Can You See How I Learn? Human Observers' Inferences about Robot Learning Processes

Bernhard Hilpert[1,2], Muhan Hou[2], Kim Baraka[2] and Joost Broekens[1]

***Abstract*— Interactive robot learners often exhibit learning behaviors that are not intuitively interpretable by human observers, which can result in suboptimal feedback in collaborative teaching settings. Yet, how humans perceive and interpret robot learning behavior is largely unknown. In a bottom-up approach with two experiments, this work provides a data-driven overview of the factors influencing human observers' understanding of the robot learning process. A novel, observation-based paradigm to directly assess human inferences about robot learning was developed. In an exploratory interview study (*N*=9), four common themes in human inferences were identified: Agent Goals, Knowledge, Decision Making, and Learning Mechanisms. A second confirmatory study (*N*=34) applied an expanded version of the paradigm across two tasks (navigation/manipulation) and two Reinforcement Learning (RL) algorithms (tabular/function approximation). Analyses of 816 responses confirmed the reliability of the paradigm and refined the thematic framework, revealing how these themes evolve over time and interrelate. Our findings provide a human-centered understanding of how people make sense of robot learning, offering actionable insights for designing interpretable RL systems and improving transparency in Human-Robot Interaction (HRI).**

I. Introduction

Hybrid Intelligence (HI) refers to collaborative human–machine systems that combine strengths to accomplish tasks neither could perform alone [1]. A key example is Human-Interactive Robot Learning (HIRL), where human teachers guide learning robots using signals such as feedback or demonstrations [2]. While humans can provide insight, expert knowledge, or common sense [3], feedback is often suboptimal: delayed, inconsistent or based on misinterpretation of the robot's behavior [4]. Recent surveys emphasize explainability and human involvement across all stages of the HIRL lifecycle [5]. However, while mental models [6] and their mismatches [7] have been a key area of research in HRI, existing literature predominantly examines teacher inferences about robot learning indirectly via interaction patterns [8] or simulated settings [9]. Thus, existing paradigms confound observation with interaction: when humans actively teach, their expectations shape not only the feedback but also the robot's behavior, making it hard to isolate what they actually understand [8]. This may obstruct an unbiased examination on the human's perspective of the learning progress, especially at later learning stages. Open questions include the correct setting, how to even query inferences about a robot's learning process from human observers, and which aspects of learning humans attend to when making sense of robot behavior.

The main Research Question for this study therefore is: *What do human teachers infer about RL-based robot learning processes from observing the robot's behavior?* To investigate how observers naturally conceptualize robot learning, this work takes a human-centered, bottom-up approach, resulting in two main contributions: 1) a novel observation-only experimental paradigm designed to isolate and directly elicit intuitive observer inferences about robot behavior, 2) a granular, empirical framework identifying four recurring inference themes (Goals, Knowledge, Decision Making, and Learning Mechanisms) with 16 subclusters, empirically collected across two experiments, spanning different tasks and Reinforcement Learning (RL) algorithms. A detailed analysis further examines how they evolve over time and contribute to observers' understanding of robot learning[1].

II. Background and Related Work

The mechanisms of RL are inherently difficult for human observers to interpret from behavior alone, due to algorithmic complexity [10] and the challenge of intuitively interpreting concepts like reward functions [11]. In interactive settings, a teacher's perception of a RL-based robot's behavior can recursively shape their interpretation of the learning process and thereby influence the course of the interaction [12]. For instance, [8] demonstrate that teachers are strongly inclined to use rewards in the teaching process intuitively as communicative signals rather than a reinforcement signal, indicating a persistent misalignment between users' assumptions and actual RL mechanisms. Although RL agents are designed to optimize cumulative rewards, the observed interaction pattern suggests that human teachers seem to interpret their actions as if the agent were responding to the communicative intent rather than reinforcement signals. The authors explained this through assumed differences in how teachers interpret the agent's state space, and that teachers seem to evaluate the agent by comparing its inferred policy to their own desired policy [8]. While such judgments may reflect a reasonable teaching strategy from the human perspective, the formation of positive reward cycles in the teaching process based on the resulting feedback [8] highlights a disconnect between human interpretations of learning and the underlying mechanisms of RL.

[1]LIACS, Leiden University, Leiden, The Netherlands
[2] VU Amsterdam, Amsterdam, The Netherlands

[1]A preliminary version was presented as an extended abstract at AAMAS 2025 and as a non-archival paper at the AAMAS ALA Workshop 2025 (available as a preprint on arXiv).

Further work indicates that these communicative assumptions about feedback generalize to Human-Human Interaction, where punishment and reward are also interpreted through inference [13]. More broadly, the Inferential Social Learning framework describes behavior interpretation as a process of "probabilistic inferences, guided by an intuitive understanding about how people plan and act" [14]. This suggests that individuals acquire knowledge about their interaction partners by inferring insights from behavior observation. Understanding how this process applies in the context of Human-Agent Interaction (HAI) settings, is therefore crucial in order to design effective HIRL settings. However, research in HAI has primarily studied teacher inferences indirectly, by analyzing interaction behaviors. The inferences themselves have rarely been studied directly. The present work addresses this gap by directly examining how human observers infer and interpret the learning processes of RL-based robots from behavior observation alone.

III. Exploratory Interview Study

In this first exploratory study, an observational experimental paradigm is developed and applied, through which this question is studied and its exploratory results presented.

A. Methodology

1) Participants: A total of *n*=9 (*n*=3 female, *n*=6 male) were recruited. Five were bachelor students with RL exposure, four had no RL background.

