

# Decision-Making for Bidirectional Communication in Sequential Human-Robot Collaborative Tasks

Vaibhav V. Unhelkar\*

unhelkar@csail.mit.edu

Massachusetts Institute of Technology  
Cambridge, Massachusetts

Shen Li\*

shenli@csail.mit.edu

Massachusetts Institute of Technology  
Cambridge, Massachusetts

Julie A. Shah

julie\_a\_shah@csail.mit.edu

Massachusetts Institute of Technology  
Cambridge, Massachusetts



**Figure 1: The costs and benefits of execution-time communication for human-robot collaboration. (left) On the one hand, communication is useful for sharing intent and improving coordination among teammates [6]. (right) On the other hand, unchecked use of communication can overload the human’s limited attentional resources, thereby negatively impacting their performance and trust [35]. How should the robot decide if, when, and what to communicate when collaborating with humans?**

## ABSTRACT

Communication is critical to collaboration; however, too much of it can degrade performance. Motivated by the need for effective use of a robot’s communication modalities, in this work, we present a computational framework that decides *if, when, and what* to communicate during human-robot collaboration. The framework, titled CommPlan, consists of a model specification process and an execution-time POMDP planner. To address the challenge of collecting interaction data, the model specification process is *hybrid*: where part of the model is learned from data, while the remainder is manually specified. Given the model, the robot’s decision-making is performed computationally during interaction and under partial observability of human’s mental states. We implement CommPlan for a shared workspace task, in which the robot has multiple communication options and needs to reason within a short time. Through experiments with human participants, we confirm that CommPlan results in the effective use of communication capabilities and improves human-robot collaboration.

## KEYWORDS

planning under uncertainty; human-in-the-loop learning; human-robot communication

\*Both authors contributed equally to this research.

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 '20, March 23–26, 2020, Cambridge, United Kingdom

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-6746-2/20/03...\$15.00

<https://doi.org/10.1145/3319502.3374779>

## ACM Reference Format:

Vaibhav V. Unhelkar, Shen Li, and Julie A. Shah. 2020. Decision-Making for Bidirectional Communication in Sequential Human-Robot Collaborative Tasks. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20)*, March 23–26, 2020, Cambridge, United Kingdom. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3319502.3374779>

## 1 INTRODUCTION

The evolutionary origins of human communication are grounded in cooperative activity [57]. Indeed, communication has evolved to be critical for collaboration. Members of successful human teams exhibit anticipatory information-sharing strategies to accomplish collaborative tasks [9, 13, 35]. For robots to become beneficial members of human teams, thus, they must be capable of effectively communicating with their human teammates.

In this work, we present models and algorithms for leveraging effective communication to achieve fluent human-robot teamwork. We focus on sequential collaborative tasks with known task dynamics and objectives. Examples of such human-robot collaborative tasks are found across domains, including in healthcare [25, 33], education [45], manufacturing [67], and outer space [17].

**Research Question.** Recognizing the importance of communication for human-robot interaction (HRI), significant research focusing on human-robot communication has been conducted in the last two decades [10, 11, 14, 15, 27, 37, 40, 56, 61]. Several modalities, both verbal and non-verbal, are being developed for enabling the sharing of information among humans and robots [1, 5, 18, 28, 29, 36, 53, 55, 64]. Depending on the collaboration context, these modalities can enable robots to either convey information to the human, interpret the information received from the human, or both (i.e., bidirectional communication).

Despite the active focus on the development of communication modalities, there exists a relative lack of approaches that enable a robot to utilize its communication capability purposefully [37]. As exemplified in Fig. 1, while modalities are a prerequisite for communication, purposeful use of the modality is essential for achieving the motivation behind their development, namely, fluent collaboration. Communication provides benefits for teamwork, but also incurs costs [9, 13, 35, 60]. Thus, while the development of a modality addresses the question, “*How to communicate?*”, an equally important question that remains to be addressed is “*If, when, and what to communicate?*”

**Solution Overview.** The primary contribution of this work is a framework, titled CommPlan, that resolves this question for a subclass of sequential collaborative tasks (described in Sec. 3) and a variety of communication types (described in Sec. 4). The research question is mathematically formalized in Sec. 5. As depicted in Fig. 2, CommPlan consists of a model specification process and an execution-time planner to compute the robot’s policy for collaboration. Both actions and communications affect the task progress and the human’s mental states. Thus, CommPlan jointly reasons about the robot’s actions and communication to arrive at its policy.

