Proceedings of the Seventh AAAI/ACM Conference on AI, Ethics, and Society (AIES 2024)

PPS: Personalized Policy Summarization for Explaining Sequential Behavior of Autonomous Agents

Peizhu Qian,[*] Harrison Huang,[*] and Vaibhav Unhelkar

Rice University, Houston, Texas, USA
pqian@rice.edu, hhuang@rice.edu, vaibhav.unhelkar@rice.edu

Abstract

AI-enabled agents designed to assist humans are gaining traction in a variety of domains such as healthcare and disaster response. It is evident that, as we move forward, these agents will play increasingly vital roles in our lives. To realize this future successfully and mitigate its unintended consequences, it is imperative that humans have a clear understanding of the agents that they work with. Policy summarization methods help facilitate this understanding by showcasing key examples of agent behaviors to their human users. Yet, existing methods produce "one-size-fits-all" summaries for a generic audience ahead of time. Drawing inspiration from research in pedagogy, we posit that personalized policy summaries can more effectively enhance user understanding. To evaluate this hypothesis, this paper presents and benchmarks a novel technique: **P**ersonalized **P**olicy **S**ummarization (PPS). PPS discerns a user's mental model of the agent through a series of algorithmically generated questions and crafts customized policy summaries to enhance user understanding. Unlike existing methods, PPS actively engages with users to gauge their comprehension of the agent behavior, subsequently generating tailored explanations on the fly. Through a combination of numerical and human subject experiments, we confirm the utility of this personalized approach to explainable AI.

1 Introduction

As artificial intelligent (AI)-enabled agents become increasingly capable, they are playing an ever-expanding role in various facets of our lives. From recommender systems to robots, the ability of these AI agents to autonomously make decisions can greatly benefit humans. However, agent decisions might not align with user expectations in practice, leading to confusion, reluctance to use, or even unintended consequences (Parasuraman and Riley 1997; de Graaf, Allouch, and Klamer 2015; Amodei et al. 2016).

Explainable AI methods aim to address this gap by offering interpretable explanations of an agent's behavior, even if its internal workings are complex (Sakai and Nagai 2022; Chakraborti, Sreedharan, and Kambhampati 2020; Tjoa and Guan 2020; Halilovic and Lindner 2023; Rong et al. 2023). These explanations are especially crucial in high-stakes domains like healthcare and disaster response (Seo et al. 2021; Orlov-Savko et al. 2024). In such domains, users must be adequately trained before agent deployment. It is vital that they are able to predict agent behavior in both familiar and new situations to ensure safe and responsible use of AI.

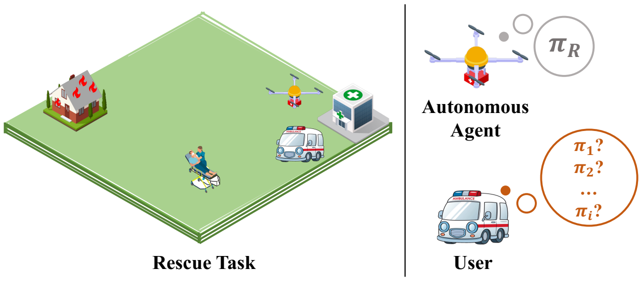

Figure 1: Motivating Scenario: A first responder wants to deploy the autonomous rescue robot to assist in disaster response. However, she is unsure how the robot behaves. To effectively deploy the robot, the first responder must be able to predict robot behavior in both familiar and new situations.

An emerging paradigm for this training involves showing users key demonstrations of agent behavior. **Policy Summarization** algorithms provide a principled approach to select these key demonstrations based on the agent's behavioral policy (Zhan et al. 2014; Amir and Amir 2018; Huang et al. 2019; Lee, Admoni, and Simmons 2021). While these algorithms lay the essential groundwork, they target a general audience and generate "one-size-fits-all" summaries. In absence of personalization, such generic summaries are not as effective as they could be.

To address this research gap, we have made a series of contributions to realize personalization in policy summarization. In (Qian and Unhelkar 2022), we present a policy summarization approach that leverages a cognitive model of the user to algorithmically generate policy summaries. An interactive user interface allows users to customize these summaries, which are communicated to the user via a virtual environment. In (Qian and Unhelkar 2024), we extend our work to consider multiple communication modalities, physical robots, and out-of-distribution scenarios. As the third work in the series, this paper aims to computationally generate user-specific policy summaries by identifying users' assumptions about the agent behavior through interaction.

[*] These authors contributed equally.

1.1 Motivating Scenario

To further motivate the problem setting, consider a scenario of human-robot collaboration in disaster response (Qian and Unhelkar 2022). As shown in Figure 1, the scenario models a first responder who has access to a robot drone. This drone is adept at aiding the responder in specific tasks, such as extinguishing fires or distributing first aid kits. Here, autonomy means the drone determines the sequence of actions to complete the assigned tasks using its autonomous policy, and carries out these actions autonomously. However, the human first responder needs to decide which tasks to assign to the robot drone.

For this human-robot collaboration to be seamless, especially during time-critical emergency scenarios, it is paramount that the first responder understands the robot's capabilities and limitations as well as its decision-making processes. This is where policy summarization techniques step in. Their aim is to offer users a clear understanding of robot behavior before the actual deployment of the robot. With an accurate mental model of the robot's behavior, the responder can anticipate its actions, allowing for a more efficient human-robot collaboration and effective disaster response.

1.2 Scope

In this work, we focus on sequential tasks where a human user has to work alongside an autonomous robot or AI agent.[1] The agent performs components of the task autonomously, with the goal of assisting the human. We limit our scope to tasks that can be modeled as a Markov decision process (MDP) (Puterman 2014). We consider the standard MDP formalism defined by the tuple $M \doteq (S, A, T, R, \gamma)$, where $s \in S$ is the set of states; $a \in A$ is the set of actions available to the agent; $T(s'|s, a)$ is the Markovian state transition probabilities; $R(s, a)$ the reward function; and γ the discount factor.

Agent behavior in this task is specified by the deterministic policy, $\pi_R(a|s)$. We do not make any assumptions about the optimality or source of this policy; e.g., the policy can be generated through planning, reinforcement learning, or even rule-based methods. While the robot has access to this policy, the human user does not. Instead, the human maintains a mental model of robot behavior (Mathieu et al. 2000; Johnson-Laird 2004). Following Bayesian Theory of Mind (Baker, Saxe, and Tenenbaum 2011), we consider that the human maintains a set of candidate models Π regarding possible robot behaviors.

1.3 Contributions

Informed by pedagogical research, we hypothesize that personalized policy summaries can more effectively bolster user understanding (Truong 2016; Miller et al. 2006; Hsieh and Chen 2016; Lin et al. 2013). To test this hypothesis, this paper offers three main contributions. **1) Estimating User Belief:** We introduce an active querying algorithm to deduce a user's prior belief about agent behavior. **2) Personalized Policy Summarization:** Using the querying algorithm as a sub-routine, we adapt a recent policy summarization algorithm to develop PPS – a technique that provides personalized policy summaries (Qian and Unhelkar 2022). Unlike existing methods that pre-generate summaries, PPS engages with users to gauge their comprehension of the robot and then generates tailored explanations on the fly. **3) Comprehensive Evaluation:** We rigorously test PPS through a series of simulation experiments, supplemented by a randomized controlled study with 30 human participants. Our evaluations shed light on the utility of personalized explanations and motivate future directions for explaining behavior of autonomous agents.

2 Related Work

Our work is grounded in existing research on policy summarization, personalized learning, and mental models. Before describing our approach, we briefly review these areas.