2) Set-up & Procedure: In the absence of a validated paradigm to directly examine observer inferences about RL processes (as a proxy for HIRL), an adapted version of the Grounded Theory framework was used to develop a reliable experimental procedure [15]. An exploratory qualitative interview setting was selected, due to the suitability of this method for studying undirected perceptual data [16]. Participants observed an RL agent's learning process with a framing of being a teacher, but no possibility for intervention. Firstly, they received a short introduction to RL together with an example video similar to the later stimuli and were asked to imagine helping the agent learn the task by teaching. Secondly, they were presented with videos of an RL agent learning to solve a task and were asked to share their observations about the learning process in a think aloud protocol [17] to elicit a verbal report that is as undirected, undisturbed and constant as possible [18]. Interviews (ca. 45 minutes) were conducted in person, audio-recorded; recordings were erased after an anonymized transcription. A written consent for this procedure was obtained from the participants beforehand; they were debriefed after the experiment.

3) Stimuli: A tabular Q-learning agent with Temporal Difference update was implemented. Stimuli consisted of rendered recordings of this agent solving the default 4x4 and 8x8 grids of the OpenAI gym *FrozenLake* environment [19]. To support coherent perception of the learning process, but also assess whether chunking affects richness of observations, two presentation formats were used. One group viewed two full-length videos (entire learning process); the other viewed six shorter clips, each showing a third of the process. Participants were balanced across these conditions and RL backgrounds (2 per condition).

4) Measurements: Participants were instructed to explicitly report on the agent's learning process (rather than behavior) while watching the videos. With no validated questionnaire available, a pilot session with an RL expert guided question phrasing in line with Grounded Theory principles [15], specifically formulations that queried most sensibly the human inferences about the learning process. Based on their input, two open-ended questions were conceived: "*What do you think is going on in the brain of the agent?*" and "*Why do you think the agent is doing what it is doing?*"

5) Analysis: Audio recordings were transcribed and participant answers were pooled per video, to ensure full anonymization. Individual statements were then isolated to enable systematic analysis of the think-aloud data. Then, an inductive, reflexive thematic analysis was conducted based on an adaptation of [20], following a three-phase pipeline: (1) Data Reduction: condensing statements into core semantic units and grouping by synonymy; (2) Theme Generation: clustering core meanings into latent themes and assigning descriptive labels; and (3) Refinement: cross validating themes against the full dataset to ensure internal consistency. To maintain conceptual continuity across datasets, primary coding was executed by a single investigator, following the recommendation of [20]. To establish systemic trustworthiness and counteract individual coder bias, the evolving thematic structure was subjected to 30 hours of systematic peer-debriefing and critical review with senior HRI researchers, followed by an external validation workshop ($N = 26$) to verify the accuracy of the generated framework.

B. Results

Data collection with the presented paradigm resulted in 264 isolated statements. Thematic analysis revealed four common emergent themes in observers' inferences about agent learning processes.

Agent Goals: includes 23 inferences about the learning agent's goals. Examples include "*The agent wants to explore and understand the environment that it is in*", "*It is trying to avoid the ice*", "*It is trying to find a way that it can navigate the environment as quick as possible*" and "*seems not to realize what the idea is*".

Agent Knowledge: inferences that the agent's learning process involved acquiring and using knowledge about the environment or task (36 mentions). Examples include "*It seems to have gathered some sort of sense of the second row being a bad row*", "*it remembers, "alright I was in that tile and then I moved to that tile and there was a lake right there"*", "*he knows exactly which tiles to not step onto, because he memorized it*".

Agent Decision Making: inferences about how the agent takes decisions during learning, including individual actions and overall policies (50 mentions). Examples include "*It is just trying out different moves*, "*He follows a system: 1 down, one right, then 2 down one right, then three down*" and

"*He realized he cannot take the first second rows and thinks whether he can take the third*".

Agent Learning Mechanisms: inferences about mechanisms by which the agent updated its internal representations (18 mentions). Examples include "*it is just mapping certain actions to certain states*", "*He has moved two tiles to the right in the first try and realized it was not a mistake*".

Other: includes additional themes that were too infrequent to be classified as separate themes like perception (e.g., vision), feelings, communicative signals.

C. Discussion

This experiment introduced a new observational paradigm to examine observer inferences of agent's learning process from its behavior. A data-driven analysis revealed four recurring themes in these inferences — Agent Goals, Knowledge, Decision Making, and Learning Mechanisms — offering insight into how observers made sense of learning agents over time.

Agent Goals emerging as a distinctive theme indicates that task understanding plays a key role in observer inferences of the learning process. It reveals observer assumptions about the agent's understanding of the task structure, but could also reflect the observers' own interpretation of the task and environment. This is especially important, as every user approaches each task with their own set of assumptions, which can shape or even distort their interpretation of learning. For instance, inferring that the agent aims to reach the exit "as safely as possible," despite safety not being encoded in the reward structure. This highlights a broader challenge in human-agent interaction, where users project task goals that may misalign with an agent's actual goals.

The emergence of Agent Knowledge indicates that observers inferred some form of internal representation of its past experiences or task structure. Unlike Goals, knowledge is a latent, less directly observable concept, yet observers appeared to attribute it readily. This indicates a more cognitively rich, anthropomorphized mental model of the agent, using assumed knowledge to explain how learning unfolds.

The emergence of Agent Decision Making indicates that observers constructed a procedural view of the agent's learning process, beyond only structural components like goals and knowledge. Rather than seeing the agent as passively reacting to stimuli, observers infer it actively weighing options, selecting actions, or following strategies. This suggests that observers interpret learning not just as an internal update, but as an active process of deliberation with a sense of control, similar to human-like decision making. This may hint at a potential over-attribution of deliberate strategy to the agent's processes.