The model specification process, detailed in Sec. 6, requires the robot developer (person/s tasked with specifying the robot’s policy) to specify modules for the task, human behavior, and communication capability. Given these specifications, CommPlan analytically generates a decision-making model for the robot and its policy (see Sec. 7). To consider the impact of human’s latent mental states, CommPlan uses the framework of partially observable Markov decision processes (POMDPs). As described in Sec. 8, via experiments with human participants, we demonstrate the utility of CommPlan for sequential human-robot collaborative tasks with multiple communication types and short planning times ( $< 1$  s).

## 2 RELATED WORK

Multiple HRI user studies have helped identify the utility of communications for fluent collaboration [48, 51, 52, 61]. These studies, which use a hand-crafted communication policy, motivate the design of computational approaches for effective human-robot communication. Kaupp et al. provide one of such first computational approaches. Mavridis and Dong integrate planning for motor control and speech to address the question “to ask or to sense?” Both approaches consider tasks where the robot is requesting help from the human. In contrast, this article examines more general mixed-initiative tasks.

Communication during mixed-initiative tasks has also been considered in recent research [10, 15, 40, 61]. Here, we briefly discuss approaches that explicitly model the human’s mental states for robot’s communication decision-making. Devin and Alami provide a framework to implement theory-of-mind (ToM) and improve the execution of human-robot shared plans. In their approach, the estimate of mental model divergence is used to reduce communication overhead through generalizable rules for decision-making.

The approach closest to ours is that of Nikolaidis et al. Their approach introduces *compliance*, an important latent parameter for modeling the impact of robot communications on human behavior, and performs planning using a mixed-observability MDP.

However, their method considers a single communication type (either state-conveying actions or commands) during task execution. In contrast, CommPlan considers multiple communication types, thereby enabling planning for bidirectional communication. Further, we extend the formalization of compliance and demonstrate communication planning at execution time and for collaborative tasks with state spaces significantly larger than the prior art.

## 3 TASK MODEL

We utilize a multi-agent variant of the Markov decision process (MDP) to describe the human-robot collaborative tasks of interest<sup>1</sup> [42]. The choice of an MDP-based model is guided by our focus on sequential tasks, the broad expressivity of Markovian models, and their heritage in computational HRI research [8, 19, 41, 58]. Specifically, we describe the collaborative tasks of interests using the following factored, multi-agent MDP (MMDP).

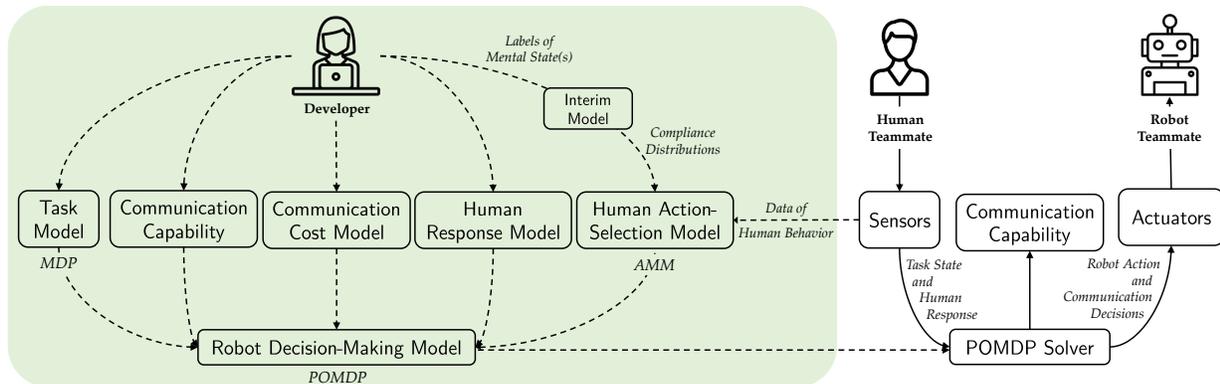
- the task consists of two agents, one human (denoted by the subscript  $H$ ) and one robot (denoted by the subscript  $R$ );
- $S$  denotes the set of states  $s \doteq (s_H, s_R, s_E)$ ; the state is factored:  $s_H$ , denoting the human-specific variables,  $s_R$ , denoting the robot-specific variables, and  $s_E$  denoting remaining variables;
- $A \doteq A_H \times A_R$  denotes the set of joint actions  $a \doteq (a_H, a_R)$ , where  $A_H$  and  $A_R$  denote the sets of human actions ( $a_H$ ) and robot actions ( $a_R$ ), respectively;
- the state dynamics are Markovian and are described by the transition model  $T(s'|s, a) : S \times A \times S \rightarrow [0, 1]$ ;
- similar to the state, the transition model  $T$  is factored, i.e.,  $T \doteq T_H(s_H|s, a) \cdot T_R(s_R|s, a) \cdot T_E(s_E|s, a)$ ; the factors  $T_H$ ,  $T_R$  and  $T_E$  describe the dynamics of  $s_H$ ,  $s_R$  and  $s_E$ , respectively;
- the team receives a shared reward at each timestep specified by the reward function,  $R(s, a)$ ;
- $\gamma \in [0, 1]$  denotes the discount factor; and
- (optionally)  $h \in \mathbb{Z}^+$ , denotes the finite time horizon.

