Human-Aware Epistemic Task Planning for Human-Robot Collaboration

Shashank Shekhar[1], **Anthony Favier**[1,2], **Rachid Alami**[1,2]

[1] LAAS-CNRS, Universite de Toulouse, CNRS, INSA, UPS, Toulouse, France
[2] Artificial and Natural Intelligence Toulouse Institute (ANITI)
firstname.lastname@laas.fr

Abstract

We present a novel human-aware epistemic planning framework designed for collaborative human-robot interactions, specially tailored for situations where the agents' shared execution experiences can be interrupted by the uncontrollable nature of humans. Our objective is to generate a robot policy that accounts for such uncontrollable behaviors, thus enabling the anticipation of potential progress achieved by the robot when the experience is not shared, e.g., when humans are briefly absent from the shared environment to complete a subtask. But this anticipation is considered from the perspective of humans who keep an estimated robot's model. As a first step to address it, we propose a general planning framework and build a solver based on AND/OR search which integrates knowledge reasoning; this includes assessing situations by perspective taking. Our approach dynamically models and manages the expansion or contraction of potential worlds while tracking whether or not agents share the task execution experiences. This helps the planner (or the robot) to prepare itself with a set of worlds that humans would consider possible. The robot *assesses* the situation from the human perspective and removes the worlds that it has reason to think are impossible. However, there might still be an impossible world that is indistinguishable from the real world. In different situations, thanks to our planning framework, the robot's policy built offline can determine an appropriate course of action, such as answering human *queries*, explicitly *communicating* some fact without being annoying, or *taking* appropriate action in the presence of the human to help them narrow down the possibilities further, facilitating collaboration. Our preliminary experiments show that the framework is effective for behavior planning in different situations. We discuss the practical issues in different problem settings.

Introduction

The increasing number of robot-assisted applications has led to a focus on human-robot collaboration (HRC) research (Baratta et al. 2023; Semeraro, Griffiths, and Cangelosi 2023). Collaborative robots are beneficial in real-world scenarios like construction engineering (Liang et al. 2021), workshops (Coupeté, Moutarde, and Manitsaris 2015), and nursing care (Nieto Agraz et al. 2022).

Planning and decision-making are crucial for successful collaboration and task completion in multi-agent scenarios. However, uncontrollable human behavior in Human-Robot Collaboration (HRC) can disrupt shared execution experiences. For example, when humans temporarily leave the shared environment, it poses unique challenges for collaborative robots, potentially leading to false beliefs or impacting overall task performance. So, automated planning is essential to overcome these challenges and maintain efficiency.

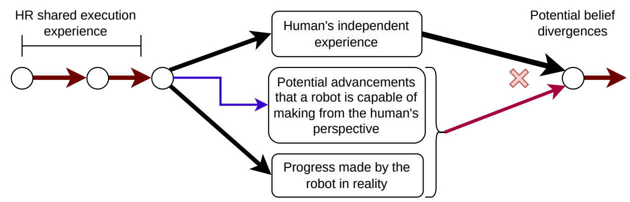

Figure 1: *Our planning framework is endowed with the ability to make the difference between **H&R** shared and individual execution experiences in the planned activities. It can anticipate potential belief divergence between **H&R** and also estimate the updated beliefs of **H** when they meet again* (situation assessment) *based on a distinction between observable and non-observable facts. This will be used to plan communicative actions or adapt the plan to ensure the shared experience of some actions.*

Recent efforts such as Dynamic Epistemic Logic (DEL) based framework introduced in (Bolander, Dissing, and Hermann 2021), propose an epistemic planning approach for human-robot collaboration. In addition, efforts have been made to tackle challenges related to human absence from the environment. For instance, addressing first- and higher-order false beliefs in (Dissing and Bolander 2020), drawing inspiration from the well-known Sally and Ann test (Wimmer and Perner 1983). In the context of planning (Favier, Shekhar, and Alami 2023b) and shared plan execution (Devin and Alami 2016), approaches have been proposed, with the latter focusing on a reactive strategy to manage the unpredictability of human absence. Favier, Shekhar, and Alami (2023) propose an offline planning approach that considers predictable human absence. When the humans are back, they still hold the outdated (false) beliefs that they possessed earlier. Situation assessment allows them to update the progress made by the robot if any, however, those false beliefs remain that cannot be corrected by situation assessment. They suggested methods to tackle collaboration under false beliefs.

We propose a novel human-aware epistemic planning framework. It enables the robot to estimate, anticipate, and adapt to the scenarios in which human and robot partners have disrupted shared execution experiences. However, it considers the human's perspective and estimation regarding the potential progress achieved when the exact progress is not experienced directly by them. Figure 1 provides a high-level illustration of what happens when agents share execution experience and when they do not. This proactive strategy ensures that the robot interacts in a way that corresponds with the human's expected perceptions, hence improving the effectiveness of task performance.

To systematically track instances of shared execution experiences, we present our concepts abstractly and symbolically. For instance, we adapt the notion of *"Places"* to denote locations within the environment. When an action occurs at a place, only agents present there share this experience, either by observing the execution or as the actor (Shekhar et al. 2023; Favier, Shekhar, and Alami 2023b).

As the first step, we build an AND/OR search-based offline planner that facilitates Theory of Mind (ToM) by integrating knowledge reasoning and incorporating perspective-taking to assess situations. It dynamically manages the evolution or contraction of estimated possible worlds from the human's point of view. It helps the planner to prepare itself with a set of worlds that humans would consider possible.

The robot takes the human's perspective for situation assessment, thus discarding estimated worlds deemed impossible. To allow appropriate situation assessment, we define the concept of *observability* of state property, extending (Shekhar et al. 2023). An appropriate course of action based on the situation at hand is planned. The policy built offline enables the robot to take proactive steps, such as waiting for the human to *inquire* about a fact, *communicating* relevant information without being annoying (*e.g.*, not verbalizing a fact already known to them), or *executing* suitable actions to allow the human to narrow down possibilities upon their return or in presence.

The paper is structured as follows. A use-case scenario is presented next, followed by the background information, covering tools necessary for our work. Next, we delve into our proposed planning framework, followed by the section describing the AND/OR search-based planning algorithm. The section next to that discusses related work in the field. With our preliminary experiments in the section followed, we show the effectiveness of the proposed framework for behavior planning for a collaborative robot in diverse scenarios. We conclude by summarizing our work in the end.