The emergence of Learning Mechanisms suggests that observers inferred not just behavior, but processes of adaptation, seeing them as actively learning entities.

The analysis did not reveal any differences in inferences that were based on demographic variables.

Together, the four themes suggest that observers employ an integrated, procedural view of agent learning, that appears to be anthropomorphized, aligning with prior findings in cognitive science [14], [8] and Human-Agent studies [21], [22], where the attribution of known concepts made complex behavior more interpretable.

Limitations and Lessons Learned

Given the small sample size and exploratory nature of the study, the findings should be viewed as preliminary. A larger-scale study is needed to validate these initial results in a more systematic and robust manner. Secondly, participants in the split-up video condition produced approximately 3 times as many statements as those in the full-length video condition, suggesting that this format supports more differentiated and richer inferences. To expand the broad concepts found so far, the next step requires refining the experimental paradigm with a more structured, differentiated questioning format. Next, the framework proved to be highly anthropomorphized. This could in part be due to the human tendency to apply pre-existing knowledge to virtual agents to facilitate understanding [23], but also due to the stimulus material (i.e. the human-like agent in FrozenLake) and the question formulation (i.e. referring to the agent's brain processes). A next step should employ a modification of the paradigm with less anthropomorphized, robot-centric stimuli and questioning. Lastly, this study examined observer inferences about an agent's learning process only for a very simple tabular agent. Since most interactive teaching settings in robotic applications employ more sophisticated forms of RL, future work should also validate this framework for other types of agents.

IV. Confirmatory Questionnaire Study

Building on the exploratory findings of Experiment 1, this second experiment aimed to confirm those results while explicitly extending their generalizability in a larger-scale study. Specifically, it pursued three primary objectives:

1) Validate the initial findings on a larger sample
2) Add granularity to the four inference themes
3) Generalize findings by expanding the experimental paradigm to other RL algorithms and tasks

To address these objectives, a larger-scale confirmatory questionnaire study was designed and implemented[2].

A. Methodology

In line with open research practices [25], this experiment was preregistered and all materials (Questionnaire, instructions, stimuli, annotated data corpus and the preregistration protocols) are openly available on the OSF[3].

1) Participants: A total of 35 participants was recruited through Prolific of which one was excluded due to inattentive responses, resulting in a final sample of $N = 34$ (26 Male, 8 Female, 0 Non-binary, self-assigned or undisclosed). Inclusion criteria required participants to be 18 years or older, with a minimum of 95% approval rate on Prolific, 100 completed tasks, and self-reported fluency in English, confirmed by

[2]Data collection was conducted jointly with [24]

[3]https://osf.io/fumd8/?view_only=9cec60dccbd446f08bd818d0b3612705

country of residence. N = 34 (26 Male, 8 Female, 0 Non-binary, self-assigned or undisclosed). The participants ranged from 19 to 65 years old (M = 38.79 years, SD = 10.68 years). 23 participants owned a pet, 20 participants reported to have children. Participants received on average £10.42.

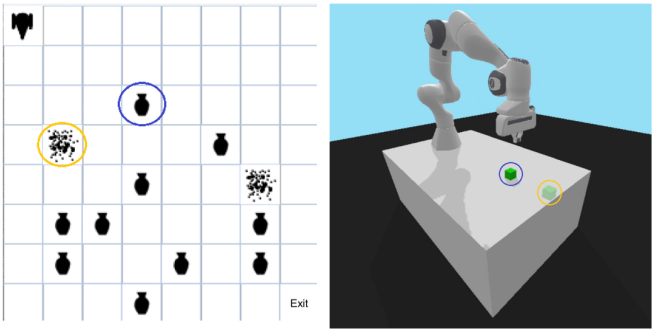

Fig. 1. a): NT: robot needs to clean dirt (yellow) and avoid breaking furniture (blue) b) MT: the assistive robot needs to push "medication" (blue) to target position (yellow)

2) Expanded experimental paradigm: To test the generalizability of the findings, the experimental paradigm was expanded in two key ways. First, a second task involving a Deep Deterministic Policy Gradient (DDPG; an off-policy, actor-critic algorithm for continuous action spaces) agent was introduced, to assess inferences about agent learning extend to function approximation settings. Second, to reduce anthropomorphism and increase ecological validity, both tasks were reframed as service robots performing everyday activities (see Figure 1).

In the navigation task (NT), the original tabular Q-learning agent operated in a modified 8x8 FrozenLake environment. Visual elements were adapted to depict a cleaning robot navigating a room: the elf was replaced with a conic robot icon (indicating orientation), ice tiles with white floor tiles, holes became vases, and the goal tile marked with an "exit" label. Two dynamic dirt patches were added, granting the agent an additional reward (+1) upon visit and disappearing afterwards, providing visible milestones to support observer inferences. Task logic and algorithm remained unchanged.

The manipulation task (MT) featured a DDPG agent framed as an assistive robot pushing a box of medication across a table to a target zone (where a patient that has trouble reaching could more easily pick it up). This was implemented using a modified Push task from the PandaGym environment [26], with fixed object/target positions near the table's edge. The reward function combined distance-based shaping rewards with a large terminal reward (+1000) upon successful delivery.