The team’s objective is to maximize the expected cumulative reward. We emphasize that the task model is a specification of the collaborative scenario. Importantly, it does not prescribe how the human or robot should act or collaborate. Consequently, the MMDP state consists of observable variables; it does not model the teammate’s mental states. Further, the action space is limited to task-specific actions that affect the MMDP state and, generally, does not include communications. Communications influence mental states (e.g., belief or intent) and are modeled in the following sections.

*Scope.* We limit our scope to collaborative tasks that can be modeled by the task model. While the human and the robot cannot observe each other’s mental states, by definition of an MMDP, both have full observability of the task state  $s \doteq (s_H, s_R, s_E)$ .

**Example Task.** We utilize the shared workspace task [22], instantiated in a kitchen environment, as an example of collaborative tasks that can be represented using the MMDP model. Consider a human-robot team preparing meals in a kitchen (see Fig. 1 and Appendix D). Each meal includes one sandwich and one cup of juice. Key landmarks in the kitchen include two cooking stations (denoted as L1 and L2), two wrapping stations (L3 and L4), four cups (one each at L1-4), a pantry (L5), and juice refill area (L6).

<sup>1</sup>Please see Appendix A for a list of symbols used in this work.



**Figure 2: CommPlan, a framework for generating a robot’s policy (action and communication decisions) for human-robot collaboration. (left, model specification process) Prior to task execution, the developer specifies five modules using data, domain expertise, and learning algorithms. These modules are analytically combined to arrive at the robot’s decision-making model. (right, execution-time planner) During task execution, the robot computes its policy using sensors, the decision-making model, and a POMDP solver. The policy is executed using the robot’s actuators and communication modality.**

The objective of the team is to prepare four meals as soon as possible. At an abstract level, the task requires the teammates to visit the landmarks and perform temporally-extended activities (such as make a sandwich, wrap a sandwich, pour juice). For instance, preparing a sandwich requires picking ingredients from the pantry (L5), assembling them at the cooking station (either L1 or L2), and then wrapping them at a task-specific wrapping station (L3 after L1 or L4 after L2). The dexterous human is tasked with preparing the sandwiches, while the robot is tasked with pouring juice into cups. However, owing to constraints of the shared workspace, the human and robot cannot work at the same landmark at the same time (e.g., if the robot is pouring juice into the cup at L1, the human cannot use the cooking station at L1). Thus, the agents have to coordinate their motion and activities to complete the task efficiently.

*Task Specification.* Teammates’ location and task progress are essential features for specifying the task objective (safety and efficiency in a shared workspace). Thus, in our implementation and experiments, the meal preparation task is modeled as follows.

- $s_H$  represents human position in a 2-dimensional grid;
- $s_R$  represents robot’s joint angles and its motion progress;
- $s_E$  tracks which and how many cups have been filled;
- $A_H$  includes human’s motion primitives in the 2-D grid (e.g., left, right, up, down, and wait).
- $A_R$  includes both robot’s motion primitives (in the joint angle space) and task-specific macro-actions (e.g., for pouring juice);
- $T_H$  and  $T_R$  represent the motion dynamics of the human and robot, respectively.  $T_E$  represents task recipes.
- $R$ : To encode efficiency, the team receives a negative reward at each timestep until task completion. In addition, to emphasize safety, the reward penalizes human-robot collisions.

We use the meal preparation task as a running example and the task for the experiments. However, the discussion applies to other domains, such as assembly tasks in factories or medicine delivery in hospitals. For the experiments, the task is described by an MMDP with 39, 816 states.