The Cube Organization Case Study

Take the case illustrated in Figure 2, in which the job of arranging cubes in boxes is shared by a human and a robot.

Assume that only **H** is capable of moving around and can exhibit unpredictable behavior (*nondeterminism*), such as moving to the other table (*ot*) to retrieve cubes, while **R** may continue to act in the environment. From the **H**'s perspective, **R** may move some or all of the cubes from the main table (*mt*) and place them into one of the boxes, or it may choose to take no action at all. Upon returning to the main table *mt*, **H** may discover that some, none, or all of the cubes originally on *mt* are missing, indicating that they have been placed in one of the boxes.

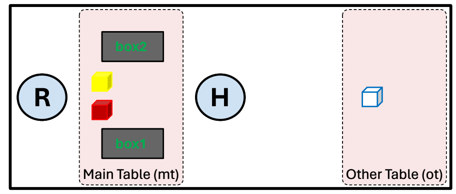

Figure 2: *The cube organization scenario involves three cubes: c_r (red), c_y (yellow), and c_w (white). c_r and c_y are placed on mt (main table), while c_w is on ot (other table). There are two boxes, box_1 and box_2, placed on mt, which can be either transparent or opaque. The shared objective is to organize the cubes in such a way that cubes from one table are placed in one box. The choice of which box is flexible as long as each table's cubes end up in separate boxes.*

If **R** places some cubes from *mt* into one of the boxes, **H** will only learn about this decision upon encountering the transparent box. But when opaque, **R** has several options: it can communicate, wait for **H** to inquire, or select one of the remaining cubes from *mt* and place it in the correct box in the presence of **H**, thus implicitly communicating its choice.

We explore such collaborative scenarios and plan from the robot's point of view by taking into account its estimated model of the environment and its human partner. Similarly, the human collaborator has an approximation of the robot's model, enabling them to anticipate the robot's action. We provide more details on these models and our assumptions about their accuracy and falsity. These models are contained within the robot and are used in planning such that human behavior can only be estimated and emulated.

Background

We outline the necessary tools for our framework, starting with the basic concepts of Dynamic Epistemic Logic (DEL) (Bolander, Dissing, and Herrmann 2021; Engesser et al. 2017). Specifically, we cover epistemic states and actions, explaining how they apply to a state and the process of transitioning to the next state. Additionally, we clarify the conditions under which a formula holds in a given state. To address issues such as communication between **H** and **R**, our development adapts these concepts.

We then describe human-aware task planning as presented in (Buisan et al. 2022; Favier, Shekhar, and Alami 2023b).

Dynamic Epistemic Logic

We will start with standard definitions from the DEL literature [1], followed by examples to explain different concepts.

[1] The description is drawn from the DEL literature (Bolander and Andersen 2011; Bolander, Dissing, and Herrmann 2021), with a few adjustments to suit our specific requirements. We also ensure technical simplicity by allowing accessibility of all worlds and

We define epistemic languages, epistemic states, and epistemic actions. All of these are defined relative to a given finite set of agent names (or simply agents, *e.g.*, **H** - human and **R** - robot) $\mathcal{A}$ and a given finite set of atomic propositions *P*. The epistemic language $\mathcal{L}_{\mathcal{K}}$ is:

$$\varphi ::= \top \mid \bot \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_i\varphi$$

As usual, $K_i\varphi$ is read as agent i knows φ. We can generalize it to $C\varphi$ which represents common knowledge.

We *evaluate* a formula in an *epistemic model*, $\mathcal{M} = \langle W, (\sim_i)_{i\in\mathcal{A}}, V\rangle$. Here, W is non-empty finite set of *worlds*, $\sim_i \subseteq W \times W$, called the *indistinguishability* relation w.r.t. agent i, and V the *valuation* function that maps W to 2^P.

Definition 1. *An (epistemic) state is $s = \langle \mathcal{M}, W_d\rangle$, where $\mathcal{M}$ is the epistemic model as described earlier and W_d is the set of designated worlds. Here, $W_d \subseteq W$ and for each world $w \in W$ and $w_d \in W_d$, we require that $(w_d, w) \in (\cup_{i\in\mathcal{A}} R_i)^*$. Note that (concerning our current context), a state s_a can be represented as $\langle \mathcal{M}, w_d\rangle$ such that $W_d = \{w_d\}$ is known as a* global *epistemic state.*

The truth of epistemic formulas is defined as follows:

$$\mathcal{M}, w \models p \text{ iff } p \in V(w) \text{ for } p \in P$$

$$\mathcal{M}, w \models \neg\varphi \text{ iff } \mathcal{M}, w \not\models \varphi$$

$$\mathcal{M}, w \models \varphi \wedge \psi \text{ iff } \mathcal{M}, w \models \varphi \text{ and } \mathcal{M}, w \models \psi$$

$$\mathcal{M}, w \models K_i\varphi \text{ iff } \mathcal{M}, v \models \varphi \text{ for all } v \text{ such that } w \sim_i v$$

Also, $\langle \mathcal{M}, W_d\rangle \models \varphi$ iff $\langle \mathcal{M}, w_d\rangle \models \varphi$ for all $w_d \in W_d$.

We represent epistemic states as graphs, where nodes represent worlds and edges represent indistinguishability relations. Each world is labeled with the propositions that hold within it. Typically, we omit world names and edges that are implied by reflexivity or transitivity. Designated worlds are denoted by a circled marker.

Definition 2. *An event model $\mathcal{E}$ is a tuple $\langle E, (\sim_i)_{i\in\mathcal{A}}, pre, post\rangle$, where E is a non-empty finite set of events, $\sim_i \subseteq E \times E$ shows an equivalence relation, $pre : E \to \mathcal{L}_{\mathcal{K}}$ defines precondition of every event, and $post : E \to \mathcal{L}_{\mathcal{K}}$ defines the effect, defining the conjunction of literals $p \in P$.*

An epistemic action $a = \langle \mathcal{E}, E_d\rangle$, where $\mathcal{E}$ is the event model and $E_d \subseteq E$ is a non-empty set of designated events. A state allows an action to be *applied* only if there is a designated event with a satisfied precondition for each designated world in the state. Similar to a state, an epistemic action can also be depicted, where each node represents an event with precondition and postcondition labels, and edges depict indistinguishability relations.