3) Set-up & Procedure: Participants observed the RL agents' learning in a non-interactive, teaching framing. After recruitment via Prolific, they were directed to a study landing page outlining the procedure, ethical rights, and withdrawal options, followed by an informed consent form and commitment check. Next, participants then received a lay-level introduction to RL concepts, designed to ensure baseline comprehension.[4] They were then presented with two task scenarios (see Section IV-A.2), each comprising three videos of a "robot's" learning trajectory (see Section IV-A.4), for a total of six videos. Following each video, participants completed a modular block of questions (see Section IV-A.5, median: 57 minutes). After this, participants were redirected to Prolific and received compensation within 24 hours. The study was approved by the university's ethics board.

4) Stimuli: To reflect genuine learning processes, stimuli were generated by training each robot to convergence (defined as 5 consecutive successful episodes) and rendering roll-outs from saved checkpoint policies. For the NT, the agent was trained for 10000 episodes, with checkpoints saved every 500 episodes. For the MT, the agent was trained for 100000 steps, with checkpoints saved every 1,000 steps. This yielded two full-length videos: 94 (NT) and 192 seconds (MT). Each source video was divided into three equal-length clips representing the early, middle, and late stages of learning, resulting in six stimulus videos in total.

5) Measurements: To examine observer inferences about the robots' learning process in greater detail, a modular questionnaire was designed, combining qualitative and quantitative assessments, divided into two equal blocks, one per task scenario. Each block began with a briefing, describing the robot's objectives and the environment's reward structure. This was followed by 11 question pairs, each comprising a qualitative and a corresponding quantitative query, presented after each stimulus video, which was available for rewatching above the questions. The first item served as a manipulation check: participants rated, in percent, how much they believed the robot had learned the task, to assess perceived learning progress. A baseline qualitative prompt—*"What do you think is happening in this video?"*—followed, ensuring participants had engaged with the stimulus. Each of the four inference themes (Goals, Knowledge, Decision Making, and Learning Mechanisms) was then assessed using a 7-point relevance rating (1 = *irrelevant*, 7 = *very relevant*), followed by a theme-specific open-ended question. For example: *"What goals/knowledge do you think the robot has in this video?"* or *"How do you think the robot makes decisions/learns in this video?"*. Finally, participants were asked to describe their teaching intuitions, reinforcing the interactive framing of the setting. This block of questions was repeated for each of the six videos. The questionnaire concluded with demographic questions assessing age, gender, and yes/no questions on pet ownership and having children.

6) Analysis: The analysis procedure of experiment 1 was reapplied to cluster the statements regarding each of the four common themes.

B. Results

This experiment examined observers' inferences about robot learning processes across two task scenarios and three learning stages. Subjective ratings on perceived learning

[4]Instructions, questionnaire and stimuli were piloted with N=4 test participants before recruiting in order to ensure understandability

progress (LP), theme relevance, and open-ended responses linked to the four common inference themes were collected.

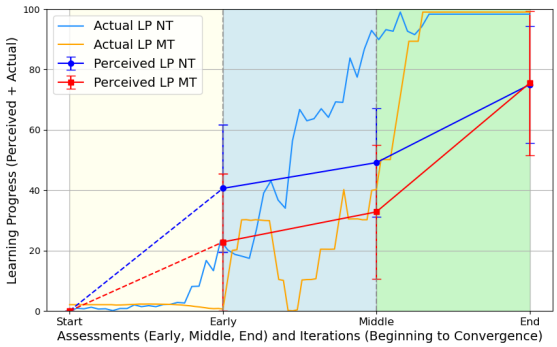

Fig. 2. Perceived vs. Actual Learning Progress

First, perceived LP was assessed to confirm that observers recognized learning improvement, validating the subsequent analysis. Figure 2 shows, that perceived LP increased steadily over time in both tasks. Early in the learning, observers rated the NT robot as learning more quickly; later, this pattern reversed, which mirrored the agents' actual LP curves based on cumulative rewards.

A second validation step assessed whether the common inference themes are also perceived as influential by the observers themselves (see Figure 3). One-sample t-tests (across both tasks) revealed that for each of the 3 time points, observers assigned a relevance significantly above a "neutral" rating of 4. Additionally, relevance ratings showed an increasing trend over time, suggesting growing salience of these themes as learning progressed.

Finally, the qualitative analysis examined thematic clusters of inferences within Goals (G), Knowledge (K), Decision Making (DM) & Learning Mechanisms (LM) (see Table I). From 816 participant responses, a total of 1,105 individual statements were extracted and coded, reflecting instances where responses contained multiple distinct inferences. Figure 4 presents the distribution of subcluster frequencies by task and learning phase.

Goals : Observer inferences regarding robot goals could largely be clustered into four categories.

(A) *Outcome-related Goals:* Inferred goals concerning task achievement, that define what constitutes success and failure in the task respectively. For the NT, this included all statements about achieving subgoals such as "*find the dirt and clean it*", "*avoid breaking any furniture*" and "*find the exit*" but also "*finding a pathway*". For the MT, such subgoals included "*locate the medicine*", "*hold onto it*", "*not to have it fly off of the table*" and "*move the box to the target zone*".

(B) *Execution-related Goals:* Inferred goals concerning improvement and optimization of task execution, that relate to how the robot aimed to improve or optimize its task performance. For both tasks, this included specific statements about efficiency (e.g. "*to clean the room efficiently*") or speed (e.g. "*the robot wants to push the box quick*"). Additionally, in the MT, this encompassed statements about solving the task with a high accuracy ("*to achieve accuracy of movement*") and safety ("*safely place it across the table for the patient*") as well as control in executing the pushing ("*To improve the force, angle and direction in which it pushes the box*").