*Solving the MMDP.* Given the task specification, planning algorithms can be used for *solving* the MMDP and generating a joint human-robot team policy [42]. This team policy can inform task allocation and execution [7, 15, 32]. However, despite the specification of team policy, the human and robot often have significant flexibility while doing the task, thereby necessitating modeling of preferences and mental states [8, 19, 41, 58]. Further, the utility of directly utilizing the MMDP team policy is limited for HRI, since it requires the human to memorize and follow the joint policy in lock-step with the robot. This is an unrealistic requirement and limits human autonomy. Thus, approaches that enable a robot to coordinate its actions with the human, without limiting human autonomy, are necessary. As detailed next, communication can address this requirement and facilitate the coordination among autonomous robots and humans, who are accomplishing sequential tasks.

## 4 COMMUNICATION CAPABILITY

Having described the collaborative task model, we discuss the different communication types for which CommPlan enables communication decision-making. CommPlan is designed around communication types (i.e., the type of information being communicated) and, thus, is applicable across communication modalities (i.e., it does not limit how the information is being communicated).

*Communication Types.* Guided by the context, multiple taxonomies for communication types have been developed across research on human-human, robot-robot, and human-robot communication [4, 9, 16, 26, 34, 43, 61, 62, 65]. Informed by these taxonomies, in this work, we consider the following types of communications:

- **inform**, the robot informs the human about a latent state of its decision-making; since the task state is observable, this type corresponds to the robot’s mental state (e.g., robot’s intent  $s_{RI}$ );
- **ask**, the robot queries the human about the latent state of her decision-making; since the task state is observable in our model, the ask message corresponds to a question regarding the human’s mental state (such as human’s intent); and

- **command**, the robot commands the human to perform a specific action or plan; this communication type can also be used to represent requests or suggestions; and
- **answer**, the robot answers the human’s question.

We do not explicitly consider the type *answer* in our decision-making model. Instead, we assume that readily answering the human’s queries is optimal in collaborative tasks. By considering these communication types, CommPlan is capable of reasoning about bidirectional human-robot communications. We demonstrate CommPlan using speech as the communication modality; however, it is equally suitable for other modalities that map to the above communication types. For instance, the robot can inform the human of its intention using visual signals or text.

A challenge while developing a modality is that of symbol grounding [12, 54]. Computational approaches that facilitate grounding have been developed to address this challenge [3, 21, 63]. We view symbol grounding as a critical but complementary problem. Hence, here, we assume that both the human and the robot can associate the messages to the attributes of the collaborative task without any ambiguity. In our demonstrations, we achieve this by utilizing a finite communication vocabulary with hand-crafted grounding.

**Communication Space.** Based on these communication types, the developer needs to specify a set of communications that the robot and the human can make during the task execution. We denote the finite set of robot’s communications  $a_{RC} \in A_{RC}$  and that of human’s communications  $a_{HC} \in A_{HC}$ .

*Example.* For instance, in the example task, knowledge of teammate’s intent can be useful. Thus, we specify the following set of robot communications  $a_{RC} \in A_{RC}$  using task-specific actions and key landmarks of the shared workspace,

- (inform) “I am going to do *action* at *landmark*.”
- (ask) “Where are you going?”
- (command) “Please make the next sandwich at *landmark*.”

Similarly, to model the human’s response, the set of human communications is specified as “I am going to *landmark*.” The human can provide this information as a response to the robot’s question (i.e., as an answer) or offer it on her own accord (i.e., inform). Grounding for the key landmarks (L1-6) is manually specified using the MMDP state variables (specifically,  $s_H$  and  $s_R$ ) to enable communication understanding. Similarly, action grounding is available from the MMDP action space.

## 5 PROBLEM STATEMENT

Given the specification of the collaborative task and communication space, our objective is to arrive at the robot policy  $\pi_R$  that maximizes the expected cumulative reward of the human-robot team. The policy  $\pi_R$  is formalized as a function that maps robot’s sensory information to its action  $a_R$  and communication  $a_{RC}$  decisions.

## 6 MODEL SPECIFICATION PROCESS

The MMDP task model depends on the team’s joint actions. However, to computationally *solve* for the robot’s policy, we need a single-agent model that only depends on robot actions. Here, we discuss the steps involved in specifying this model, which we refer to as the robot’s decision-making model. To streamline the

specification process, CommPlan requires the developer to specify three additional modules, namely, for communication cost, human’s action-selection (i.e., policy for  $a_H$ ), human’s communication response (i.e., policy for  $a_{HC}$ ).

