

# Active Explicable Planning for Human-Robot Teaming

Akkamahadevi Hanni and Yu Zhang

ahanni,yzhan442@asu.edu

School of Computing, Informatics and Decision Systems Engineering

Arizona State University

Tempe, Arizona, USA

## ABSTRACT

Intelligent robots are redefining autonomous tasks but are still far from being fully capable of assisting humans in day to day tasks. An important requirement of collaboration is to have a clear understanding of each other's expectations and capabilities. Lack of which may lead to serious issues such as loose coordination between teammates, ineffective team performance, and ultimately mission failures. Hence, it is important for the robot to behave explicably to make themselves understandable to the human. One of the challenges here is that the expectations of the human are often hidden and dynamically changing as the human interacts with the robot. Existing approaches in plan explicability often assume the human's expectations are known and static. In this paper, we propose the idea of **active explicable planning** to address this issue. We apply a Bayesian approach to model and predict dynamic human beliefs to be more anticipatory, and hence can generate more efficient plans without impacting explicability. We hypothesize that active explicable plans can be more efficient and more explicable at the same time, compared to the plans generated by existing methods. From the preliminary results of M-Turk study, we find that our approach effectively captures the dynamic belief of the human which can be used to generate efficient and explicable behavior that benefits from dynamically changing expectations.

### ACM Reference Format:

Akkamahadevi Hanni and Yu Zhang. 2021. Active Explicable Planning for Human-Robot Teaming. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21 Companion)*, March 8–11, 2021, Boulder, CO, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3434074.3447154>

## 1 INTRODUCTION

Intelligent robots are redefining autonomous tasks but are still far from being fully capable of assisting humans in day to day tasks. As these agents further embed into our lives, the need to build agents capable of working not just by themselves but also with humans increases. Human-robot collaborations can occur in simple household chores or complex industry-level tasks. Whether the task is simple or complex, it is essential to understand the expectations of all members of a team which is critical for the success of a

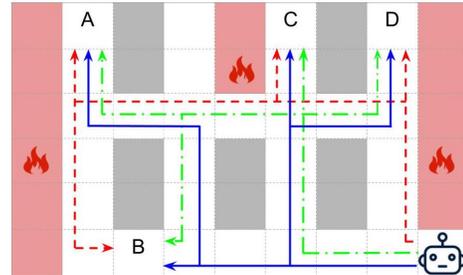
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*HRI '21 Companion*, March 8–11, 2021, Boulder, CO, USA

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-8290-8/21/03...\$15.00

<https://doi.org/10.1145/3434074.3447154>



**Figure 1: USAR Domain: building layout indicating the path taken by Optimal plan - OP (Green), Explicable plan - EXP (Red), Active Explicable plan with dynamic modeling - ActiveEXP (Blue)**

ActiveEXP (Blue)	EXP (Red)	OP (Green)
Collect C	Collect D	Collect C
Collect D	Collect C	Collect D
Collect B	Collect A	Collect A
Collect A	Collect B	Collect B

**Table 1: Illustrative Plans ActiveEXP, EXP and OP**

team's mission. Inconsistency between expectations and reality can lead to loose coordination among the team members or even failure of a mission. Typically, these are caused by discrepancies in their understanding of the domain. To avoid such inconsistencies, it is desirable that the robot behaves in an explicable manner by considering the differences in the models [7, 14, 16]. This requires the knowledge of the human's understanding of the robot. Prior work, however, often assumes that such an understanding of the robot is known and static.

In this paper, we introduce the notion of active explicable plan to relax this strong assumption. The motivation is that the human's understanding and expectations of the robot are hidden and subject to change as a result of observing the robot performing actions. We apply a Bayesian approach to model and predict the human's dynamic belief to capture the expectations of the human at any given time. We hypothesize that active explicable plans generated with dynamic modeling of human's expectations (ActiveEXP) can be both more efficient and explicable than the plans generated by the existing plan explicability approaches (EXP), and more explicable than the optimal plans (OP). Let us illustrate this using a motivational example below.