(C) *Learning about Environment:* Inferred the explicit goal of learning about the environment. For both tasks, this included learning the overall layout (e.g. "*Map the area out in more detail*") and about individual components (e.g. "*Learn what constitutes furniture*"). For the MT, this included limit testing (e.g. "*how far its arm can extend*") and understanding possible movements (e.g. "*how the box moves on the table*").

(D) *Lack of Goals:* Statements in which observers explicitly inferred a Lack of goals (e.g. "*I'm not sure this robot has any idea of a goal in mind*").

9 statements were classified as *Unclear*.

Knowledge: For observer inferences regarding agent knowledge a similar pattern of four clusters emerged.

(A) *Outcome Knowledge:* Understanding of overall objectives and task including specific goals within the environment. For both tasks this included statements of overall purpose (e.g. "*It knows that it needs to pick up the dirt and to get out*", and action consequences (e.g. "*it's dangerous to go near furniture in case it breaks.*").

(B) *Procedural Knowledge:* Inferences describing robot understanding of task execution, i.e. on how to move (e.g. "*Being able to stretch*"), implement behavior (e.g. "*It knows that it needs more deliberate and softer pushes and to stay down on the table more*") and how to solve the task (e.g. "*There's knowledge here in that it learns to do the dirt first before it heads for the exit*").

(C) *Knowledge about the Environment:* Inferences describing robot understanding of its environment. For both settings, this included the "*knowledge of the layout of the room*", certain elements (e.g. "*it has learned what furniture is*") and their location (e.g. "*location of the target and destination of the target*") and for the MT, its boundaries (e.g. "*using its knowledge on boundaries*").

(D) *Lack of Knowledge:* Statements which explicitly attested a lack of knowledge (e.g. "*The robot currently does not seem to have enough knowledge*").

A total of 12 statements were classified as *Unclear*.

Decision Making (DM): Observer inferences regarding robot DM could be clustered into four categories.

(A) *Undirected DM:* Inferences in both settings about DM processes, described as undirected (e.g. "*By trying to move about*"), random (e.g. "*mostly moves at random*"), explorative (e.g. "*The robot learns completely through discovery in this video*") or trial-and-error (e.g. "*It is making its decision through trial and error*").

(B) *Experience-based DM:* Inferences in both settings describing DM that repeats previous behavior or relies on previous experiences. For the NT, a strong focus on past mistakes seemed to prevail (e.g. "*having experienced many mistakes i.e wrong moves, breaking vases*"), while for the MT, it included more general inferences (e.g. "*by following what it has learned overtime*"). For both tasks, repetition (e.g. "*by deciding to repeat movements that have been successful*")

TABLE I

FOUR COMMON INFERENCE THEMES AND THEIR RESPECTIVE SUBCLUSTERS

Goals	Knowledge	Decision Making	Learning Mechanisms
Outcome-related Goals	Outcome Knowledge	Undirected DM	Learning by Exploring
Execution-related Goals	Procedural Knowledge	Experience-based DM	Learning by Feedback
Learning (about Environment)	Knowledge about Environment	Expected Outcome-based DM	Learning by Reasoning
Lack of Goals	Lack of Knowledge	Absence of DM	Absence of Learning

and sensing ("*The robot is making its decisions based seeing the object there*") were inferred to be the basis of DM.

(C) *Expected Outcome-based DM:* Inferences that assume DM to be based on expected outcomes. For both settings, this included targeting subgoals (e.g. "*Once it cleans both spills, it makes the decision to go to the exit as quickly as possible.*"), systematic exploration (e.g. "*systematically learning the shape of the room and the objects within*") or planning expected outcomes (e.g. "*Calculating the right path where the dirt is"*)

(D) *Absence of DM:* Explicitly inferred absence of DM (e.g. "*I don't think the robot is making decisions*").

A total of ten statements were classified as *Unclear*.

Learning Mechanisms (LM): Observer inferences regarding robot LM could be clustered into 4 categories.

(A) *Learning by Exploring:* Inferences that assume the robot to be learning by exploring. This includes learning by trial-and-error and exploration behavior. It indicates a specific focus on learning through new experiences. For both settings, statements such as "*It learns through trial and error*", "*By exploring around*" and "*The robot learns by trying different patterns*" appeared.

(B) *Learning by Feedback:* Inferences that assume the learning from feedback, including general feedback (e.g. "*learning from the decisions it has been making throughout the whole learning process*"), reward and punishment ("*Via positive and negative feedback*"), but also learning by repeating successful patterns ("*by repeating its moves*"), practice (e.g. "*the robot learns by practicing*") and learning from assumed sensory feedback (e.g. "*it learns by perception*").

(C) *Learning by Reasoning:* Inferences that assume the learning from processes of higher reasoning (e.g. "*tries to detect the furniture which it breaks to understand where it is*"), internal representation (e.g. "*remembering where the vases, patches and exit are located*"), abstraction (e.g. "*it split the room into four quadrants and went from right to left learning where the obstacles were*") or from pieces of memory (e.g. "*storing the data of everything learnt*").

(D) *Absence of Learning:* Statements in which observers explicitly denied the robot to engage in learning (e.g. "*I see no evidence of learning here*").

Additionally, in 8 cases observers explicitly attributed Machine Learning or RL mechanisms to the robot, 10 statements were classified as *Unclear*.