Definition 3. *Given an epistemic state s and action a the product update $s \otimes a$ defines a new epistemic state $s' = \langle \mathcal{M}', W'_d\rangle$, such that,*

- $W' = \{(w, e) \in W \times E \mid \mathcal{M}, w \models pre(e)\}$
- $\sim'_i = \{((w, e), (w', e')) \in W' \times W' \mid w \sim_i w', e \sim_i e'\}$
- $V'((w, e)) = \{p \in P \mid post(e) \models p$ *or* $(\mathcal{M}, w \models p$ *and* $post(e) \not\models p)\}$
- $W'_d = \{(w, e) \in W' \mid w \in W_d$ *and* $e \in E_d\}$

events from designated ones in epistemic states and actions.

To exemplify these concepts, we will explore examples grounded in our use case scenario.

Example 1. *Consider our use case with three cubes. (Note that we use the same illustration to convey our points on DEL-related concepts; later on, we will see that planning for such scenarios is significantly more involved when we introduce false beliefs.) Suppose the status of the task is in a situation s_i shown in Figure 3, in which c_r is inside box_1 and both the boxes are opaque, and the robot holding c_y and the human comes back with c_w, and assess the situation. We assume that the human can see the robot holding c_y.*

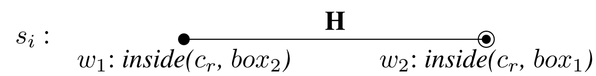

Figure 3: *It represents the epistemic state $s_i = \langle \mathcal{M}, \{w_2\}\rangle$. Worlds and accessibility relations are represented by nodes and edges, respectively, such that a designated world is shown by a double circle. This depicts that $s_i \models K_{\boldsymbol{R}} inside(c_r, box_1)$, but concerning the human partner, $s_i \models \neg K_{\boldsymbol{H}} inside(c_r, box_1) \wedge \neg K_{\boldsymbol{H}} inside(c_r, box_2)$.*

Example 2. *The next state $s_{i+1} = s_i \otimes a_i$ while the agents are aware of the action the robot will execute in s_i is a_i which is placing c_y in the correct box. We describe how s_{i+1} looks like when the human is co-present or when not with the robot during execution (Bolander 2014):*

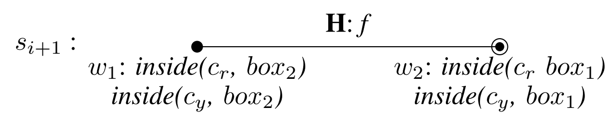

Figure 4: *The next state s_{i+1}, resulting from the robot placing c_y in box_1 (the correct box). The indistinguishability relation is only for humans when the formula f that is not(copresent(H, R)) holds. The robot always knows that the designated world is w_2. For simplicity, common facts for both the worlds like* opaque(box_1) and opaque(box_2) *are not shown.*

We now describe perspective shift or perspective taking. For a state $s = \langle \mathcal{M}, W_d\rangle$ and agent j, it represents $s^j = \langle \mathcal{M}, \{v \mid (w, v) \in \sim_j \text{ and } w \in W_d\}\rangle$. For example, $s_i^{\mathbf{H}} = \langle \mathcal{M}, \{w_1, w_2\}\rangle$, represents how the epistemic state s_i looks from the human's perspective.

Human-Aware Task Planning

The synopsis of the human-aware task planning (HATP) paradigm is discussed here. HATP/EHDA (Buisan et al. 2022) comprises a dual Hierarchical Task Network (HTN (Ghallab, Nau, and Traverso 2004)) based task specification model. It is a recently proposed planner that **e**stimates and emulates **h**uman **d**ecisions and **a**ctions for HRC, formalized in (Favier, Shekhar, and Alami 2023a). The language that follows is easier to understand by adhering to the latter.

Definition 4. ***(Human-Aware Task Planning Problem.)*** *The HATP problem, which extends HTN specifications, is a 2-tuple $\mathcal{P}_{rh} = (\langle s_0^r, tn_{r,0}, D_r\rangle, \langle s_0^h, tn_{h,0}, D_h\rangle)$ where s_0^r is the initial belief state of the robot, while s_0^h is the belief state human begins with. Here, $tn_{r,0}$ is the initial task network that the robot has to solve, similarly $tn_{h,0}$ —for the human. D_r (D_h) represents the domain available for the robot (human) containing its operators and methods.*

Each agent has their action model, task network (agenda), plan, and triggers, denoted as $\mathcal{P}_{rh} = \langle \mathcal{M}_R, \mathcal{M}_H\rangle$ (Buisan 2021). (Other components of $\mathcal{P}_{rh}$ are ignored for brevity in Definition 4.) The robot with an estimated human model $\mathcal{M}_H$, also estimates what it believes s_0^r. We consider this estimation the ground truth in the reference of the planner (or, in the context of DEL, the robot's *knowledge*) versus what the robot estimates to be believed by the human s_0^h by perspective taking. The latter may include facts that are not true from the robot's perspective (called false beliefs), which can be corrected, setting this framework apart from Fully Observable Non-Deterministic (FOND) planning[2], in principle. This distinction will become clearer as we progress.

The framework uses agents' action models and beliefs to decompose agents' task networks into primitive tasks (actions). The planning scheme assumes that a single agent decides to act at a time and which action it performs, and uses specific actions to synchronize the agent's plans. First, it builds the whole search space by considering all possible, feasible decompositions. Then, considering plan evaluation with action and social costs, it can adapt off-the-shelf search algorithms to determine the best robot policy.

Definition 5. ***(Implicitly Coordinated Joint Solution.)*** *The solution for $\mathcal{P}_{rh}$, is represented as a tree or a graph, i.e. $G = (V, E)$. Each vertex ($v \in V$) represents the robot's belief state, starting from the initial belief. Each edge ($e \in E$) represents a primitive task that is either a robot's action o^r, or a human's estimated and emulated action o^h. G gets branched on the possible choices (o_1^h, o_2^h, ..., o_m^h).*

Its branch is a sequence of primitive actions, say $\pi = (o_1^r, o_2^h, o_3^r, ..., o_{k-1}^h, o_k^r)$, that must satisfy all the solution conditions. Here, each o_i^h represents a decision the human could make, frequently from a range of options. This is the factor that determines the robot's execution strategy.

A planner extending HATP/EHDA has been proposed recently. This enhancement enabled the planner to effectively anticipate humans' incorrect beliefs and ensure a smooth collaboration (Favier, Shekhar, and Alami 2023b). Implicitly coordinated plans that incorporate both robot and human actions can be generated by the planner.