### 1.1 Motivational Example

Consider you and a robot working together as a team in an Urban Search and Reconnaissance scenario. Assume you are the supervisor

and the task of the robot is to collect certain objects (A, B, C, D) from the wrecked building as shown in Fig. 1. Initially, without prior knowledge, you would expect the robot to take the shortest path to collect the objects. This corresponds to the EXP plan (red) depicted in Fig. 1, where the robot picks up D, C, A, B in sequence. Now, imagine that you observe the robot first went to pick up C and then D, as shown in the blue path. You would probably infer that the robot does not wish to be exposed close to a fire. Having observed as such, your expectation would change for the robot’s later steps. In particular, you would expect the robot to avoid the fire location in the middle to pick up B next and finally A. This corresponds to the ActiveExp (blue) plan. Note that your expectation changes after observing the robot’s first few actions! The optimal plan in this case, however, corresponds to the green path. This is due to the fact even though exposure close to fire incurs a cost for the robot, it is less costly to be exposed on a short path than taking a long detour as in the blue path. The observation here is that the ActiveEXP plan is more efficient than the EXP plan while being as explicable as the EXP plan, due to the dynamic nature of the human’s understanding.

The contribution in our work is three folds: we provide 1) a method for dynamic modeling of the human belief, 2) an active explicable planning framework that generates active explicable plans that are simultaneously more efficient, 3) preliminary analysis.

## 2 RELATED WORK

In general, explicable plan generation falls under the umbrella of explainable planning. Several methods have been proposed under the pretext of explicability [7, 16], interpretability, predictability, legibility [3, 8], and transparency [11] which focus on generating plans that assist in understanding the plans or goals of the executing agent.

Our work on active explicable plan is connected to plan explicability [16] where the aim is to generate behavior that is expected by the human teammate assuming the goal is already known. This is critical in human-robot teaming tasks since inconsistency between the robot’s behavior and human’s expectation can lead to the loss of trust and safety risks to the humans. In [7], the authors assume a human’s model and choose a plan in the robot’s model that is closest (in terms of plan distances) to the optimal plan in the human’s model. In contrast, authors in [16] and [14] relax the assumption of a known human model. In the former method they learn the human’s expectations via a sequential labeling scheme while in the latter, the authors define a Bayesian model of the observer’s belief.

The dynamic modeling of the human’s understanding of the robot in our work is similar to work on plan recognition [2, 6, 9, 12], where observations inform the recognition. It is particularly similar to dynamic trust and intent modeling. Most approaches [4, 10, 15] rely on HMM or Beta models for modeling dynamic changes. However, these models are not directly applicable to our problem formulation.

## 3 PROBLEM SETTING

In this work, we follow the representation of a model  $\mathcal{M}$  from the definition of Classical Planning [13] represented by the PDDL [5] schema. A model  $\mathcal{M} = ((\mathcal{F}, \mathcal{A}), \mathcal{I}, \mathcal{G})$  where  $\mathcal{F}$  and  $\mathcal{A}$  are a finite

set of fluents and actions respectively where  $\forall a, a \in \mathcal{A}$  is defined in terms of precondition list  $\text{Pre}(a)$ , add effect list  $\text{Add}(a)$ , delete effect list  $\text{Del}(a)$  and cost  $C(a)$ ; all of which are subsets of  $\mathcal{F}$ . Let  $S$  be the set of states where  $s \in S$  is an instantiation of a unique set of fluents  $\mathcal{F}$ . The initial state  $\mathcal{I} \in S$  and the goal state  $\mathcal{G} \in S$ .

**Definition 3.1.** We define the explicable planning problem setting as a tuple  $\mathcal{P} = \langle M, \mathcal{I}, \mathcal{G}, b_0(\mathcal{M}_H) \rangle$  where,

- $M$  is the model space, which is finite.
- $\mathcal{I}$  is the initial state.
- $\mathcal{G}$  is the goal state.
- $b_0(\mathcal{M}_H)$  is the initial belief of the human’s understanding of the robot.