C. Discussion

The experimental paradigm in this study was designed to collect data in a direct way, unbiased by interaction and

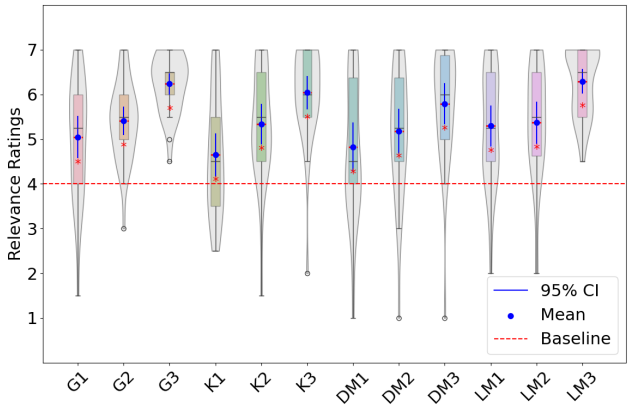

Fig. 3. Relevance ratings: G,K,DM,LM across tasks at 3 time points; Red line: a neutral rating of 4; red stars: significantly different from 4

resulting expectations. Across both task domains (navigation and manipulation) and underlying RL architectures (tabular and DDPG), observers consistently differentiated their ratings across the four primary themes and their respective subclusters (see Fig. 4), further supporting the robustness of the experimental paradigm for eliciting observer inferences. The finding that all four inference themes were rated as highly relevant to the robots' learning process, with perceived relevance generally increasing across learning stages (see Fig. 3) provides empirical support for the thematic structure of Goals, Knowledge, Decision Making and Learning Mechanisms found across both experiments as not merely conceptual abstractions but as subjectively important dimensions of human inference. While individual aspects have been reported before [27], this work offers a systematic framework of observer inferences about RL processes that indicates an integrated representation of robot learners. The first experiment suggested that observers distinguish between what the agent intends to do and what it knows. Notably, they reflect a meaningful distinction between structural assumptions (G and K) and procedural ones (DM and LM). However, the larger study revealed considerable overlap: observers made both task-related and execution-based inferences within each theme, attributing to the robot a detailed understanding of its environment. Crucially, our granular analysis demonstrates that this internal architecture is neither rigid nor static. Responses were differentiated enough to identify at least four conceptual subclusters within each theme. The distribution of these specific subclusters over time illustrates a fluid temporal dynamic within this inference framework. For instance, early-stage inferences regarding a 'Lack of Knowledge' (K-D) or 'Learning about the Environment' (G-C) plummet to near extinction by the

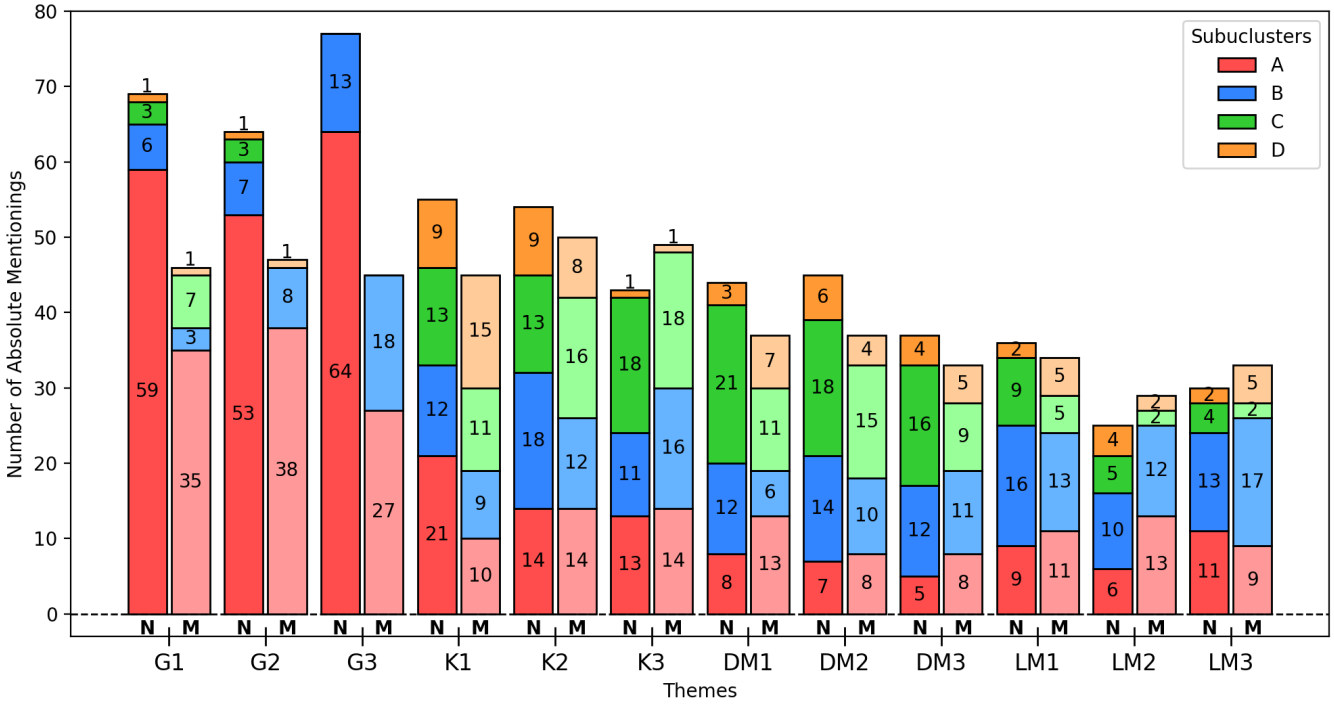

Fig. 4. Themes+Subclusters across time split by task. Letters correspond to subclusters within themes, see IV-B