Policy Summarization Recognizing the importance of enhancing user understanding of AI agents, several policy summarization techniques have been proposed in recent years (Zhan et al. 2014; Amir and Amir 2018; Huang et al. 2019; Lee, Admoni, and Simmons 2021; Qian and Unhelkar 2022, 2024). The goal of these techniques is to maximize users' ability to predict robot action in unseen scenarios by providing summaries of robot behavior. Following the "explanation-by-example" paradigm, these summaries are given as a set of examples of robot behavior; we refer to each example as an explanation of robot behavior. To select the most informative examples, some methods use information inherent in the robot policy such as its Q-values or Shannon entropy (Amir and Amir 2018; Zhan et al. 2014; Watkins et al. 2021; Huang et al. 2019). Others simulate users' mental model of the robot and select the examples that can maximally improve this mental model (Huang et al. 2018; Qian and Unhelkar 2022; Lee, Admoni, and Simmons 2021; Tabrez, Agrawal, and Hayes 2019). The selected examples are typically presented via a user interface as (s, a)-trajectories to the user. These foundational works, however, put relatively little emphasis on individual differences between users. Despite the consideration of mental models in more recent work, most existing techniques generate identical policy summaries for all users. To our knowledge, only the work of Qian and Unhelkar (2022, 2024) allows for tailored policy summaries. However, in their work, the onus is on the user to design customized summaries through an interactive user interface; the algorithmically generated summaries are not personalized. Recognizing this relative lack of emphasis on individual differences between users, this work introduces personalization in policy summarization.

Personalized Learning Our focus on personalized policy summaries is informed by research in pedagogy. Several works indicate that providing the same content to students with different backgrounds is inadequate, and promote personalized learning to be the solution for meeting individuals' needs and prior experience (Truong 2016; Miller et al. 2006; Hsieh and Chen 2016; Lin et al. 2013). In human-to-human settings (e.g., human tutoring), personalized learning can be an efficient approach to improve learning (Gómez

[1] We use the terms robot and AI agent interchangeably.

et al. 2014). In computer-based tutoring, the research field of Intelligent Tutoring System (ITS) shows that personalized learning enhances students' performance (Ghali et al. 2018; Akyuz 2020; Foshee, Elliott, and Atkinson 2016) and improves students' engagement and time management (Andersen 2011; Pontual Falcão et al. 2018; Shemshack and Spector 2020) when students receive customized instructions and feedback from a computer teaching program. To track the knowledge of students and tailor their learning experience, many computer-based tutoring systems employ a technique known as Knowledge Tracing (KT). Classical KT algorithms include Bayesian Knowledge Tracing (Corbett and Anderson 1994; Khajah et al. 2014; Lee and Brunskill 2012; Pardos and Heffernan 2010; Yudelson, Koedinger, and Gordon 2013), Factor Analysis Models (Cen, Koedinger, and Junker 2008; Pavlik, Cen, and Koedinger 2009; Wilson et al. 2016), and neural-network-based Deep Knowledge Tracing (Piech et al. 2015; Minn et al. 2018; Xiong et al. 2016; Yeung and Yeung 2018). More recently, work in interpretable machine learning for image classification tasks is also shifting to generating personalized explanations (Yang et al. 2021; Rong et al. 2024). In this work, we integrate ideas from KT to model users' learning processes.

Mental Model Estimation To realize personalized summaries, we explore techniques for estimating human beliefs about robots. Related to this effort, there is a rich literature on estimating human cognitive states (Bethel et al. 2007; Neubauer et al. 2020; Kulic and Croft 2007; Heard, Harriott, and Adams 2018; Qian et al. 2024). For instance, methods have been proposed to estimate and model cognitive variables such as intent, belief, goals, rewards, plans, and policies (Osa et al. 2018; Croft 2003; Van-Horenbeke and Peer 2021; Meneguzzi and Pereira 2021; Qian et al. 2024). These techniques utilize a variety of observations arising from behavior, physiology, self-reports, among others. Among these, active techniques to estimate human rewards and policies from demonstrations are closest to our problem setting (Daniel et al. 2014; Cui and Niekum 2018; Basu et al. 2019; Orlov-Savko et al. 2022; Quintero-Pena et al. 2022; Seo and Unhelkar 2024).

3 Problem Statement

We next formalize the problem setting of interest. The objective of a policy summarization algorithm is to maximize users' ability to predict behavior of a given robot. Recall that robot behavior depends on its policy π_R, which defines the action a selected by the robot in a given state s. For this state s, a user can have different hypotheses regarding robot's action a. We capture these hypotheses using a set of candidate stochastic policies $\Pi = \{\pi_i\}$. A candidate policy has the same state space and action space as π_R, but $\pi_i(s)$ might not be the same as $\pi_R(s)$. There are different approaches to generating the set of candidate policies, such as sampling from π_R, handcrafting, and consulting domain experts. Here, we assume we have access to Π and detail our approach to generating Π in the evaluations section.

One way to represent a user's understanding of the robot policy is by modeling their belief $b(\pi_i) \doteq Pr(\pi_i)$ in candidate policies as a probability distribution. Assuming $\pi_R \in \Pi$, the user's belief in robot policy is denoted as $b(\pi_R)$. Users' prior knowledge and experiences will influence their initial belief in the robot policy and other candidate policies. Borrowing terminology from Bayesian statistics, we refer to this initial belief as user's **prior** belief. The belief will correspond to a uniform distribution if the user does not have any preconceived notions regarding the robot's behavior. After the user receives an explanation e, the user will update her belief to $b(\pi_i|e) \doteq Pr(\pi_i|e)$.[2] This updated belief describes the user's belief in a specific candidate policy π_i after seeing e. Given these notations, we paraphrase the objectives of a personalized policy summarization algorithm in mathematical terms: **estimate** $b(\cdot)$ **and select summaries that maximize** (π_R). Next, we discuss our solution to this problem.

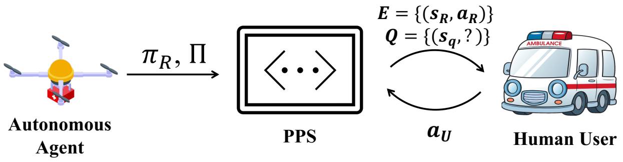

Figure 2: PPS generates personalized explanations E by estimating $b(\pi_R)$ using user's response a_U to questions Q.

4 Personalized Policy Summarization (PPS)

Figure 2 introduces our **P**ersonalized **P**olicy **S**ummarization (PPS) framework, which includes multiple rounds of interactions with the user. Within each round, PPS alternates between asking questions to estimate user belief $b(\cdot)$ and generating a set of explanations $E = (e_1, e_2, \cdots, e_{n_E})$ on the fly to maximize $b(\pi_R|E)$. As noted in Algorithm 1, PPS assumes as input the robot policy π_R and a set of candidate policies Π. Additionally, it includes hyperparameters: n_E, n_Q, n_q, and p_C defined further in this section. The algorithm maintains an estimate of user belief, denoted as $b(\cdot)$, which is initialized as a uniform distribution.

There exist two key challenges in personalizing policy summaries. The first challenge lies in selecting maximally informative questions that estimate user understanding from a small number of (question, answer)-interactions. To address this, PPS includes sub-routines for selecting questions (lines 10-11) and updating the belief estimated based on user responses (line 13). The second challenge lies in selecting n_E number of personalized explanations of robot behavior, denoted as E, that maximize user belief in robot policy. To address this, PPS leverages concepts from knowledge tracing and policy summarization (lines 17-18). The algorithm alternates between addressing these two challenges until user's belief in robot policy reaches 1.[3] We describe the subroutines that address this challenges next.

[2] Recall that, in policy summarization, explanation (e) refers to one example of robot behavior $(s, a{=}\pi_R(s))$ in a chosen state (s).

[3] In practice, the termination condition may depend on real-world constraints such as available time. In our evaluations, we limit the total number of explanations to ensure fair comparisons with baselines and timely completion of experimental tasks.

Algorithm 1: PPS: Personalized Policy Summarization

```
Input: $\pi_R$, robot policy; $\Pi$, set of candidate policies
Initialize: $b(\cdot) = \mathcal{U}$, user belief
Initialize: $p_c = 0.5$, probability of comprehension
Initialize: $n_E$, number of explanations per round
Initialize: $n_Q$, number of quizzes per round
Initialize: $n_q$, number of questions per quiz
repeat
    for $j = 1 : n_Q$ do
        for $k = 1 : n_q$ do
            Generate a question $(s_q, ?)$ using Eq. (1)
            Obtain user response $a_U$ to this question
        end for                          // end of one quiz
        Update belief based on $(s_q, a_U)$ via Eqs. (2)–(4)
        Update probability of comprehension $p_C$
    end for                  // end of one round of quizzes
    for $k = 1 : n_E$ do
        Generate an explanation $e$ using Eq. (5)
        Update belief based on $e$ using Eq. (6)
    end for            // end of one round of explanations
until $b(\pi_R) = 1$
```

4.1 Estimating User Belief

While designing PPS, we explored two approaches to estimating user belief: a direct approach that asks user to describe their belief and an indirect approach that asks user to answer questions regarding robot behavior and then uses the responses to estimate user belief. While seemingly straightforward, we find the direct approach challenging to realize due to multiple reasons. Often users do not have sufficient understanding of robot behavior *a priori* to verbalize their beliefs. For instance, in our human subject experiments, we find textual description of robot objectives alone to be insufficient for correctly informing user beliefs. And, even in cases when they do, it is difficult to accurately translate their understanding to relevant mathematical terms (b, π) in an efficient manner. Nonetheless, direct approaches to estimating user belief remains an interesting avenue for future work.