A candidate plan space  $\Pi = \{\pi_1, \pi_2, \dots, \pi_p\}$  is a finite set of  $p$  plans which is a union of all plans generated by each possible model  $m \in M$ , bounded by a cost threshold  $\zeta$  (arbitrarily set). Note that  $\zeta \geq C(\pi^*)$ , where  $C(\pi^*)$  is the cost of the optimal plan for  $\{\mathcal{I}, \mathcal{G}\}$  under the true robot’s model  $\mathcal{M}_R$ .  $\Pi_R \subset \Pi$  is the set of all plans generated by the true model of the robot  $\mathcal{M}_R$  bounded by the cost  $\zeta$ .

## 4 METHODOLOGY

Explicable planning in the setting described above is about how to plan while considering the human’s understanding of the robot, which may differ from the robot’s true model. The objective there is how to trade off the action cost of a plan with its explicability score. In this work, we extend explicable planning to consider the change in human’s understanding as observations are made about the robot.

### 4.1 Active Explicable Plan

Plan explicability is commonly viewed as whether a human can interpret the robot’s plan [1, 7, 16]. While it has been defined in various ways in prior work, we use a Bayesian formulation and define the explicability of a plan as the probability of the plan being in a set of candidate plans that can be generated by the human model. More specifically:

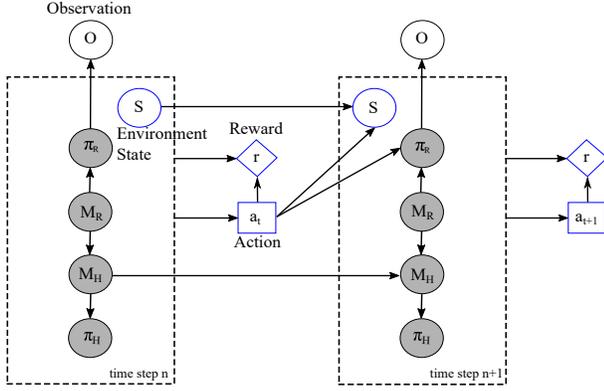
$$\mathcal{E}(\pi_R) = P(\pi_R \in \Pi_H \mid \mathcal{M}_H) \quad (1)$$

where  $\Pi_H$  is the set of candidate plans in  $\mathcal{M}_H$ . However, in the above formulation  $\mathcal{M}_H$  remains unchanged as in the prior work. In our work, when  $\mathcal{M}_H$  changes dynamically, the active explicability of a plan is defined as the average probability of actions that are explicable in  $\pi_R$  following the work in [16]. Therefore, the active explicability score is given by:

$$\mathcal{E}_{\mathcal{A}}(\pi_R) = \frac{1}{T} \sum_t P(\pi_R[t] \in \Pi_H \langle \pi_R[1:t-1] \rangle \mid \mathcal{M}_H^t) \quad (2)$$

where  $\pi_R[t]$  represents the action at step  $t$  in  $\pi_R$  and  $\Pi_H \langle \pi_R[1:t-1] \rangle$  represents the set of candidate plans generated by  $\mathcal{M}_H^t$  at time  $t$  after executing the first  $t-1$  actions in  $\pi_R$ . Essentially, this means that the action at step  $t$  must be explicable in the human model.

When we assume that the human observe is noisily rational, the human believes the robot is noisily rational, and when the human’s model is uncertain, the above equation changes its form respectively.



**Figure 2: Influence Diagram for active explicable planning for modeling dynamic human belief.**

In such cases, to simplify computation, the explicability score can be optimized by minimizing the entropy between the distributions of plans  $\pi_R$  and  $\pi_H$  generated by  $\mathcal{M}_R^t$  and  $\mathcal{M}_H^t$  respectively. In such a case, an active explicable plan  $\pi_{ae}$  for dynamically changing human belief is given by:

$$\pi_{ae} = \operatorname{argmin}_{\pi_R} \frac{1}{T} \sum_t H(\Pi_R^t, \Pi_H^t) \quad (3)$$

The benefit of the approximation above is that we can compute the plans for each candidate model a priori and avoid computing plans for many sub-planning problems.

## 4.2 Human belief update

To simplify the problem, we assume that the environment states and robot's actions are fully observable by the human and that the actions of the robot are deterministic.  $O = o_1, o_2, \dots, o_T$  is the sequence of actions that are observed by the human at respective time steps.