final observation phase as the agent's policy stabilizes (see Fig. 4). Rather than viewing these categories as distinct, static computational modules, observers continuously adapt and re-weight these interconnected components as the interaction unfolds. Ultimately, these shifting dimensions and granular subclusters together provide empirical evidence of a coherent inferential framework that human teachers naturally leverage to flexibly construct and iteratively update their mental models of learning robots, consistent with prior work in HRI [6], [7]. Notably, this framework strongly aligns with social learning theories in human-human settings, which suggest that the inferential frameworks humans natively employ to decode human minds [14], [13] are fundamentally projected onto artificial agents. However, because this framework relies heavily on anthropomorphized categories as an explanatory bridge, it underlines significant interactive vulnerabilities. Whenever these fluid human inferences misalign with a robot's actual formalized state, they risk triggering critical expectation gaps and counterproductive pedagogical interventions [8]. This highlights the importance of designing systems that anticipate and adapt to such a human inference framework and the subsequent assumptions, laying the groundwork for architectures that actively engage with the human's evolving mental model.

V. Limitations and Future Work

While the thematic analysis followed a structured, reflexive method, future psychometric validation could incorporate inter-rater reliability to further strengthen the framework. While exploratory analyses across tracked demographics (including parenting and pet ownership) revealed no major differences in theme salience, the male-skewed sample (76%) remains a key limitation, given known gender differences in anthropomorphization tendencies. To address the main RQ of what teachers infer from observing robot learning behavior, this work found four common inference themes across tasks and algorithms. Rather than a static taxonomy, researchers should interpret this as an empirically-founded, dynamic interpretive framework of human interpretation. While this controlled, observation-based paradigm serves as a critical precursor to live HRI, a key next step is to validate this framework in quantitative, more complex, real-world settings where observer inferences interact with task realism, embodiment, and time constraints, to examine how these themes shift in active teaching settings where expectations about how robots receive and use teaching signals come into play. In practical terms, these findings enable a human-first approach to Explainable HRI. Rather than deriving explanations only from the algorithmic architecture [10], [28], the framework allows modeling observer interpretation for real-time perspective alignment [29], predict and interpret communication breakdowns, and generate targeted hypotheses for complex, interactive robotic teaching settings. Further, the identified subclusters suggest specific areas where users form strong, but potentially inaccurate assumptions. Designers can utilize these insights to engineer transparency mechanisms around user-inferred concepts to address such mental model mismatches [6], [7]. For instance, designing interfaces that explicitly clarify the distinction between experience-based and outcome-based decision making can proactively reduce observer misconceptions and align robot behavior along observer inferences [30].

VI. Conclusion

This work investigated how humans infer robot learning processes through observation alone. To this end, a novel experimental paradigm was introduced and applied across two

tasks and algorithms. The results revealed a coherent framework of observer inferences structured around four common inference themes, with a detailed structure of interrelated subclusters, along which observer inferences about robot learning processes are organized: Goals, Knowledge, Decision Making, and Learning Mechanisms. These findings offer a data-driven foundation for designing more interpretable learning robots. By aligning robot communication and behavior with these human inference patterns, future systems can foster better understanding, reduce misalignment, and enable more effective collaboration between human teachers and robot learners. Taken together, this work marks a step toward more human-centered systems and contributes to the broader vision of Hybrid Intelligence, combining human and machine intelligence to solve tasks neither can tackle alone.

VII. Acknowledgements

This research is sponsored by the Hybrid Intelligence project, grant number 024.004.022. Special thanks to Jonne Goedhart, Felix Kleuker, David Hyland, Tom Kouwenhoven, Anna Lea Reinwarth, Merle Reimann, Saloni Singh and Kevin Godin-Dubois for their support. Generative AI tools were utilized solely for proofreading and language polishing.

References

[1] Z. Akata, D. Balliet, M. De Rijke, F. Dignum, V. Dignum, G. Eiben, A. Fokkens, D. Grossi, K. Hindriks, H. Hoos, *et al.*, "A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence," *Computer*, vol. 53, no. 8, pp. 18–28, 2020.

[2] K. Baraka, I. Idrees, T. Kessler Faulkner, E. Biyik, S. Booth, M. Chetouani, D. H. Grollman, A. Saran, E. Senft, S. Tulli, A.-L. Vollmer, A. Andriella, H. Beierling, T. Horter, J. Kober, I. Sheidlower, M. E. Taylor, S. Van Waveren, and X. Xiao, "Human-Interactive Robot Learning: Definition, Challenges, and Recommendations," *ACM Transactions on Human-Robot Interaction*, vol. 15, no. 2, pp. 1–31, Mar. 2026. [Online]. Available: https://dl.acm.org/doi/10.1145/3779297

[3] C. Celemin, R. Pérez-Dattari, E. Chisari, G. Franzese, L. de Souza Rosa, R. Prakash, Z. Ajanović, M. Ferraz, A. Valada, J. Kober, *et al.*, "Interactive imitation learning in robotics: A survey," *Foundations and Trends® in Robotics*, vol. 10, no. 1-2, pp. 1–197, 2022.

[4] C. Arzate Cruz and T. Igarashi, "A survey on interactive reinforcement learning: Design principles and open challenges," in *Proceedings of the 2020 ACM designing interactive systems conference*, 2020, pp. 1195–1209.

[5] C. O. Retzlaff, S. Das, C. Wayllace, P. Mousavi, M. Afshari, T. Yang, A. Saranti, A. Angerschmid, M. E. Taylor, and A. Holzinger, "Human-in-the-loop reinforcement learning: A survey and position on requirements, challenges, and opportunities," *Journal of Artificial Intelligence Research*, vol. 79, pp. 359–415, 2024.