Selecting the Question Format Thus, in PPS, we opt for asking users questions regarding robot behavior. While operationalizing the indirect approach, one could use a variety of question formats, such as asking users to predict robot behavior, rank different behaviors, among others. In PPS, we focus on questions of the format $(s_q, ?)$, where s_q denotes a state and ? denotes the question "What do you think the robot will do in the state?" For this format, selecting questions correspond to selecting states $s_q \in S$ in which to query the user. User response to these questions is denoted as a_U. An advantage of this question format is that it has the same format as a single explanation (s, a) of the policy summary. As such, this question can be delivered using the same interface and modality used for providing explanations. To facilitate question selection, we define a "quiz" as a set of n_q questions during which PPS does not perform belief updates. PPS designs a quiz, solicits user responses, and updates estimate of user belief at the end of each quiz.

Selecting Informative Questions We now discuss the process of designing a quiz. To enable efficient estimation of user beliefs, PPS aims to select questions that are maximally informative. As detailed next, we pose this as a problem of entropy minimization. We assume that while answering a quiz the user first samples a candidate policy $\pi_h \sim b(\cdot)$ based on its belief. Then, the user answers each question of the quiz using the sampled policy, i.e., $a_{Uk} \sim \pi_h(s_{qk})$, where k denotes the k-th question within a quiz. The questions within a quiz are selected to efficiently infer the index h of the candidate policy π_h selected by the user. In particular, PPS selects s_q that minimizes the conditional entropy of h conditioned on the user response a_U

$$s_q = \arg\min_{s \in S} H(h | a_U; s), \tag{1}$$

where h and a_U are modeled as latent random variables; since, while designing a question, PPS knows neither the user's selected index h nor the user's potential response a_U. The selected question is posed to the user to solicit their response. Subsequent questions of a quiz are selected in a similar fashion based on the user response to previous questions.

Soliciting User Response In our implementation, the selected question is presented to the user via a user interface that includes a visual depiction of the task. Each question is accompanied with the prompt: "What do you think the robot will do in this scenario?", along with a visual representation of the task and robot in the selected state s_q. Our experimental evaluations include tasks with discrete action spaces. For these tasks, the questions are presented as single-select multiple choice questions, where the choices correspond to robot actions. Through this interface, the user needs to input their prediction of robot action for the state s_q.

Updating Belief based on User Response We pose the problem of updating belief based on a user responses as one of Bayesian inference. Once a quiz is complete, PPS derives a composite measurement z corresponding to the inferred value of h based on the user's responses to the quiz:

$$z = \arg\max_{h} \Pr(h | s_{q1}, a_{U1}, \cdots, s_{qn_q}, a_{Un_q}). \tag{2}$$

This measurement z is used to update the estimate of user's belief. The belief is modeled as a categorical random variable $b \sim \text{Dir}(\alpha)$, with a conjugate Dirichlet prior. The conjugate prior enables closed-form computation of the posterior $Pr(b|z)$. We set the concentration parameter $\alpha \doteq (\alpha_1, \cdots, \alpha_{|\Pi|})$ of the Dirichlet distribution as

$$\alpha_i = m \cdot b(\pi_i) + 1, \tag{3}$$

where b is the current estimate of the belief and m is a hyperparameter. This definition ensures that the mode of the Dirichlet prior equals the the current estimate of the belief. Based on simulation trials, we choose a value of $m = 2$ for experiments. Given the prior and composite measurement z computed in Eq. 2, the maximum a posteriori (MAP) estimate of belief is obtained as:

$$b \leftarrow b^{\text{MAP}}(\pi_i | z) = \frac{\mathbb{1}(i = z) - 1 + \alpha_i}{1 - |\Pi| + \sum_i \alpha_i}, \tag{4}$$

which is used as the updated estimate of user belief in order to generate the next quiz or explanation.

4.2 Generating Personalized Explanations

Recall that our overall goal is to maximize user belief in the robot policy $b(\pi_R)$. Hence, PPS generates explanations by solving the optimization problem

$$E_* = \arg\max_E b(\pi_R|E). \tag{5}$$

Since this optimization problem depends on user beliefs, its maxima will inherently generate personalized explanations provided user-specific estimates of prior beliefs are used to compute the objective. PPS includes quiz-based mechanisms for estimating this user-specific prior. In this section, we discuss PPS's approach to solving this optimization problem given user-specific prior beliefs.

Estimating Changes in Belief based on Explanations To solve this optimization problem, we first need a method to compute $b(\pi_R|E)$; that is, changes in user belief in response to receiving explanations. Following (Qian and Unhelkar 2022), we model the user as a Bayesian agent that updates their belief according to the Bayes rule. Informed by prior work in intelligent tutoring and our experience, we argue that to generate effective explanations it is essential to consider human factors that influence learning (Corbett and Anderson 1994; Rong et al. 2023). Hence, additionally, our user model considers one such human factor. Specifically, prior work on policy summarization assumes that a user accurately comprehends a given explanation. However, human learning theory suggests a potential for mismatch between the generated explanation e and its comprehension by the user e_U. We model this mismatch using the scalar term $p_c \doteq Pr(e = e_U)$, referred to as user's probability of comprehension. Incorporating p_c, PPS estimates the updated user belief $b(\pi_i|e)$ upon receiving an explanation e as:

$$b(\pi_i|e) = \Pr(\pi_i|e_U{=}e)\cdot p_c + \Pr(\pi_i|e_U{\neq}e)\cdot(1-p_c), \tag{6}$$

where $\Pr(\pi_i|e_U{=}e)$ and $\Pr(\pi_i|e_U{\neq}e)$ denote user's belief with and without accurate explanation comprehension, respectively. PPS approximates $\Pr(\pi_i|e_U{\neq}e) = b(\pi_i)$ as the user's prior belief. The term $\Pr(\pi_i|e_U{=}e)$ is computed via a Bayesian update in a manner identical to that of (Qian and Unhelkar 2022). More concretely, $Pr(\pi_i|e{=}e_U) = \lambda\pi_i(a|s)Pr(\pi_i)$, where λ is the normalization factor and (s, a) are the states and actions contained in the explanation e. The term p_c is initialized at 0.5 and then updated after each quiz as the percentage of questions that the user has answered correctly so far.

Selecting and Delivering Personalized Explanations In theory, given the aforementioned approach to compute the objective function $b(\pi_R|E)$, a variety of methods can be used to solve Eq. 5. For instance, (Qian and Unhelkar 2022) pose this as a sequential decision-making problem and solve it using Monte Carlo Tree Search. In PPS, we approximate the optima using greedy search. This approach while approximate enables quick computation, which is essential for generating personalized explanations on the fly. Through a suite of experiments, we confirm that this quick but approximate approach performs satisfactorily in practice. Upon generating an explanation $e{=}(s, a)$, PPS generates visualization of robot trajectories that begin in the state (s) and show these animations to the users through a graphical user interface.

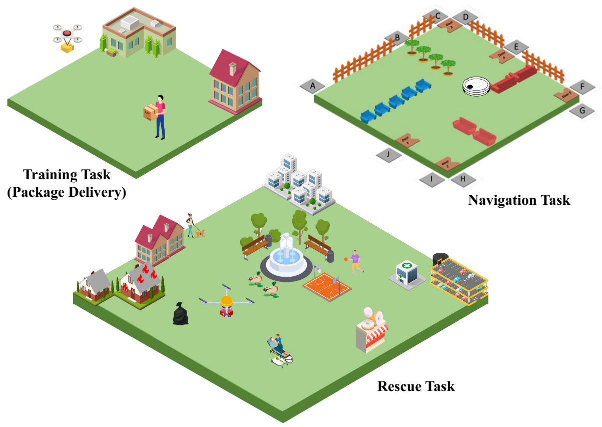

Figure 3: We design three domains to conduct thorough evaluation of our framework: Training, Navigation, and Rescue.