To accomplish that, the authors modeled situation assessment processes based on co-presence within the HATP/EHDA's planning workflow, thus providing a symbolic approach for specifying these abilities. It enabled the planner to be more pertinent to capture what agents can observe and infer in their surroundings. Due to this, the planning process assesses the detrimental effects of humans' incorrect beliefs on the task at hand. As a result, the robot's plan can be to communicate minimally and proactively. The current limitation of this approach is that it systematically communicates all those facts that may have a bad impact, and there is no way humans can initiate the communication. Moreover, if this false belief is the result of an unobserved robot action, the robot's plan might be to delay this action until the human observes it, preventing the incorrect belief from being formed. The planner was shown to be effective.

[2] It extends classical planning by addressing events beyond control, where actions lead to a set of potential outcomes (Daniele, Traverso, and Vardi 1999; Muise, Belle, and McIlraith 2014). While actions in FOND planning are non-deterministic, their outcomes become observable once executed, allowing agents to adapt their strategies based on observed results.

The HAETP Planning Framework

We consider that the human maintains an estimated model for the robot $\mathcal{M}_H^R$, which can be incorrect compared to $\mathcal{M}^R$. We observe that while the majority of the model's components remain static during the planning phase, the task network tn_φ and agent beliefs $Bel(\varphi)$ are dynamic, with φ denoting an agent (or agent perspective). Moreover, in line with the described previous research, we consider that only $Bel(H)$ and $Bel(R_H)$ can contain false beliefs, while other components, e.g., agents' action models and task networks, are accurately estimated.

Our framework considers all three models $\mathcal{M}^R$, $\mathcal{M}^H$, and $\mathcal{M}_H^R$, is called *human aware epistemic task planning (HAETP)*. $\mathcal{M}^R$ is used to plan the robot's actions, while $\mathcal{M}^H$ to estimate and emulate humans' decisions and actions, while $\mathcal{M}_H^R$, to predict the possible actions the robot could do from human's perspective in their presence or absence.

Planning Workflow

Roughly, the new planning system works as follows: We focus only on the dynamic parts of what follows. An *epistemic state*, s_0 is provided as an input to our system. Each world w_j in an epistemic state s_i represents $\langle (Bel(R), tn_r), (Bel(H), tn_h), (Bel(R_H), tn_{r_h})\rangle$. It also includes the only designated world w_d always known to the robot. We note that these worlds are indistinguishable from the human point of view, but human knows that the robot can distinguish them and that the robot can identify w_d. Additionally, humans also believe that, if w_j is the real designated world, then $Bel(R_H)_{ij}$, is the reality as they do not have access to $Bel(R)_{ij}$. In this paper, we consider that $Bel(H)$ is equal to $Bel(R_H)$, but they can be different from $Bel(R)$, hence can contain false beliefs.

The robot, a particular epistemic state s_i, and possible worlds w_j are considered. We compute the set of all possible primitive actions, computed by all allowed decompositions, based on $(Bel(R), tn_r)_{ij}$, and whether it is different than the set of primitive actions based on the allowed decompositions w.r.t. $(Bel(R_H), tn_{r_h})_{ij}$. The goal is to align these decompositions in a way that the human can correctly estimate the progress the robot may achieve, hence utilizing the human's capacity for anticipating. If there is a difference, we identify the *relevant* facts in $Bel^{ij}(R)$ that need to be corrected in $Bel^{ij}(R_H)$, to make the decompositions similar. To achieve that, we adapt to what is being followed in (Favier, Shekhar, and Alami 2023b). That is, one can plan minimal communication, possible to schedule ahead of time during offline planning when communication is allowed. Eventually, communication will also fix $Bel^{ij}(H)$, accordingly. However, $Bel^{ij}(H)$ and $Bel^{ij}(R_H)$ can still have non-relevant false beliefs compared to the ground truth ($Bel^{ij}(R)$) with respect to the world w_j.

Next, the planner computes the robot's next real action based on its task network tn_R^{id} in the designated world w_d of s_i, we call it the *designated* event. It also computes other non-designated events based on respective decompositions in each world w_j of s_i. An event and a possible real action including *noop*s, are the same and are used interchangeably. In other words, the planner computes a set of all possible decompositions based on what humans can anticipate, that means by taking into account $(Bel(R_H), tn_{r_h})_{ij}$. These are all the anticipated events that can take place due to the robot is acting and the execution is not shared. Next, all these decompositions (*i.e.*, the set of the first primitive action in each refinement) together form an epistemic action, a_i.

Executing an Epistemic Action in a State: Following the formula provided for the *cross-product operation* ($\otimes$), it computes $s_{i+1} = s_i \otimes a_i$. We model within the planning algorithm (Algorithm 1, Line 8) as, if human is *co-present* – an idea which is adopted from the literature and will be described later, then they can distinguish between the designated action performed by the robot with the other possible actions, otherwise humans consider each as the robot's next possible action, and accordingly the indistinguishability relations are managed in s_{i+1}. Note that, within each world of the new epistemic state, individual beliefs, i.e., $Bel(R)$, $Bel(H)$, and $Bel(R_H)$ are updated corresponding to the possible robot action (either *real* or *anticipated*) that is a part of epistemic action a_i. Also, the task networks concerning $\mathcal{M}^R$ and $\mathcal{M}_H^R$ are updated in each world, accordingly.

When humans are co-present, they assess the execution of robot actions through *situation assessment* process, which further narrows down the possibilities. However, this narrowing effect can also be seen as a direct consequence of the cross-product operation ($\otimes$), where co-presence is managed at the representation level (Bolander 2014).

Now, we similarly focus on the human partner. They act only if they believe that their next real ontic action, corresponding to a possible decomposition, is applicable in all possible worlds. That means, for each w_j in s_{i+1}, applicability is examined in $(Bel(H), tn_h)_{i+1,j}$. At this stage, there are two issues: humans can act based on a false belief (if consistent through all the worlds), or a true belief w.r.t. the ground truth in every w_j. We handle the false belief scenario the way it is addressed in the past work, that is, by finding out relevant belief divergence and handling it via *communication* and/or *delay* (Favier, Shekhar, and Alami 2023b).

However, we also know that facts that they are uncertain about, which are true in some worlds and not in others, are due to actions performed by the robot in their absence. If such a fact is a part of the precondition for task refinements, then humans can initiate communication, or the robot can inform them. And, if co-present, the robot can act in reality to implicitly share this fact.

A joint solution that is implicitly coordinated in relation to the task is the result of this framework (Def. 5).