We consider that the human's expectation of the robot  $\mathcal{M}_H$  is not *directly* influenced by the observations made, instead, it is influenced by the robot's compatible model  $\mathcal{M}_R$  (the model that derives the robot's behavior) and the human's prior belief as shown in Fig. 2.

When a new action  $o_{t+1}$  performed by the robot is observed by the human teammate, the human's belief of the robot's compatible model  $\mathcal{M}_R$  is updated.

$$\phi(\mathcal{M}_R^{t+1}) \propto P(\mathcal{M}_R^{t+1} | O^{t+1}) \quad (4)$$

This in turn influences the human's expectations. Initially, when the robot does not have any information of the human's expectations, the robot may assume a uniform distribution for the human's belief  $b_0$  ( $\mathcal{M}_H$ ). Given the prior belief  $b_t$  ( $\mathcal{M}_H$ ), the observer's belief is updated as follows:

$$\begin{aligned} b_{t+1}(\mathcal{M}_H) &= \text{BeliefUpdate}(b_t(\mathcal{M}_H), \phi(\mathcal{M}_R^{t+1})) \\ &= P(\mathcal{M}_H^{t+1} | \mathcal{M}_H^t, \mathcal{M}_R^{t+1}) \end{aligned} \quad (5)$$

Assuming the human observer is noisily rational and that he/she believes the robot is NOT a rational agent, we consider a Boltzmann

distribution for the robot's plan  $\pi_R$  and the human's expected plan  $\pi_H$  given by:

$$P(\pi_R^t | \mathcal{M}_R^t) \propto e^{-\beta \times C(\pi_R^t)} \quad \text{and} \quad P(\pi_H^t | \mathcal{M}_H^t) \propto e^{-\beta \times C(\pi_H^t)} \quad (6)$$

The likelihood of observing  $o_t$  given a plan  $\pi_R^t$  takes a binary form given by:

$$P(O^t | \pi_R^t) = [0, 1] \quad (7)$$

where the probability is 1 if the observation  $o_t$  is the  $t^{\text{th}}$  action in the plan  $\pi_R$  and 0 otherwise.

As shown in Fig. 2, after every belief update the robot receives a reward  $r \in \mathbb{R}$ . Following the approximation in Eq. 3 we define a reward function that balances explicability and plan cost given below:

$$r = -H(\Pi_H^t, \Pi_R^t) - \gamma C(a_t) \quad (8)$$

where  $C(a_t)$  is the cost of the action observed by the human at time  $t$  weighted by a constant  $\gamma$ . Notice that in contrast to a POMDP, this reward function is not static as it depends on the belief.

The expected cumulative reward obtained for a plan  $\pi_i$  for starting in a state  $\mathcal{I}$  is defined as the sum of rewards obtained for every action  $a_t$  in plan  $\pi_i$  with length  $T$  given by:

$$\text{cumReward}(\pi_i) = E [\sum_{t=0}^T R(s_t, a_t)] \quad (9)$$

The goal of this system is to find a plan that maximizes the expected cumulative reward.

## 5 EVALUATION AND RESULTS

### 5.1 Evaluation Domain

To evaluate our approach we have designed a Treasure Collection domain. The layout of this domain is shown in Fig. 4. It includes a jackal robot and a set of color-coded objects, each having a different utility value. The initial location of the robot, objects in the layout along with their utilities, and the destination location where the objects have to be deposited are known to both the human and robot. The goal of the robot is to pick up a target number of objects in the layout and place them at the destination location such that the reward of the robot is maximized. Also, the robot incurs a cost factor for navigating through each cell in the layout which is in the range  $[0.1, 0.7]$ . The range is known to the human teammate but the actual value is not. This constitutes the discrepancy between the human's understanding of the robot and the robot's model.

