Tertiary Education, CP:22–CP:26.

[3] Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. 2018. Social robots for education: A review. *Science Robotics* 3, 21 (2018), eaat5954.

[4] Sidney D'Mello, Blair Lehman, Reinhard Pekrun, and Art Graesser. 2014. Confusion can be beneficial for learning. *Learning and Instruction* 29 (2014), 153–170.

[5] Sidney D'Mello and Arthur Graesser. 2014. Confusion. In *International Handbook of Emotions in Education*. Routledge, 289–310.

[6] Na Li, John D. Kelleher, and Robert Ross. 2021. Detecting Interlocutor Confusion in Situated Human-Avatar Dialogue: A Pilot Study. In *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue*. 190–193.

[7] Na Li and Robert Ross. 2023. Invoking and identifying task-oriented interlocutor confusion in human-robot interaction. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. ACM/IEEE, 578–587.

[8] D. Martinovic, G. H. Burgess, C. M. Pomerleau, and C. Marin. 2016. Computer games that exercise cognitive skills: What makes them engaging for children? *Computers in Human Behavior* 60 (July 2016), 451–462. https://doi.org/10.1016/j.chb.2016.02.063

[9] Andreas Naoum. 2024. Repository for Experiment: Learning Through Confusion in HRI. https://github.com/andreasnaoum/Learning-Through-Confusion-HRI

[10] Andreas Naoum. 2024. Video Demonstration of Experiment Learning Through Confusion in HRI. https://youtu.be/JnoAcd8MO3khttps://youtu.be/JnoAcd8MO3k

[11] Diyi Yang, Robert E. Kraut, and Carolyn P. Rose. 2016. Exploring the Effect of Student Confusion in Massive Open Online Courses. *Journal of Educational Data Mining* 8, 1 (2016), 52–83.