The discussion in this section highlights the focal point: for each world in an epistemic state, denoted by $\langle (Bel(R), tn_r), (Bel(H), tn_h), (Bel(R_H), tn_{r_h}) \rangle$, our attention in what follows is directed towards the scenario where $Bel(R) = Bel(H) = Bel(R_H)$. Addressing the case of $Bel(R) \neq Bel(H) = Bel(R_H)$ can be achieved through technical adjustments of established methodologies from prior research. However, if all three beliefs are distinct, it would present a non-trivial challenge beyond the scope of this paper and we leave it for future work.

AND/OR Search based HAETP Planner

Terminologies Used

We are interested in domains in which some world's properties can be observed directly by being present in the relevant part of the world. For example, if box_1 is transparent, **H** can observe *inside(c_r, box_1)* by visiting mt and is not needed to be co-present with **R** when it places c_r in box_1. However, if box_1 is opaque, humans can assess the red cube being placed in box_1 by the robot and determine that *inside(c_r, box_1)* holds. If human misses this action execution, they may need another means to know that *inside(c_r, box_1)* holds.

In the same way, we use the definition of the observation process from (Shekhar et al. 2023). We say that assessing the status of an environmental feature depends on a broader context to determine whether it can be observed, or can only be evaluated by attending the action execution changing it. Knowledge rules were used to address this aspect. For example, an agent can view the current status of the variable *inside*(c_r, box_1) as *true* if they meet the requirements of a rule's antecedent formula, such as being at the main table, box not being opaque, and c_r being inside box_1. If the antecedent includes a dynamic variable that holds in the first scenario but not the second, then certain rules may apply in one and not the other.

Definition 6. *The* situational assessment *(SA) process considers the observation process described above and an epistemic state s_i, producing an updated epistemic state s_i'. This process iterates over each world w_j in s_i, removing it if it can be distinguished from w_d by the human.*

Consider the following scenario: let $w_1 = \langle (...), (\{inside(c_r, box_1)\}, ...), (...) \rangle$ and $w_2 = \langle (...), (\{inside(c_r, box_2)\}, ...), (...) \rangle$, where w_1 and w_2 represent distinct worlds within an epistemic state s_i, with w_1 as the designated world. When boxes are transparent and the human is co-located with the main table, the updated epistemic state s_i' contains only w_1.

Our Planning Algorithm

Algorithm 1 takes the HAETP problem as input and produces an output as either a *failure* or a *joint solution* with the optimal-worst case plan. We propose that primitive agent actions, including auxiliary actions, e.g., *NOOP*, are instantaneous and of equal cost at the moment, as in classical

Algorithm 1: *AND/OR Planner using Breadth-First Search. Two key subroutines are* Situation Assessment *and* Expand.

1: **Input:** *A HAETP task*
2: **Output:** *A joint solution* or *failure*
3: $root\_epi\_state \leftarrow \langle \mathcal{M}, w_d \rangle$ ▷ (focusing just on the essential parts) each world in $w \in W$ contains $(\langle s_0^r, tn_{r,0}, D_r\rangle, \langle s_0^h, tn_{h,0}, D_h\rangle, \langle s_0^{r_h}, tn_{r_h,0}, D_{r_h}\rangle)$
4: $queue$.enqueue($root\_epi\_state$)
5: **while** $queue$ is not empty **do**
6:     $curr\_node' \leftarrow$ queue.dequeue()
7:     $curr\_node \leftarrow$ *Situation Assessment($curr\_node'$)*
8:     $successors \leftarrow$ *Expand($curr\_node$)*
9:     **if** $successors \neq \emptyset$ **then**
10:         **for** successor **in** successors **do**
11:             $queue$.enqueue($successor$)
12:         **end for**
13:     **else**
14:         *eval*($curr\_node$) ▷ assign it *DONE* or *DEAD*
15:         *propagate_revised_status*($curr\_node$)
16:     **end if**
17:     **if** *root_solved*($root\_epi\_state$) **then**
18:         **return** *extract_joint_solution()*
19:     **end if**
20: **end while**
21: **return** *failure*

planning. Our algorithm is an implementation of the classic AND/OR search algorithm using rooted graphs. Following the search and assuming that the *root* node is *DONE*, the joint solution policy is extracted, *extract_joint_solution()*. This process is delayed until the end (Line 18).

We consider the root node ($root\_epi\_state$) and the subsequent actor, either **R** or **H**, to begin the plan exploration (Line 3). Within the loop, in Line 6, we select a node/state from the $queue$, and next call the ***Situation Assessment( )*** subroutine. What we present in Line 7 is a *lazy* approach for doing situation assessment. At this stage, the planner already knows whether agents were co-present and whether the designated action could be assessed by the human. This helps the planner ignore those worlds that can be distinguished from the designated world (*Definition 6*). The situation where a human *moves* to *place* where the robot is present and then emerges as co-present is especially well-suited for it. Another significant subroutine, ***Expand( )***, which we previously discussed in the HAETP framework's planning workflow, is called in Line 8. The children created after the robot agent expands this popped node are *AND* nodes, *OR* nodes, and vice versa for the case where the human agent expands this node.

If there are no successors for the current node, it indicates either a goal node or a dead end. In Line 14, we evaluate the current node, s_i, where s_i roughly captures $\{\langle (Bel(R), tn_r), (Bel(H), tn_h), (Bel(R_H), tn_{r_h})\rangle\}$. If both tn_r and tn_h are fully decomposed in the designated world of s_i, we execute an auxiliary action with a precondition that the task network is solved. If both agents can execute this action, it signifies that both the human and the robot are aware that the shared task has been achieved. (Line 15) We propagate the status of this node to its immediate parent, who then propagates its status to its parent depending on whether it is an *AND* node or an *OR* node.

(Lines 17 & 18) We verify whether the root node is *DONE*. If the task is achieved, we extract the joint solution.

The Post-processing Step Once we have an executable AND/OR policy, we post-process it depending on whether the agents are co-present. When agents are co-present, we follow a turn-taking approach, but when they are not co-present, we parallelize their actions. This involves executing the AND/OR joint policy, and identifying where agents separate and reunite. We then group all the actions in between to form pairs of human and robot actions. It is important to note that in the original policy, one agent always waits while the other acts. This post-processing step ensures that actions performed in parallel when agents are apart, do not interact and can occur simultaneously and independently.