These modules can be specified either entirely manually, learned from data, or both. To reduce the effort required for prototyping and developing a collaborative robot, learning models from data is desirable. However, collecting *communication-based* interaction data (which is essential for learning approaches) is challenging in the real world. Hence, a hybrid approach – where part of the model is learned, while the remainder can be manually specified using the developer’s domain knowledge – is emphasized. The specification process, detailed next, is also summarized in Appendix B.

### 6.1 Communication Cost Model

Research on human teams and interruption management has identified various latent causes of communication costs for humans [24, 35]. Consequently, communication cost is often task- and context-specific, and novel cost models are needed for HRI.

As one solution, for CommPlan, we provide a parametric model to specify communication costs. For a particular collaborative task, the developer can specify these parameters based on domain knowledge to inform the robot’s decision-making. In order to stress the cost of too many communications, the proposed cost model is nonlinear. Specifically, the model requires three parameters: minimum and maximum cost of communication ( $\rho_a, \rho_b$ ), and the duration of the interruption interval ( $h_{RC}$ ). The communication reward (negative cost)  $R_c$  is then given as follows:

$$R_c = -\rho_a \exp \left[ \left( 1 - \frac{t_{RC}}{h_{RC}} \right) \log \rho_b \right] \quad (1)$$

where  $t_{RC}$  denotes the time since the last communication. The interruption interval begins after communication is made and lasts for a duration of  $h_{RC}$ . For communication outside this interval, the cost corresponds to the minimum  $\rho_a$ . The model penalizes multiple communications during the interruption interval.

Depending on the task and context, the cost could be tailored to different communications by specifying different parameters for each  $a_{RC}$ . For instance, to emphasize polite communications during equal partners teamwork, a developer could specify a higher cost for commands as compared to suggestions. Similarly, the parameters of cost models could be tuned to account for cultural and lingual differences [2, 20, 39, 49]. We reiterate that the proposed model is just one alternative to specifying cost and highlight that modeling communication cost is a crucial but under-explored problem in HRI. However, irrespective of the model choice, we posit that the nonlinear characterization of communication cost is critical when the team is performing a sequential task.

*Example.* For the example task, we specify  $h_{RC} = 3$  s. The minimum and maximum costs are set as  $\rho_a = 1$  and  $\rho_b = 15$ , based on their relative magnitude to the task reward. Finally, we tailor the cost model for *inform*, in which the robot commits to action and landmark (i.e., intent). To emphasize that the robot follows through its commitment, denoted as  $s_{RI}$ , the model includes a penalty if the robot does not complete its communicated commitment and a positive reward (negative cost) if it does.

**Communication State.** Additional states are required to enable Markovian decision-making with a temporal model of communication cost. We denote this additional state as communication state,  $s_{RC} \in S_{RC}$ . The communication state includes the latest robot communication, the time since last communication  $t_{RC}$ , and the robot’s committed intent  $s_{RI}$ .

## 6.2 Human Action-Selection Model

The need and utility of anticipating the human teammate’s actions ( $a_H$ ) have been demonstrated across multiple HRI studies [56]. Consequently, a model of the human’s behavior is essential for making effective communication decisions. Further, multiple data-driven approaches have been utilized to learn this model, including variants of supervised learning and inverse reinforcement learning [31, 40, 68]. CommPlan includes an approach for specifying human action-selection model that builds upon this prior work.

However, instead of relying on data alone, CommPlan enables the developer to hand-craft part of the model for which training data might be insufficient or unavailable. The ability to combine data and domain expertise is especially relevant for communication decision-making. In practice, it is challenging to collect interaction data for training models, even more so when the interaction involves communications. Further, collecting training data for communication-based interaction requires the robot to follow a communication policy. However, training with a standard communication policy is also undesirable for HRI, as it can negatively impact the user’s trust and subjective perception of the collaborative robot.

**Representation.** We utilize the Agent Markov Model (AMM) to represent human’s sequential decision-making behavior [59]. AMM represents the behavior of an agent whose decisions are contingent on both observable ( $s_F$ ) and latent ( $x_H$ ) decision factors, and the dynamics of the decision factors are Markovian. Our choice for using the AMM is motivated by the need to model human’s mental states and their dynamics (for which the AMM includes explicit parameters). Further, the AMM can be learned using both data and domain expertise [59]. Formally, an AMM is defined by the tuple  $(X_H, S_F, A_H, b_X, T_X, T_F, \pi_H)$ , where

- $X_H$  denotes the state space of human’s mental states  $x_H$ ;
- $S_F$  denotes the space of human’s observable states  $s_F$ ;
- $A_H$  denotes the set of human’s task-specific actions  $a_H$ ;
- $b_X$  denotes the probability of the initial latent state;
- $T_X$  and  $T_F$  denote the transition models of  $x_H$  and  $s_F$ ; and
- $\pi_H \doteq \pi_H(a_H|s_F, x_H)$  represents the human’s policy, i.e., the probability of choosing action  $a_H$  in the state  $(s_F, x_H)$ .