5 Model Evaluations with Simulated Users

To evaluate our framework, we perform thorough experiments to test both its belief estimation and explanation generation components. In this section, we describe the experiments we run with simulated users. We emphasize the importance of testing with simulated users, which enables

1) computing quantitative metrics (due to access to simulated users' mental models) that are not possible in human subject experiments;
2) resource-effective evaluation in different settings and with a variety of simulated users.

5.1 Experimental Domains

We designed two domains to be used both in simulation and in experiments with human users (Figure 3). In addition, we designed a training domain to familiarize users with the experiment protocol and the user interface (UI) used during human experiments.

Domain 0: Training Task In this domain, a drone is tasked with delivering a package to green houses. There can also be red houses on the map but the drone avoids them. In a given scenario, either type of houses can be present or absent (representing different neighborhoods based on the number of houses on the map); as such, if the green house does not exist, the drone will select the action of staying still unless it is over the red house, in which case it will move away. This domain is modeled as a 3×3 grid-world with the houses in fixed locations for a state space of size $3 \times 3 \times 2^2 = 36$. The robot's state consists of its (x, y) position and two Boolean variables indicating whether the green house and red house exist in the current neighborhood. The drone also prioritizes movement in the following order: stay still, move east, move north, move west, and move south, for an action space of size 5. If the drone must move up and to the right to reach the green house, it will always select move right over move up. We hand-design 4 candidate policies for this training domain, modeling the drone's behavior if its goal were the green house, red house, both, and neither.

Domain 1: Navigation Task A robot navigates in a room containing 5 doors to dock at a charging station at the closest odd-numbered door. The robot may not pass through any of the furniture or plants, and must enter each door from a particular side. Furthermore, the robot must orient itself to enter the door while facing forward (i.e., the robot cannot enter a door facing backward or sideways). This discrete environment is modeled as a 10×10 grid with state represented as the (x, y) position of the robot in addition to its orientation $\theta \in \{0°, 90°, 180°, 270°\}$, where $\theta = 0°$ means the robot is facing east and each $90°$ increment turns the robot counterclockwise by $90°$. At each time step, the robot may either stay still, move forward, move backward, rotate left, or rotate right, in order of priority. In other words, if the rescue robot must move forward and turn left to reach its goal, it will always prioritize moving forward over rotating left. The task models includes $10 \times 10 \times 4 = 400$ states and 5 actions. We approximate candidate user policies by training $2^5 = 32$ policies, each representing a possible combination of doors for the robot to navigate to.

Domain 2: Rescue Task A robot is deployed to perform several rescue tasks in a city. This discrete environment is modeled as a 10×10 grid with up to four tasks that the robot performs in the following order: putting out a fire, removing debris/trash, dropping off a first-aid kit for a patient, and picking up medicine from a hospital. The location of each task is fixed but, depending on the scenario, some tasks may be absent. For example, there may not be trash/debris, and as such, the rescue robot will put out the fire then move to dropping off the first-aid kit for the patient. There is also a 2×2 body of water that the rescue robot avoids. In this domain, the robot's state space is its position and a Boolean variable indicating whether it has performed each of the tasks, which is represented as $(x, y, s_1, s_2, s_3, s_4)$, where (x, y) is the robot's current position and $s_i \in \{0, 1\}$ for $i \in \{1, 2, 3, 4\}$. At each time step, the robot can either—in order of priority—stay still, move east, move north, move south, or move west. For example, if the robot needs to move up and to the right, the robot will prioritize moving right over moving upwards. The environment is modeled using $10 \times 10 \times 2^4 = 1600$ states and 5 actions. We approximate $4! = 24$ candidate user policies, each representing a different order of performing all the tasks.

5.2 Simulated Users

We evaluate our algorithm in the navigation and rescue domains with different designs of simulated users. To arrive at a simulated user, we need three components: their prior belief, a mechanism to simulate their response to questions posed by PPS, and a mechanism to update their belief given an explanation generated by PPS. We describe these three components for simulating users next.

Prior Belief To ensure that our algorithm is robust to different types of users, we create 6 simulated users with different prior beliefs for each domain. To arrive at these six priors, we first consider two categories of simulated users: those who have high belief in one candidate policy and those who cannot differentiate between two candidate policies. For all simulated users with high belief in a single candidate policy, π_i, we set $b(\pi_i) = 0.7$ and normalize the belief equally across the remaining candidate policies. This simulated user represents users with strong preconceived notions of what they expect the robot to do. For simulated users with high belief in two candidate policies, denoted as π_{i_1}, π_{i_2}, we set $b(\pi_{i_1}) = b(\pi_{i_2}) = 0.35$ and normalize the belief equally across the other candidate policies. From there, we define the notion of **policy similarity**: the number of states in which a policy takes the same action as another policy. The higher this value, the more similar the two policies are. The simulated users are named and created as follows:

- **Robot**: This user has a prior belief with a near complete understanding of the robot policy, $b(\pi_R) = 0.7$.
- **Similar**: This user has a high prior belief in a different policy that is highly similar to π_R, denoted as $\pi_{similar}$.
- **Similar 2**: This user has equally high prior belief in two policies: the robot and the highly similar policy, $b(\pi_R) = b(\pi_{similar}) = 0.35$.
- **Dissimilar**: This user has a high prior belief in a policy that is highly dissimilar to π_R, denoted as $\pi_{dissimilar}$.
- **Dissimilar 2**: This user starts with high belief in two policies that are highly dissimilar to the robot policy π_R, denoted as $\pi_{dissimilar}$ and $\pi_{dissimilar-2}$.
- **Random**: This user has high belief in a randomly selected policy that is neither the robot policy nor any of the similar or dissimilar policies.

Question Response We use two mechanisms for simulating user response to questions, titled "Deterministic User" and "Probabilistic User." For a question $(s_q, ?)$, a deterministic user always answers based on the policy over which it has the highest belief; a probabilistic user chooses its answer by based on a policy sampled based on its belief.

Belief Update For updating users' belief after they receive explanations $b(\pi_i|E)$, we use the belief update method described in Section 4.2 for all simulated users.

5.3 Results: Belief Estimation

To evaluate belief estimation performance of PPS, we fix the user belief and ask them to answer questions posed by the algorithm. We observe that PPS is able to infer the user's true prior for all types of simulated users with high accuracy. Results for four of the user types are shown in Figure 4, where each user is asked 50 quizzes. Each quiz is composed of $n_q = 4$ questions of type $(s_q, ?)$. The user answers these questions as per the response policy of "Probabilistic User." For each user, PPS is able to obtain an acceptable estimate of the user's true belief after ≈ 3 quizzes. With additional questions, PPS converges toward the user's true belief. In practice, asking users large number of questions is expensive and can lead to poor user experience; hence, these results are especially promising as they show that PPS is able to estimate users' belief through a small number of quizzes.

5.4 Results: Impact of Personalized Explanation

Next, we compare PPS's overall performance against a baseline algorithm that does not select personalized examples. In