Related Work

Human Robot Collaboration (HRC): Automated planning the robot's behavior while also considering the existence of a human partner, known as human-aware planning and decision-making (Cirillo, Karlsson, and Saffiotti 2009; Cramer, Kellens, and Demeester 2021; Unhelkar, Li, and Shah 2020; Lallement, de Silva, and Alami 2018), parallel (online) planning and dispatching plans (Bezrucav and Corves 2020), and negotiating role allocation (Roncone, Mangin, and Scassellati 2017). Our work aligns with this research focus, dealing with human-aware planning and decision-making in collaborative human-robot scenarios. However, we are not aware of existing approaches that explicitly consider the human's anticipation abilities when direct experience is not shared in the environment during collaboration on a shared task.

Models and Planning Approaches: Many planning models are applied in the context of HRC planning, including HTNs (Lallement, de Silva, and Alami 2018; Roncone, Mangin, and Scassellati 2017), POMDPs (Partially Observable MDPs) (Unhelkar, Li, and Shah 2019; Roncone, Mangin, and Scassellati 2017; Unhelkar, Li, and Shah 2020; Görür et al. 2017), AND/OR graphs (Darvish et al. 2021), etc. HTNs use both abstract and non-abstract tasks to form a hierarchical network, while AND/OR graphs cover the causal links among subtasks (Gombolay et al. 2016), and depth-first search is used for planning.

Theory of Mind in HRC: Several variants of ToM are explored in executing shared global plans. However, the main focus lies on perspective-taking, where a robot reasons about what humans can perceive in the environment. This involves constructing a world from the human's perspective and managing the agents' beliefs accordingly while executing (Berlin et al. 2006). The framework proposed in (Devin and Alami 2016) enables the robot to estimate the mental state of the human at execution time, encompassing their beliefs, actions, goals, and plans. This framework facilitates the execution of shared plans in an object manipulation context and illustrates how a robot can adapt to human decisions and actions, and use communication when necessary.

This work serves as a loose inspiration for integrating Theory of Mind (ToM) into the decision-making process of social robots, as shown in (Görür et al. 2017), to better adapt to stochastic intentions, behaviors, and expectations over a series of repeated interactions.

Epistemic Planning: The DEL-based epistemic planning framework, as demonstrated in (Bolander, Dissing, and Herrmann 2021), holds promise for capturing key elements of ToM in autonomous robots. This framework lays the groundwork for implicit coordination through perspective shifts in human-robot collaboration. By adopting this planning framework and focusing on the robot's perspective, it could serve as a basis for addressing the core problem we aim to solve with the shared mental model (Nikolaidis and Shah 2012), albeit without considering false beliefs.

Explainable AI Planning (XAIP): In general, XAIP deals with human-aware systems *explaining* some aspect of their behavior during plan generation or execution (Kambhampati et al. 2022). Humans can have a disparate robot model ($\mathcal{M}_R^H$ for $\mathcal{M}_R$). The model reconciliation approach (proposed in (Sreedharan, Chakraborti, and Kambhampati 2021)) avoids soliloquies, considering the exact differences between the two models to generate needed explanations only. They proposed a planning approach to compute the optimal explanations. We adopt a similar approach, thus computing relevant divergences to communicate only what is necessary to align the decompositions.

Empirical Evaluation

We implemented our planning system using Algorithm 1 in Python. It is based on the latest version of HATP/EHDA code (Buisan et al. 2022).

As far as we are aware, there are no baseline planners to compare. But, when it makes sense, we do compare our planner to the one from (Favier, Shekhar, and Alami 2023b).

We conduct an initial assessment of our planner within the context of the cube organization domain. The algorithm highlights that the rapid growth in the size of the epistemic state in terms of the number of worlds directly correlates with the *number of actions* (K) the robot can perform when the execution experience is not shared. Furthermore, the sequencing of these actions significantly influences the range of potential worlds the human might expect to encounter.

We consider the parameter K, to assess its effect on the planner's overall performance. During planning, we assume that whenever the shared execution experience is disrupted, the robot can execute a maximum of K real actions, with the option of not acting at all. For example, when the human is away to fetch the cube and has a *fixed* length and sequence of actions to perform. The exact number of real ontic actions the robot performs ranging from 0 to K, including which of those allowed ones and their exact sequence, will depend on the scenario at hand, environment dynamics (e.g., the observability factor), and the optimization criteria. Currently, all the actions are of unit cost and instantaneous during planning. The option for the robot to limit its real actions whenever required is integrated into the task description, aligning with the turn-taking nature of the underlying

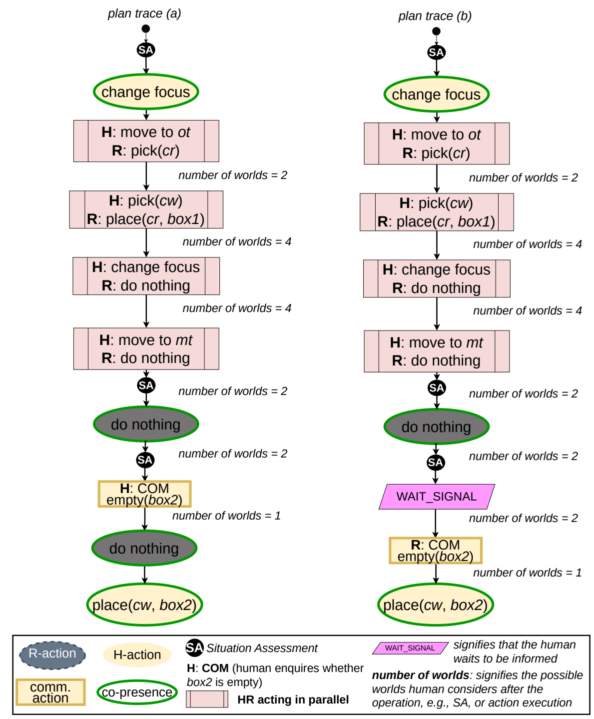

Figure 5: *Two branches from an AND/OR joint solution are shown: (a)* ***R*** *informs* ***H*** proactively*, thus leaving only the designated world for them to continue with* place(c_w, box_2)*. (b)* ***R*** waits *to inform* ***H*** *about the condition* empty(box_2).

planner. Consequently, the planner is engineered to optimize the robot's policy tree branching on uncontrollable human choices, which may include a communication action initiated by either the human or the robot, to meet our objective.

Qualitative Analysis

We explore different plan traces the planner can come up with depending on scenarios that arise. We start with two cubes, c_r and c_w, placed initially on tables mt and ot, respectively. Initially, there is only one designated world, w_d, in the initial epistemic state, s_0. The environment otherwise remains unchanged. The human can decide to go and retrieve the white cube, while the robot begins to work on other part(s) of the shared task.