[6] G. Bansal, B. Nushi, E. Kamar, W. S. Lasecki, D. S. Weld, and E. Horvitz, "Beyond accuracy: The role of mental models in human-ai team performance," in *Proceedings of the AAAI conference on human computation and crowdsourcing*, vol. 7, 2019, pp. 2–11.

[7] P. Richter, H. Wersing, and A.-L. Vollmer, "Reducing mental model mismatch with intention-based feedback in human-robot teaching," in *Proceedings of the 12th International Conference on Human-Agent Interaction*, 2024, pp. 399–401.

[8] M. K. Ho, F. Cushman, M. L. Littman, and J. L. Austerweil, "People teach with rewards and punishments as communication, not reinforcements." *Journal of Experimental Psychology: General*, vol. 148, no. 3, p. 520, 2019.

[9] C. Grislain, H. Caselles-Dupré, O. Sigaud, and M. Chetouani, "Utility-based adaptive teaching strategies using bayesian theory of mind," *arXiv preprint arXiv:2309.17275*, 2023.

[10] S. Habibian, A. A. Valdivia, L. H. Blumenschein, and D. P. Losey, "A review of communicating robot learning during human-robot interaction," *arXiv preprint arXiv:2312.00948*, 2023.

[11] D. J. Hejna III and D. Sadigh, "Few-shot preference learning for human-in-the-loop rl," in *Conference on Robot Learning*. PMLR, 2023, pp. 2014–2025.

[12] J. MacGlashan, M. K. Ho, R. Loftin, B. Peng, G. Wang, D. L. Roberts, M. E. Taylor, and M. L. Littman, "Interactive learning from policy-dependent human feedback," in *International conference on machine learning*. PMLR, 2017, pp. 2285–2294.

[13] A. Sarin, M. K. Ho, J. W. Martin, and F. A. Cushman, "Punishment is organized around principles of communicative inference," *Cognition*, vol. 208, p. 104544, 2021.

[14] H. Gweon, "Inferential social learning: Cognitive foundations of human social learning and teaching," *Trends in cognitive sciences*, vol. 25, no. 10, pp. 896–910, 2021.

[15] B. Glaser and A. Strauss, *Discovery of grounded theory: Strategies for qualitative research*. Routledge, 2017.

[16] D. Silverman, "How was it for you? the interview society and the irresistible rise of the (poorly analyzed) interview," *Qualitative research*, vol. 17, no. 2, pp. 144–158, 2017.

[17] S. Oh and B. Wildemuth, "Think-aloud protocols," *Applications of social research methods to questions in information and library science*, pp. 178–188, 2009.

[18] T. Boren and J. Ramey, "Thinking aloud: Reconciling theory and practice," *IEEE transactions on professional communication*, vol. 43, no. 3, pp. 261–278, 2000.

[19] G. Brockman, "Openai gym," *arXiv preprint arXiv:1606.01540*, 2016.

[20] V. Braun and V. Clarke, "Thematic analysis: A practical guide," 2021. [Online]. Available: https://www.torrossa.com/it/resources/an/5282292

[21] D. Immel, B. Hilpert, P. Schwarz, A. Hein, P. Gebhard, S. Barton, R. Hurlemann, *et al.*, "Patients' and health care professionals' expectations of virtual therapeutic agents in outpatient aftercare: Qualitative survey study," *JMIR Formative Research*, vol. 9, no. 1, p. e59527, 2025.

[22] B. Hilpert, P. Gebhard, and T. Schneeberger, "Employing virtual agents for building trust in driving automation: A qualitative pilot study," in *Workshop on Robo-Identity: Artificial identity and multi embodiment at the 16th International Conference on Human-Robot Interaction (HRI'21)*, 2021.

[23] T. Schneeberger, M. Scholtes, B. Hilpert, M. Langer, and P. Gebhard, "Can social agents elicit shame as humans do?" in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2019, pp. 164–170.

[24] B. Hilpert, K. Baraka, and J. Broekens, "Teaching is a process: The TOSS framework for modeling human teaching decisions in human-interactive robot learning," in *2026 35th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2026.

[25] J. Wessler, T. Schneeberger, B. Hilpert, A. Alles, and P. Gebhard, "Empirical research in affective computing: An analysis of research practices and recommendations," in *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2021, pp. 1–8.

[26] Q. Gallouédec, N. Cazin, E. Dellandréa, and L. Chen, "panda-gym: Open-Source Goal-Conditioned Environments for Robotic Learning," *4th Robot Learning Workshop: Self-Supervised and Lifelong Learning at NeurIPS*, 2021.

[27] H. Kopecka, J. Such, and M. Luck, "The role of perception, acceptance, and cognition in the usefulness of robot explanations," in *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 2024, pp. 7868–7876.

[28] M. Hou, K. Hindriks, A. Eiben, and K. Baraka, "" give me an example like this": Episodic active reinforcement learning from demonstrations," *arXiv preprint arXiv:2406.03069*, 2024.

[29] R. Van Der Meulen, R. Verbrugge, and M. Van Duijn, "Common ground provides a mental shortcut in agent-agent interaction," in *3rd International Conference on Hybrid Human-Artificial Intelligence, HHAI 2024*. IOS Press, 2024, pp. 281–290.

[30] B. Hilpert, "Closing the teacher-learner loop: The role of affective signals in interactive rl," in *2024 12th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 2024, pp. 97–101.