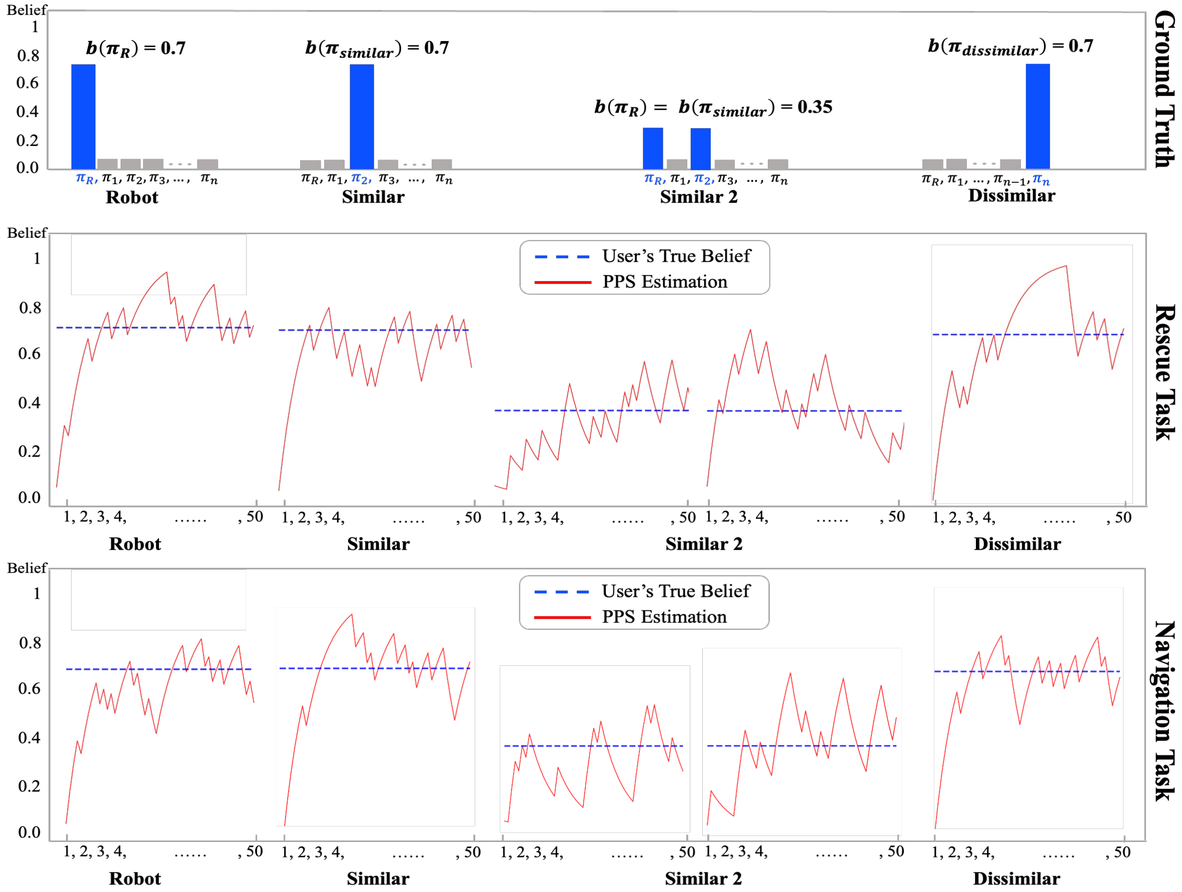

Figure 4: Belief estimation results for simulated users based on user response to PPS-generated questions. (top) Prior belief of four simulated users. (middle and bottom) Belief estimation performance of PPS for the rescue and navigation tasks. For each user, the red curves depict the estimated belief for policy in which that user has the highest belief and the dashed blue line depicts users' true belief for the corresponding policy. x-axis denotes the number of quizzes posed to the user by PPS.

these experiments, both PPS and the baseline algorithm are first deployed to summarize a robot policy, denoted as π_{old}, without estimating a user's belief. Then, PPS uses quizzes to estimate their belief and updates their probability of comprehension. In contrast, the baseline method neither asks questions nor estimates user belief to generate policy summaries, leading to lack of personalization. It also assumes a fixed $p_c = 0.75$. Next, both algorithms are deployed to summarize another robot policy, denoted as π_{new}; this policy update replicates real-life scenarios where a robot goes through a software update causing its behavior to change. In such scenarios, users will need to relearn the updated robot behavior to ensure safe use of AI and achieve seamless collaboration (Bansal et al. 2019a,b).

We compare simulated users' learning curve under each treatment. These experiments confirm that PPS is indeed capable of generating personalized examples due to its ability to estimate user belief and understanding. Figure 5 depicts the learning curves of one probabilistic simulated user. Encouragingly, by leveraging personalization, PPS enables the simulated user to learn faster in both experimental domains.

6 Experiments with Human Users

Inspired by the promising results of our simulation study, we design and conduct two experiments with human users. The experiment protocols are approved by Rice University's IRB. In the preliminary study, we evaluate the belief estimation performance of PPS. In the second study, we evaluate PPS comprehensively with emphasis on the impact of personalized explanations. Figure 6 provides a flowchart summarizing the procedures of the two studies. For both studies, we create a graphical user interface (GUI) to provide users with a visual platform to learn robot behaviors. The GUI first collects informed consent from the participants, then provides a step-by-step tutorial regarding the experiment using the training task, before proceeding to the test domains. A supplementary video demonstration of the GUI is available at https://tiny.cc/pps-experiment-demo.

6.1 Preliminary Study

We recruit 33 participants for the preliminary study using the web-based study platform Prolific, including 16 females, 16 males, and 1 prefer not to say. The participants report

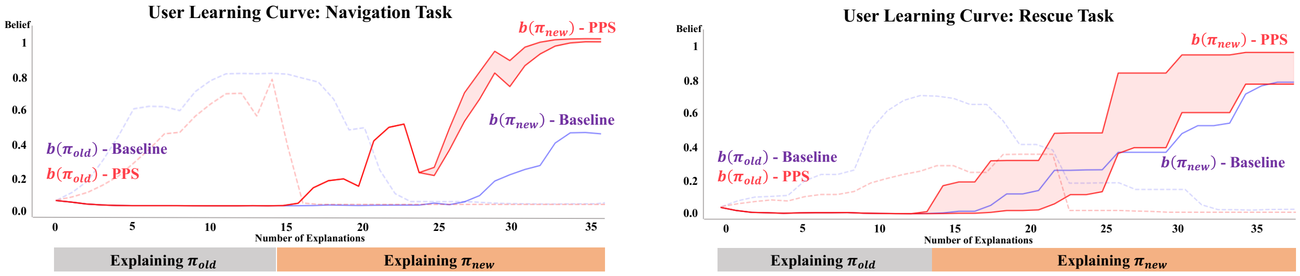

Figure 5: Learning curves of a probabilistic simulated user over five experimental trials. x-axis denotes the number of explanations provided to the user by the policy summarization algorithms: PPS (red) and Baseline (purple). Faint dashed lines denote the simulated user's belief in π_{old}. Bold solid lines denote the simulated user's belief in π_{new}.

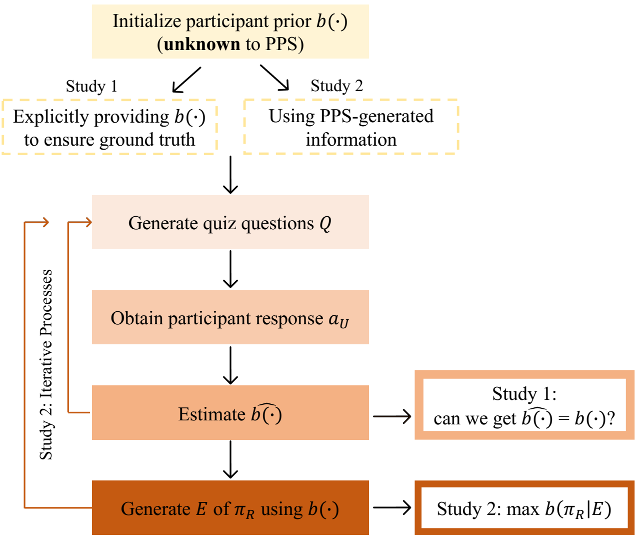
Initialize participant prior $b(\cdot)$ (**unknown** to PPS)
Study 1
Study 2
Explicitly providing $b(\cdot)$ to ensure ground truth
Using PPS-generated information
Generate quiz questions Q
Obtain participant response a_U
Estimate $\widehat{b(\cdot)}$
Generate E of π_R using $b(\cdot)$
Study 1: can we get $\widehat{b(\cdot)} = b(\cdot)$?
Study 2: max $b(\pi_R|E)$
Study 2: Iterative Processes

Figure 6: Overview of the two human experiments to evaluate PPS. The two studies focus on evaluating the belief estimation and personalization capabilities of PPS, respectively.

an average age of 40.89 (SD = 13.40), with the oldest being 76 and the youngest being 21. Each participant is provided with information about robot's high-level objective to inform their prior belief. For example, in the navigation task, the participants are informed that "The robot will go to the closest door but avoid door 4." In the rescue task, we inform the participants that "the robot will first put off the fire, then clean the trash, then pick up medicine from the hospital, and drop off the first-aid kit to the patient." These high-level objectives represent one of the candidate policies in Π and are unknown to PPS. Then the participants receive 3 sets of quizzes. Similar to the simulation experiments, each quiz is composed of 4 questions. PPS then uses user response (a_U) to estimate the participants' belief $b(\cdot)$.