Figure 5 shows two plan traces from an AND/OR joint plan tree. **H** starts to execute. **H** & **R** are co-present and the boxes are opaque. The SA process is a systematic subroutine of our planner, but it is depicted only at relevant places.

Let us focus on (a): after the human shifts focus to ot, both agents are not co-present until they reunite later in the trace, during which they act simultaneously. (*In this situation, agents must be at the same table and simultaneously focus on it to be considered co-present.*) In the first broad rectangular box, the human moves to ot. They anticipate that the robot may have picked c_r or done nothing, but in reality, the robot picks c_r, resulting in two possibilities that will be maintained within the robot. Similarly, in the following box, the human picks c_w at ot and anticipates that if the robot

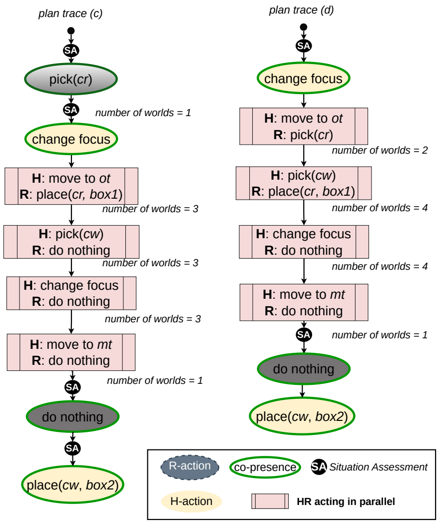

Figure 6: *Two branches are shown: (c)* ***R*** *starts the process, and (d)* ***H*** *takes the lead. Boxes are transparent. We can see that as soon as* ***H&R*** *become co-present again, SA by perspective taking ignores the impossible worlds, immediately.*

had picked c_r, it could have placed it in one of the boxes or held onto it, or c_r is still on the table. Together, these create four possibilities, with the reality being that c_r is inside box_1. At this point, the robot currently has no feasible action to execute, and the shared task has been not achieved yet, too. Upon the human's return, as per their initial agreement on K, the robot has prepared itself with four possible worlds (with a designated world that only the robot knows). Perspective-taking and situation assessment help the robot eliminate two worlds where c_r is not on ot or in its hand.

At this stage, we present two approaches to proceed with the task. The richer representation used in planning allows them. In trace (a), the robot waits for human inquiry, while in trace (b), humans wait to receive information (e.g., through nodding or eye contact with the robot). Consequently, the robot decides to inform that box_2 is empty, resulting in only the designated world remaining. Here, $empty(box_2)$ is a precondition for the human to place c_w in it, which is true in one world and not in the other. With our proposed solution, when the human waits to be informed, the robot can also choose to take action to modify the physical environment, conveying a variable's value if the execution is attended. This aspect is intended to be addressed in the future.

In our three-cube scenario, if the red cube is already in box_1 and the robot is holding the c_y, it can choose to place the c_y in box_1 in the presence of the human. This action results in the creation of a state with only the designated world as the next action ordered in the task network (tn_{r_h}) of that world does not allow the robot to execute *place*(c_y, box_1). The robot can only be clever if it can fully explore its options. Depending on the situation, it might not always be preferable to place the yellow cube while the human is away and rely on communication or other means later on.

In contrast, in (Favier, Shekhar, and Alami 2023b), the robot informs after they become co-present again, assuming that the human can choose to place the c_w in box_1 due to the outdated (false) belief w.r.t. changes they missed when the execution history was not shared.

Figure 6 shows plan traces for the case when the boxes are transparent. Their explanations follow a similar pattern as discussed in the case where the boxes were opaque.

inst	K	*comm*	*#states*	$\lvert W \rvert$	*#leaves*	*time (ms)* $\times 10^5$
P1 (2,2,T)	2	N	218	4	3	0.089
P2 (2,2,O)	2	Y	236	4	3	0.141
P3 (3,2,T)	2	N	1643	7	6	5.906
P4 (3,2,O)	2	Y	2003	7	6	9.816
P5 (3,2,T)	4	N	4107	14	5	99.81
P6 (3,2,O)	4	Y	5607	14	5	125.3

Table 1: *The planner's performance is evaluated on different metrics.* inst *is instance description. Whether communication is employed –* comm. *The metrics include the total number of states explored (*#states*), the worst-case number of worlds ($\lvert W \rvert$) evaluated in a state, the number of traces (*#leaves*), and the execution* time *(measured in 10^5 ms).*

Quantitative Results and Analysis

We refer to Table 1. The first column indicates the instance number, along with the count of cubes and boxes, and whether the boxes are transparent (T) or opaque (O), respectively. In each instance, at least one cube is positioned on the other table, which the human needs to retrieve to fulfill the objective. We show how the factor K influences the runtime. Additionally, we present the number of states expanded for each instance and, in the worst-case scenario, the count of worlds evaluated in an epistemic state. The number of branches in the obtained optimal AND/OR joint solution tree is also provided. Our observation is that both $|W|$ and K contribute to longer runtimes. Instances requiring communication tend to take slightly longer compared to those where communication is not required.

Conclusion

We note that our framework allows the robot to implement a ToM not only at execution time but also at planning time and hence explores what would be the beliefs of the human and the robot depending on which course of action. This is done thanks to the use of epistemic reasoning, the notion of shared experience, and observable and non-observable facts, which allow anticipation of human situation assessment along the various non-deterministic shared **H&R** plan traces.

In the future, we plan to evaluate our planner in diverse domains and develop faster search methods for improved scalability. While assuming $Bel(H) = Bel(R_H) \neq Bel(R)$ as in this paper, can solve many scenarios, we intend to consider distinct $Bel(R)$, $Bel(H)$, and $Bel(R_H)$, and extend our framework to allow planning with that.

We aim to adapt this framework *"to explain"* the robot's actions in different scenarios to its human partner. Note that, in this work, we assumed that except for the $Bel(.)$ components, other components of $\mathcal{M}_R^H$ are correctly estimated (and are similar to those of $\mathcal{M}_R$).

Acknowledgment

This work was supported by the Horizon Europe Framework Programme Grant: euROBIN 101070596 and Agence Nationale de la Recherche Grant: ANITI ANR-19-PI3A-0004.