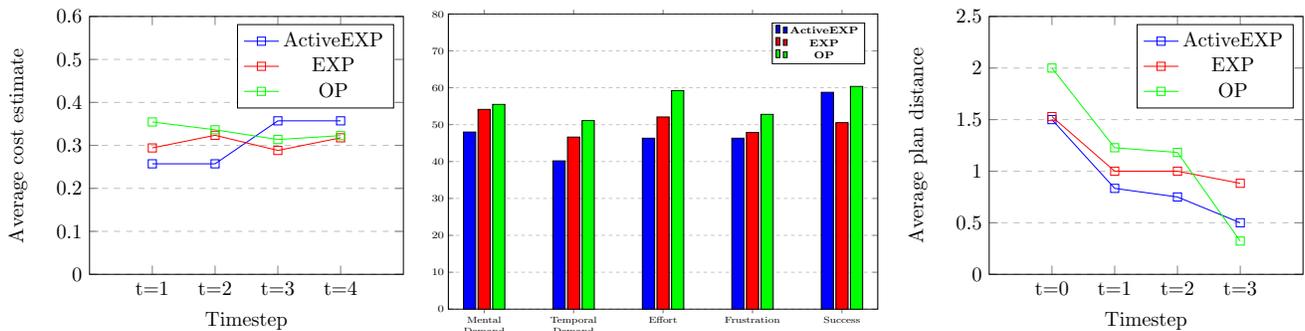
### 5.2 Experiment

To test the effectiveness of our approach, we compared the plan generated by our approach with the optimal plan (OP) in the robot model and the explicable plan generated based on the fixed human model (EXP) (that is most likely with a minimal navigation cost).

We form the following two hypotheses:

*H1 = "Both the human's belief and expectation of the robot change dynamically during the course of the robot's plan execution"*

*H2 = "The active explicable plan generated with dynamic modeling of human's expectations (ActiveEXP) is no less explicable than the plans generated by the explicable plan approach (EXP) while being more cost-efficient"*



**Figure 3: Experimental Results from left to right: (a) Changes of average cost estimated by human at every time step. (b) NASA TLX results for testing. (c) Changes in plan distance between robot plan and human expected plan.**



Plan	Cost
OP	18.2
EXP	29.4
ActiveEXP	21

**Table 2: Cost of plans: OP, EXP and ActiveEXP in the robot’s true model**

**Figure 4: Layout for Treasure Collection Domain**

We collected responses from 45 Mturkers (15 for each plan). To test the first hypothesis  $H1$ , in our Mturk study, we asked the subjects to input their expected cost of navigation after every action.

Similarly, to test the second hypothesis  $H2$ , after demonstrating each action, we asked the subjects to input their expected remaining actions and asked if the action demonstrated by the robot was explicable or not. We compared the expected plans of the subjects with the robot’s plan using plan distances. We measured the distance based on the objects collected at each time step.

### 5.3 Results

The result of the experiment conducted to test  $H1$  is shown in Fig. 3(a). We can observe that the cost of navigation predicted by the subjects initially for ActiveEXP is low and changes after observing the robot’s actions and at the end of time step  $t=4$  is higher. For EXP and OP, surprisingly, the estimates stay relatively flat throughout the course of robot actions. This may be due to the fact that these methods create less informative plans for model estimates since neither of them considers the dynamic change of human belief.

The result of the experiment conducted to test  $H2$  is shown in Fig. 3(c). We can observe from the graph that the distance between ActiveEXP and the human expected plans is relatively smaller at all time steps as compared to that of EXP and OP. It is also interesting to observe that even though OP is optimal in the robot’s model it is the farthest from what the human expects for the first three

time-steps. However, after the third time step we observe that the optimal plan becomes much closer to the human’s expectations than EXP. Intuitively, when the robot executes OP sequentially the human observes that the robot only picks up closer objects despite having low utility values. This behavior conveys to the user that the robot only cares about objects around it.

But the order in which the objects are collected in OP allowed ambiguity for the human to continue to expect something different than the robot’s plan until the last time step. This is where dynamic belief modeling plays an important role by incorporating how the human’s belief might change after observing certain actions which help the robot actively behave explicably.

In addition to  $H2$  we can also observe that the plan generated by our approach (ActiveEXP) is more efficient than the traditional explicable plan EXP as shown in Table 2. Intuitively, choosing EXP because it is the most expected plan in the human model would not be beneficial as it results in obtaining very low utility value in the true robot’s model.