We first discuss influence of textual description of robot's objective on user belief. Although we provide the participants with information to influence their initial belief, answers of some users do not match with the provided information. There could be multiple reasons behind this behavior, such as lack of attention due to experiments being conducted remotely through the web. This could also happen because we only give participants high-level (task-level) objectives and some participants need more detailed (action-level) guidance to provide accurate answers. On average, the online participants answered 5.97 out of 12 questions correctly in navigation and 9.09 out of 12 questions correctly in rescue. **These results further highlight the need of policy summarization algorithms for explaining robot behavior, as we find textual description of robot objectives alone to be insufficient for correctly informing user beliefs.**

With the participants' performance in mind, in the rescue task, PPS successfully identifies 12 out of 33 participants' ground-truth policy. Additionally, for 7 other participants, the algorithm identifies a policy extremely similar to the ground-truth policy (95.81% of the state space have the same actions as the ground-truth policy). In the navigation task, our model only identifies the correct ground-truth policy for 2 participants, but identifies the most similar policy (which is 96.47% similar to the ground-truth policy) for another 12 participants. Overall, similar to the simulation study, **we observe positive trends for belief estimation with human participants; however, PPS can find it challenging to disambiguate between highly similar policies.**

6.2 Participants

Next, we conduct a more comprehensive study to evaluate the benefit of personalized policy summaries. This study is conducted in person as a randomized control trial with two conditions. Similar to simulation experiments, the baseline condition generates user-agnostic explanations and the experimental condition generates personalized explanations using PPS. We recruit 30 participants from Rice University, 15 in each condition. The participants of this study have an average age of 21.13 (SD = 1.80), including 17 females, 12 males, and 1 prefer not to say. We survey their prior experience with AI and robots, and 14 (46.66%) of the participants report no experience with robots, and all participants report some level of experience using AI such as Siri or Alexa. Among the participants who have experience with robots, they report a wide range of experiences from “watching Roombas clean” to having “programmed motion planning algorithms.” When specifically asked about their experience as a developer (not counting being a user), we are surprised to see that more participants have programmed robots but not AI algorithms (8 vs. 2). Last, we survey the participants on their frequency of playing video games: 14 report either never or a few times a year, and the rest report monthly, weekly, or daily engagement with video games.

6.3 Experimental Procedure

We recruit participants by sending emails to various listservs and ask interested participants to email us to schedule an appointment. During the experiment appointment, we first thank the participants and start with a briefing to explain the purpose of the study. Then, the participants log into our GUI with an assigned participant ID to proceed to the consent form. Upon signing the consent form electronically, the participants start a tutorial with the training task to get familiar with the user interface, the study objectives, and the format of the study. Next, they continue to complete the navigation and rescue tasks. Before each task, they see a short video clip to help them understand the graphics that also includes a reminder for them to stay focused. Similar to simulation experiments, participants are provided summaries of two robot policies, first π_{old} and then π_{new}, as detailed next.

Stage 1 For each task, a participant first receives 6 explanations on the first policy π_{old}, followed by 3 sets of quizzes where each quiz consists of 4 questions, to measure how well the participant understands the first policy. The explanations for the first policy are the same across both conditions to ensure users acquire the same prior assumption about the robot. The quizzes are generated interactively and might differ across users based on their responses to previous questions.

Stage 2 Then, participants are informed that “The robot has gone through a software update that might have caused its behavior to change.” Now, the participants’ goal is to figure out the new robot behavior, captured by π_{new}. For simplicity, we refer π_{new} as π_R in the remaining section. In this stage, the participants in the experimental condition receive personalized explanations whereas the baseline group receives user-agnostic explanations. Each participant first receives 3 explanations, then 2 sets of quizzes each consists of 3 questions to evaluate their current belief, then another 3 explanations. Both groups receive the questions to ensure identical procedure between groups but we do not use a_U to generate personalized explanations for the baseline group.

Stage 3 Stage 2 concludes the explanation part of the experiment, and in Stage 3, we ask participants a series of questions to test their understanding of π_R. In this part, we use two types of questions: **action-level** and **task-level** questions. The action-level questions include a graphically represented state s and ask the participants to predict $a = \pi_R(s)$, “What do you think the robot will do in this scenario?”; or a teamwork scenario where the participant is collaborating with the robot and needs to provide an action a_H they would take. The teamwork questions involve both predicting robot action $a_R{=}\pi_R(s)$ and correct understanding of the task in order to determine the suitable action for the human collaborator a_H. With the teamwork questions, we aim to evaluate the benefit of policy summarization methods for not just improving user understanding of robot behavior but also for enhancing human-AI or human-robot teamwork.

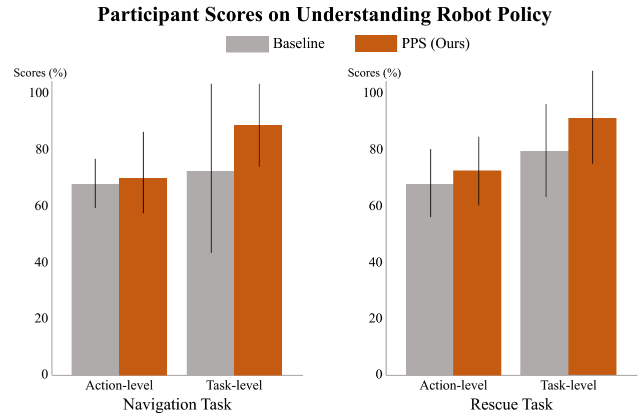

Figure 7: Results of the human experiments. x-axis denotes the type of question, y-axis denotes the participants’ scores $(\%)$. Participants receiving personalized explanations (orange) score higher on both action- and task-level questions than those receiving non-personalized explanations (gray).

The task-level questions involve asking the participants to rank a set of task-level objectives in the order of the robot’s priority. For example, in the rescue task, the participants need to rank “putting off a fire”, “cleaning up trash”, “picking up medicine from the hospital”, and “dropping off the first-aid kit to the patient” in the order they believe the robot is following. We also add incorrect objectives to the suggestion list and the participants are expected to delete them from the list. The participants are further asked to add any other objectives that they observe, such as “avoiding the furniture in the room” in the navigation task and “avoiding the water” in the rescue task. With the help of both action-level and task-level questions, we hope to gain a thorough picture of the participant’s understanding of the robot policy.

Stage 4 The experiment concludes with a post-experiment survey, which asks the participants about their experience, their perception of the role of explanations and quizzes, and their preferred method for understanding robot behavior.

6.4 Results

Next, we analyze the results from the in-person study derived using both the objective and subjective measures.

PPS generates effective personalized explanations that help users understand robot behavior. Figure 7 presents the participant scores when answering action-level and task-level questions. For both domains, we observe that those participants who received personalized summaries generated using PPS score higher for both question types. Among these, the task-level questions indicate a significant improve-

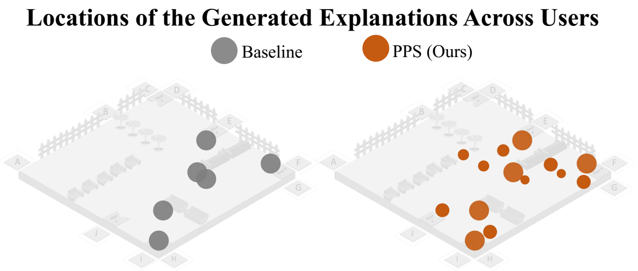

Figure 8: Explanation distribution shows that PPS generates more diverse explanations that differ across users. (left) The gray dots indicate the locations of the explanations generated by Baseline in the navigation task. (right) The orange dots indicate the locations of the explanations generated by PPS across users in the navigation task.

ment from baseline to PPS. In the navigation task, the average scores of task-level questions are 70.77% and 86.15% in baseline and PPS conditions, respectively. In the rescue task, the average scores of task-level questions are 76.92% and 88.46% in baseline and PPS conditions, respectively. To verify that PPS indeed generates different explanations for different users, we visually examine them in Figure 8. This figure depicts the locations corresponding to the generated explanations; more prominent dots indicate a higher frequency of an explanation. The quantitative results coupled with the visualization highlight PPS's ability to generate more diverse explanations that differ across users, thereby leading to improved user understanding of robot behavior.