References

Baratta, A.; Cimino, A.; Gnoni, M. G.; and Longo, F. 2023. Human Robot Collaboration in Industry 4.0: a literature review. *Procedia Computer Science*, 217: 1887–1895.

Berlin, M.; Gray, J.; Thomaz, A. L.; and Breazeal, C. 2006. Perspective taking: An organizing principle for learning in human-robot interaction. In *Proc. of AAAI*.

Bezrucav, S.-O.; and Corves, B. 2020. Improved AI planning for cooperating teams of humans and robots. In *Proc. of the PlanRob Workshop at ICAPS'20*.

Bolander, T. 2014. Seeing is Believing: Formalising False-Belief Tasks in Dynamic Epistemic Logic. In *Proc. of ECSI*, volume 1283 of *CEUR Workshop Proceedings*, 87–107.

Bolander, T.; and Andersen, M. B. 2011. Epistemic planning for single and multi-agent systems. *J. Appl. Non Class. Logics*, 21(1): 9–34.

Bolander, T.; Dissing, L.; and Herrmann, N. 2021. DEL-based Epistemic Planning for Human-Robot Collaboration: Theory and Implementation. In *Proc. of KR*.

Buisan, G. 2021. *Planning For Both Robot and Human: Anticipating and Accompanying Human Decisions.* Ph.D. thesis, INSA Toulouse, France.

Buisan, G.; Favier, A.; Mayima, A.; and Alami, R. 2022. HATP/EHDA: A Robot Task Planner Anticipating and Eliciting Human Decisions and Actions. In *Proc. of ICRA*.

Cirillo, M.; Karlsson, L.; and Saffiotti, A. 2009. A Human-Aware Robot Task Planner. In *Proc. of ICAPS 2009*.

Coupeté, E.; Moutarde, F.; and Manitsaris, S. 2015. Gesture recognition using a depth camera for human robot collaboration on assembly line. *Procedia Manufacturing*, 3: 518–525.

Cramer, M.; Kellens, K.; and Demeester, E. 2021. Probabilistic Decision Model for Adaptive Task Planning in Human-Robot Collaborative Assembly Based on Designer and Operator Intents. *IEEE Robotics AL*, 6(4): 7325–7332.

Daniele, M.; Traverso, P.; and Vardi, M. Y. 1999. Strong cyclic planning revisited. In *European Conference on Planning*, 35–48. Springer.

Darvish, K.; Simetti, E.; Mastrogiovanni, F.; and Casalino, G. 2021. A Hierarchical Architecture for Human-Robot Cooperation Processes. *IEEE Trans. Robotics*, 37(2): 567–586.

Devin, S.; and Alami, R. 2016. An implemented ToM to improve human-robot shared plans execution. In *HRI Proc.*

Dissing, L.; and Bolander, T. 2020. Implementing Theory of Mind on a Robot Using DEL. In *Proc. of IJCAI*.

Engesser, T.; Bolander, T.; Mattmüller, R.; and Nebel, B. 2017. Cooperative Epistemic Multi-Agent Planning for Implicit Coordination. In *Proc. of M4M@ICLA 2017*.

Favier, A.; Shekhar, S.; and Alami, R. 2023a. Anticipating False Beliefs and Planning Pertinent Reactions in Human-Aware Task Planning with Models of Theory of Mind. *PlanRob 2023, The ICAPS Workshop*.

Favier, A.; Shekhar, S.; and Alami, R. 2023b. Models and Algorithms for Human-Aware Task Planning with Integrated Theory of Mind. In *RO-MAN*, 1279–1286. IEEE.

Ghallab, M.; Nau, D. S.; and Traverso, P. 2004. *Automated Planning - Theory and Practice*. Elsevier.

Gombolay, M. C.; Jensen, R.; Stigile, J.; Son, S.; and Shah, J. A. 2016. Apprenticeship Scheduling: Learning to Schedule from Human Experts. In *Proc. of IJCAI*.

Görür, O. C.; Rosman, B.; Hoffman, G.; and Albayrak, S. 2017. Toward integrating Theory of Mind into adaptive decision-making of social robots to understand human intention. In *the workshop on the role of intentions in HRI'17*.

Kambhampati, S.; Sreedharan, S.; Verma, M.; Zha, Y.; and Guan, L. 2022. Symbols as a Lingua Franca for Bridging Human-AI Chasm for Explainable and Advisable AI Systems. In *AAAI*, 12262–12267. AAAI Press.

Lallement, R.; de Silva, L.; and Alami, R. 2018. HATP: hierarchical agent-based task planner. In *Proc. of AAMAS*.

Liang, C.-J.; Wang, X.; Kamat, V. R.; and Menassa, C. C. 2021. Human–robot collaboration in construction: Classification and research trends. *Journal of Construction Engineering and Management*, 147(10): 03121006.

Muise, C.; Belle, V.; and McIlraith, S. 2014. Computing contingent plans via fully observable non-deterministic planning. In *Proc. of AAAI 2014*.

Nieto Agraz, C.; Pfingsthorn, M.; Gliesche, P.; Eichelberg, M.; and Hein, A. 2022. A survey of robotic systems for nursing care. *Frontiers in Robotics and AI*, 9: 832248.

Nikolaidis, S.; and Shah, J. 2012. Human-robot teaming using shared mental models. *ACM/IEEE HRI*.

Roncone, A.; Mangin, O.; and Scassellati, B. 2017. Transparent role assignment and task allocation in human robot collaboration. In *Proc. of ICRA*.

Semeraro, F.; Griffiths, A.; and Cangelosi, A. 2023. Human–robot collaboration and machine learning: A systematic review of recent research. *Robotics and Computer-Integrated Manufacturing*, 79: 102432.

Shekhar, S.; Favier, A.; Alami, R.; and Croitoru, M. 2023. A Knowledge Rich Task Planning Framework for Human-Robot Collaboration. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*.

Sreedharan, S.; Chakraborti, T.; and Kambhampati, S. 2021. Foundations of explanations as model reconciliation. *Artif. Intell.*, 301: 103558.

Unhelkar, V. V.; Li, S.; and Shah, J. A. 2019. Semi-Supervised Learning of Decision-Making Models for Human-Robot Collaboration. In *Proc. of CoRL*.

Unhelkar, V. V.; Li, S.; and Shah, J. A. 2020. Decision-Making for Bidirectional Communication in Sequential Human-Robot Collaborative Tasks. In *Proc. of HRI*.

Wimmer, H.; and Perner, J. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1): 103–128.