Fig. 3(b) presents the subjective testing results and we can see that our approach (ActiveEXP) performs better than the other methods. We performed an independent t-test study on the data collected by ActiveEXP and OP methods and found a statistically significant difference between the two approaches with  $p$  value 0.021583. Also, this is in agreement with the objective results presented above.

## CONCLUSION

In this paper, we introduced the problem of active explicable planning with dynamic modeling of the human belief. It addressed the limitations of existing work on explicable planning. A planning framework was then proposed to generate active explicable plans. We evaluated our method against existing planning methods and showed that an active explicable plan is more efficient without suffering explicability for human-robot teaming.

## ACKNOWLEDGMENTS

We thank the anonymous reviewers for their helpful comments. This research is supported in part by the NSF grant IIS-1844524, the NASA grant NNX17AD06G, and the AFOSR grant FA9550-18-1-0067.

## REFERENCES

- [1] T. Chakraborti, Anagha Kulkarni, S. Sreedharan, David E. Smith, and S. Kambhampati. 2019. Explicability? Legibility? Predictability? Transparency? Privacy? Security? The Emerging Landscape of Interpretable Agent Behavior. In *ICAPS*.
- [2] Eugene Charniak and R. Goldman. 1993. A Bayesian Model of Plan Recognition. *Artif. Intell.* 64 (1993), 53–79.
- [3] Anca D. Dragan and S. Srinivasa. 2013. Generating Legible Motion. In *Robotics: Science and Systems*.
- [4] Ehab ElSalamouny, V. Sassone, and M. Nielsen. 2009. HMM-Based Trust Model. In *Formal Aspects in Security and Trust*.
- [5] Maria Fox and Derek Long. 2011. PDDL2.1: An Extension to PDDL for Expressing Temporal Planning Domains. *CoRR* abs/1106.4561 (2011). arXiv:1106.4561 <http://arxiv.org/abs/1106.4561>
- [6] Henry A. Kautz and James F. Allen. 1986. Generalized Plan Recognition. In *AAAI*.
- [7] Anagha Kulkarni, T. Chakraborti, Yantian Zha, Satya Gautam Vadlamudi, Y. Zhang, and S. Kambhampati. 2016. Explicable Robot Planning as Minimizing Distance from Expected Behavior. *ArXiv* abs/1611.05497 (2016).
- [8] Anagha Kulkarni, Siddharth Srivastava, and S. Kambhampati. 2019. A Unified Framework for Planning in Adversarial and Cooperative Environments. In *AAAI*.
- [9] S. J. Levine and B. Williams. 2014. Concurrent Plan Recognition and Execution for Human-Robot Teams. In *ICAPS*.
- [10] X. Liu and A. Datta. 2012. Modeling Context Aware Dynamic Trust Using Hidden Markov Model. In *AAAI*.
- [11] Aleck M. MacNally, N. Lipovetzky, M. Ramirez, and A. Pearce. 2018. Action Selection for Transparent Planning. In *AAMAS*.
- [12] M. Ramirez and H. Geffner. 2010. Probabilistic Plan Recognition Using Off-the-Shelf Classical Planners. In *AAAI*.
- [13] Stuart Russell and Peter Norvig. 2010. *Artificial Intelligence: A Modern Approach* (3 ed.). Prentice Hall.
- [14] S. Sreedharan, Anagha Kulkarni, T. Chakraborti, D. E. Smith, and S. Kambhampati. 2020. A Bayesian Account of Measures of Interpretability in Human-AI Interaction. *ArXiv* abs/2011.10920 (2020).
- [15] Y. Xia, Alexandre Parmentier, and R. Cohen. 2019. Trust Modelling in Dynamic Environments.
- [16] Y. Zhang, S. Sreedharan, Anagha Kulkarni, T. Chakraborti, Hankui Zhuo, and S. Kambhampati. 2017. Plan explicability and predictability for robot task planning. *2017 IEEE International Conference on Robotics and Automation (ICRA)* (2017), 1313–1320.