Participants understand the roles of explanations and questions, and prefer to have both. Participants positively perceive both questions and animated explanations generated by PPS. Recall that we use the questions to estimate participants' belief in order to generate personalized explanations; although for fair evaluations, participants are not provided with the correct answer to the questions, they still find them helpful. One participant writes,

> "The explanations were very helpful to me because they would often show new scenarios where I wouldn't have an idea yet, and they would give helpful insight into what the robot would do in that circumstance. The quiz questions served as more of a sanity check to make sure I had a general idea what the robot would be doing at a given time."

When asked to rate the importance of algorithm-generated explanations and questions for their learning of robot behavior on a scale of $1 - 7$ (from extremely unimportant to extremely important), the animated explanations (M=6.20) and questions (M=5.27) both receive high scores. We further survey the participants on how they would allocate their time between explanations and questions if they had the option to choose between them. We find that 16 (out of 30 participants) would like to spend 75% of the time on animated explanations and 25% of the time on questions; 8 participants prefer 50% on each; and the rest split between only explanations or 25% explanations with 75% questions.

7 Discussion and Concluding Remarks

Policy summarization methods are an important tool to make complex AI systems more transparent to human users, improve human-AI team performance, and reduce unwanted consequences caused by black-box nature of AI systems. In this work, we propose PPS, a policy summarization method that provides personalized explanations by estimating a human user's prior knowledge of a robot's behavior. PPS consists of two parts: first, a belief estimation sub-routine that infers the user's mental model of the robot through active querying; second, a policy summarization algorithm that utilizes Bayesian Theory of Mind to generate customized policy summaries to improve user's understanding of the robot policy. We conduct and report on numerical experiments and user studies to evaluate our approach as well as the role of personalized explanations. Experiment results show PPS can accurately estimate user assumptions and prior beliefs, and the generated summaries effectively improve user understanding of robot policies.

Limitations Our work assumes that users prior assumptions regarding robot behavior can be adequately captured by the set of candidate policies Π and the corresponding belief vector $b(\cdot)$. This assumption directly informs the proposed user model and influences the explanation generated by PPS. The empirical results suggest that this is not a strong assumption; however, it may not always hold in practice. Future work is needed that considers additional user models and compares their relative benefits and limitations for policy summarization. Second, we evaluate our method on discrete domains that are fully observable. An interesting avenue of future research is generating personalized explanations for domains with continuous or partially observable state spaces. Third, our work currently explores only one method of delivering quizzes and explanations, specifically through a graphical user interface. We are interested in investigating other modalities, such as language-based explanations, physical interaction, and augmented reality. Finally, both of our studies have small sample sizes (≈ 30 participants each), and the participants may not be representative of a broader population. Notably, in the second study, all participants are either graduate or undergraduate students.

Implications and Future Avenues As pedagogical research suggests, we believe that providing the same content to users with different backgrounds is not adequate. Users have diverse expectations and assumptions about autonomous systems, necessitating personalized explanations. In this work, however, we have only considered one aspect of personalization in explainable AI: namely, policy summaries. Future XAI research should explore further avenues for personalization, including for modeling users and delivering explanations, to provide more helpful explanations.

Acknowledgments

We thank the anonymous reviewers for their constructive feedback. Peizhu Qian was supported in part by NSF awards #2222876, #2326390, and Rice University funds. Harrison Huang was supported by Rice University funds.

References

Akyuz, Y. 2020. Effects of Intelligent Tutoring Systems (ITS) on Personalized Learning (PL). *Creative Education*, 953–978.

Amir, D.; and Amir, O. 2018. HIGHLIGHTS: Summarizing Agent Behavior to People. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*.

Amodei, D.; Olah, C.; Steinhardt, J.; Christiano, P.; Schulman, J.; and Mané, D. 2016. Concrete problems in AI safety. *arXiv preprint*.

Andersen, M. H. 2011. The World Is My School: Welcome to the Era of Personalized Learning. *The Futurist*, 1–6.

Baker, C.; Saxe, R.; and Tenenbaum, J. 2011. Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the annual meeting of the cognitive science society*, volume 33.

Bansal, G.; Nushi, B.; Kamar, E.; Lasecki, W. S.; Weld, D. S.; and Horvitz, E. 2019a. Beyond accuracy: The role of mental models in human-AI team performance. In *Proceedings of the AAAI conference on human computation and crowdsourcing*, volume 7, 2–11.

Bansal, G.; Nushi, B.; Kamar, E.; Weld, D. S.; Lasecki, W. S.; and Horvitz, E. 2019b. Updates in human-AI teams: Understanding and addressing the performance/compatibility tradeoff. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 2429–2437.

Basu, C.; Bıyık, E.; He, Z.; Singhal, M.; and Sadigh, D. 2019. Active learning of reward dynamics from hierarchical queries. In *Proceedings of International Conference on Intelligent Robots and Systems*, 120–127. IEEE.

Bethel, C. L.; Salomon, K.; Murphy, R. R.; and Burke, J. L. 2007. Survey of psychophysiology measurements applied to human-robot interaction. In *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, 732–737. IEEE.

Cen, H.; Koedinger, K.; and Junker, B. W. 2008. Comparing Two IRT Models for Conjunctive Skills. In *International Conference on Intelligent Tutoring Systems*.

Chakraborti, T.; Sreedharan, S.; and Kambhampati, S. 2020. The Emerging Landscape of Explainable Automated Planning & Decision Making. In *International Joint Conference on Artificial Intelligence*, 4803–4811.

Corbett, A. T.; and Anderson, J. R. 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4: 253–278.

Croft, D. 2003. Estimating intent for human-robot interaction. In *IEEE international conference on advanced robotics*, 810–815. Citeseer.

Cui, Y.; and Niekum, S. 2018. Active reward learning from critiques. In *Proceedings of IEEE International Conference on Robotics and Automation*, 6907–6914. IEEE.

Daniel, C.; Viering, M.; Metz, J.; Kroemer, O.; and Peters, J. 2014. Active Reward Learning. In *Proceedings of Robotics: Science and systems*, volume 98.

de Graaf, M. M.; Allouch, S. B.; and Klamer, T. 2015. Sharing a life with Harvey: Exploring the acceptance of and relationship-building with a social robot. *Computers in Human Behavior*, 43: 1–14.

Foshee, C. M.; Elliott, S. N.; and Atkinson, R. K. 2016. Technology-enhanced learning in college mathematics remediation. *British Journal of Educational Technology*, 47(5): 893–905.

Ghali, M. A.; Ayyad, A. A.; Abu-Naser1, S. S.; and Laban, M. A. 2018. An Intelligent Tutoring System for Teaching English Grammar. *International Journal of Academic Engineering Research*, 2.

Gómez, S.; Zervas, P.; Sampson, D. G.; and Fabregat, R. 2014. Context-aware adaptive and personalized mobile learning delivery supported by UoLmP. *Journal of King Saud University - Computer and Information Sciences*, 26(1, Supplement): 47–61.

Halilovic, A.; and Lindner, F. 2023. Visuo-Textual Explanations of a Robot's Navigational Choices. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 531–535.

Heard, J.; Harriott, C. E.; and Adams, J. A. 2018. A survey of workload assessment algorithms. *IEEE Transactions on Human-Machine Systems*, 48(5): 434–451.

Hsieh, C.-W.; and Chen, S. Y. 2016. A Cognitive Style Perspective to Handheld Devices: Customization vs. Personalization. *The International Review of Research in Open and Distributed Learning*, 17(1).

Huang, S. H.; Bhatia, K.; Abbeel, P.; and Dragan, A. D. 2018. Establishing Appropriate Trust via Critical States. In *Proceedings of International Conference on Intelligent Robots and Systems*, 3929–3936.

Huang, S. H.; Held, D.; Abbeel, P.; and Dragan, A. D. 2019. Enabling Robots to Communicate their Objectives. *Autonomous Robots*, 43(2).

Johnson-Laird, P. N. 2004. The history of mental models. *Psychology of reasoning: Theoretical and historical perspectives*, 8: 179–212.

Khajah, M. M.; Wing, R.; Lindsey, R. V.; and Mozer, M. C. 2014. Integrating latent-factor and knowledge-tracing models to predict individual differences in learning. In *Educational Data Mining*.

Kulic, D.; and Croft, E. A. 2007. Affective state estimation for human-robot interaction. *IEEE transactions on robotics*, 23(5): 991–1000.

Lee, J. I.; and Brunskill, E. 2012. The Impact on Individualizing Student Models on Necessary Practice Opportunities. In *Educational Data Mining*.

Lee, M. S.; Admoni, H.; and Simmons, R. 2021. Machine Teaching for Human Inverse Reinforcement Learning. *Frontiers in Robotics and AI*, 8: 188.

Lin, C. F.; chu Yeh, Y.; Hung, Y. H.; and Chang, R. I. 2013. Data mining for providing a personalized learning path in creativity: An application of decision trees. *Computers and Education*, 68: 199–210.

Mathieu, J. E.; Heffner, T. S.; Goodwin, G. F.; Salas, E.; and Cannon-Bowers, J. A. 2000. The influence of shared mental models on team process and performance. *Journal of applied psychology*, 85(2): 273.

Meneguzzi, F. R.; and Pereira, R. F. 2021. A survey on goal recognition as planning. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence, 2021, Canadá.*

Miller, R.; Miliband, D.; Hopkins, D.; Järvelä, S.; Spitzer, M.; Hébert, Y.; Hartley, W.; Ruano-Borbalan, J.-C.; Paludan, J.; Leadbeater, C.; and Bentley, T. 2006. *Personalising Education*. Organisation for Economic Co-operation and Development.

Minn, S.; Yu, Y.; Desmarais, M. C.; Zhu, F.; and Vie, J.-J. 2018. Deep Knowledge Tracing and Dynamic Student Classification for Knowledge Tracing. *IEEE International Conference on Data Mining*, 1182–1187.

Neubauer, C.; Schaefer, K. E.; Oiknine, A. H.; Thurman, S.; Files, B.; Gordon, S.; Bradford, J. C.; Spangler, D.; and Gremillion, G. 2020. Multimodal Physiological and Behavioral Measures to Estimate Human States and Decisions for Improved Human Autonomy Teaming. Technical report, CCDC Army Research Laboratory Aberdeen Proving Ground United States.

Orlov-Savko, L.; Jain, A.; Gremillion, G. M.; Neubauer, C. E.; Canady, J. D.; and Unhelkar, V. 2022. Factorial Agent Markov Model: Modeling Other Agents' Behavior in presence of Dynamic Latent Decision Factors. In *Proceedings of International Conference on Autonomous Agents and Multi-Agent Systems*. IFAAMAS.

Orlov-Savko, L.; Qian, Z.; Gremillion, G. M.; Neubauer, C. E.; Canady, J.; and Unhelkar, V. 2024. RW4T Dataset: Data of Human-Robot Behavior and Cognitive States in Simulated Disaster Response Tasks. In *ACM/IEEE International Conference on Human-Robot Interaction*.

Osa, T.; Pajarinen, J.; Neumann, G.; Bagnell, J. A.; Abbeel, P.; Peters, J.; et al. 2018. An algorithmic perspective on imitation learning. *Foundations and Trends in Robotics*, 7(1-2): 1–179.

Parasuraman, R.; and Riley, V. 1997. Humans and automation: Use, misuse, disuse, abuse. *Human factors*, 39(2): 230–253.

Pardos, Z. A.; and Heffernan, N. T. 2010. Modeling Individualization in a Bayesian Networks Implementation of Knowledge Tracing. In *User Modeling, Adaptation, and Personalization*.

Pavlik, P. I.; Cen, H.; and Koedinger, K. 2009. Performance Factors Analysis - A New Alternative to Knowledge Tracing. In *International Conference on Artificial Intelligence in Education*.

Piech, C.; Bassen, J.; Huang, J.; Ganguli, S.; Sahami, M.; Guibas, L. J.; and Sohl-Dickstein, J. N. 2015. Deep Knowledge Tracing. In *Annual Conference on Neural Information Processing Systems*.

Pontual Falcão, T.; Mendes de Andrade e Peres, F.; Sales de Morais, D. C.; and da Silva Oliveira, G. 2018. Participatory methodologies to promote student engagement in the development of educational digital games. *Computers and Education*, 116: 161–175.

Puterman, M. L. 2014. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

Qian, P.; and Unhelkar, V. 2022. Evaluating the Role of Interactivity on Improving Transparency in Autonomous Agents. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, 1083–1091.

Qian, P.; and Unhelkar, V. V. 2024. Interactively Explaining Robot Policies to Humans in Integrated Virtual and Physical Training Environments. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 847–851.

Qian, Z.; Orlov Savko, L.; Neubauer, C.; Gremillion, G.; and Unhelkar, V. 2024. Measuring Variations in Workload during Human-Robot Collaboration through Automated After-Action Reviews. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 852–856.

Quintero-Pena, C.; Chamzas, C.; Sun, Z.; Unhelkar, V.; and Kavraki, L. E. 2022. Human-Guided Motion Planning in Partially Observable Environments. In *Proceedings of International Conference on Robotics and Automation*. IEEE.

Rong, Y.; Leemann, T.; Nguyen, T.-t.; Fiedler, L.; Qian, P.; Unhelkar, V.; Seidel, T.; Kasneci, G.; and Kasneci, E. 2023. Towards Human-centered Explainable AI: User Studies for Model Explanations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Rong, Y.; Qian, P.; Unhelkar, V.; and Kasneci, E. 2024. I-CEE: Tailoring Explanations of Image Classification Models to User Expertise. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(19): 21545–21553.

Sakai, T.; and Nagai, T. 2022. Explainable autonomous robots: a survey and perspective. *Advanced Robotics*, 36(5-6): 219–238.

Seo, S.; Kennedy-Metz, L. R.; Zenati, M. A.; Shah, J. A.; Dias, R. D.; and Unhelkar, V. V. 2021. Towards an AI Coach to Infer Team Mental Model Alignment in Healthcare. In *International Conference on Cognitive and Computational Aspects of Situation Management*. IEEE.

Seo, S.; and Unhelkar, V. 2024. IDIL: Imitation Learning of Intent-Driven Expert Behavior. In *Proceedings of International Conference on Autonomous Agents and Multi-Agent Systems*. IFAAMAS.

Shemshack, A.; and Spector, J. 2020. A systematic literature review of personalized learning terms. *Smart Learning Envrionments*.

Tabrez, A.; Agrawal, S.; and Hayes, B. 2019. Explanation-based Reward Coaching to Improve Human Performance via Reinforcement Learning. In *International Conference on Human-Robot Interaction*, 249 – 257. IEEE.

Tjoa, E.; and Guan, C. 2020. A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE transactions on neural networks and learning systems*, 32(11): 4793–4813.

Truong, H. M. 2016. Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities. *Computers in Human Behavior*, 55: 1185–1193.

Van-Horenbeke, F. A.; and Peer, A. 2021. Activity, plan, and goal recognition: A review. *Frontiers in Robotics and AI*, 8.

Watkins, O.; Huang, S.; Frost, J.; Bhatia, K.; Weiner, E.; Abbeel, P.; Darrell, T.; Plummer, B.; Saenko, K.; and Dragan, A. 2021. Explaining robot policies. *Applied AI Letters*, 2(4).

Wilson, K. H.; Karklin, Y.; Han, B.; and Ekanadham, C. 2016. Back to the basics: Bayesian extensions of IRT outperform neural networks for proficiency estimation. In *Educational Data Mining*.

Xiong, X.; Zhao, S.; Inwegen, E. V.; and Beck, J. E. 2016. Going Deeper with Deep Knowledge Tracing. In *Educational Data Mining*.

Yang, S. C.-H.; Vong, W. K.; Sojitra, R. B.; Folke, T.; and Shafto, P. 2021. Mitigating belief projection in explainable artificial intelligence via Bayesian teaching. *Scientific reports*, 11(1): 9863.

Yeung, C.-K.; and Yeung, D. Y. 2018. Addressing two problems in deep knowledge tracing via prediction-consistent regularization. *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*.

Yudelson, M. V.; Koedinger, K.; and Gordon, G. J. 2013. Individualized Bayesian Knowledge Tracing Models. In *International Conference on Artificial Intelligence in Education*.

Zhan, Y.; Fachantidis, A.; Vlahavas, I.; and Taylor, M. E. 2014. Agents Teaching Humans in Reinforcement Learning Tasks. In *Proceedings of International Conference on Autonomous Agents and Multiagent Systems*.