**AMM Feature Specification.** In a collaborative task, the human’s behavior depends on the task-specific features (i.e., the MMDP state), mental states, actions, and communications. The robot has full observability of the MMDP state, its own actions, and communications. Hence, the specification of the human’s observable decision factors  $s_F$  and their dynamics  $T_F$  is obtained from the MMDP state, robot actions, and communications, i.e.,  $s_F = g(s, a_R, s_{RC}, a_{RC})$  and  $S_F \subseteq S \times A_R \times S_{RC} \times A_{RC}$ . The function  $g$  is an arbitrary nonlinear function, which provides the developer flexibility while specifying features influencing human behavior. AMM’s action space ( $A_H$ ) is also available from the task MMDP. The parameter  $X_H$  enables the

developer to specify important latent features (i.e., mental states) that affect human behavior. Thus, using the task definition and domain knowledge, the developer can partially specify the AMM tuple, specifically  $(X_H, S_F, A_H, T_F)$ .

*Example.* In the meal preparation task, human behavior is landmark driven. The human teammate first collects the ingredients from the pantry (L5), then assembles them at the cooking station (either L1 or L2), and then wraps them at L3 or L4. Thus, we identify these task-specific landmarks (or subgoals) as an important mental state ( $x_H$ ) for human decision-making. In other tasks, mental states may include variables such as attention, workload, and belief. The set of all the subgoals (L1-5) specifies the latent state space  $X_H$ . In addition to her subgoal, the teammate’s behavior depends on her position ( $s_H$ ) as well as robot’s communications ( $a_{RC}$ ). This completes the specification of both known decision factors  $s_F$  as  $(s_H, a_{RC})$  and latent decision factors  $x_H$  as subgoals. The human’s action space  $A_H$  is known from the MMDP definition.

**Hybrid Approach for AMM Learning.** Given the partial AMM tuple and interaction data, Unhelkar and Shah provide algorithms to recover the complete AMM tuple. The interaction data corresponds to execution traces of human behavior, namely, time series of  $s_F$ ,  $a_H$ , and (optionally)  $x_H$ . The complete AMM tuple specifies the human’s policy ( $\pi_H$ ) and temporal dynamics ( $T_X$ ) of the mental states. For the example task, this corresponds to learning how the human chooses her next subgoal ( $x'_H$ ) and actions ( $a_H$ ) based on previously completed subgoal ( $x_H$ ), current position ( $s_H$ ), and robot communications ( $a_{RC}, s_{RC}$ ). However, as noted earlier, collecting data of communication-based interaction is challenging. Hence, in CommPlan, we adopt a hybrid two-step approach. First, we learn an interim model that does not capture the influence of robot communication on human behavior. The interim model is learned using interaction data without communication. Next, we augment this model by manually specifying the effect of robot communications on human behavior.

**Interim Model.** To learn the interim model, CommPlan requires interaction data without communications. Specifically, execution traces of  $s_F$  with  $a_{RC} = s_{RC} = \emptyset$ , where  $\emptyset$  corresponds to no communication. The dataset can be optionally labeled to include sequences of the latent state  $x_H$ . Given this dataset and the partial AMM tuple, the interim model is generating algorithmically using Bayesian approaches for learning the AMM [59].

*Example.* We collect a dataset of humans performing the meal preparation task without any communications, i.e., time series of  $(s_H, a_H)$ -tuples. The dataset is collected from three humans and includes twelve training sequences and three test sequences. For all sequences, we manually provide subgoal labels  $x_H$ . Given this dataset, the interim model that captures the human’s landmark-driven behavior is obtained computationally via Bayesian AMM learning [59]. Due to our choice of AMM as the representation, the learned model includes the subgoal dynamics (i.e., probability of human’s subgoal selection) and policy for achieving a subgoal, denoted as  $T_X(x'_H|x_H, s_{F\emptyset}, a_H)$  and  $\pi_H(a_H|s_{F\emptyset}, x_H)$ , where  $s_{F\emptyset} = g(s, a_R, s_{RC} = \emptyset, a_{RC} = \emptyset)$ . We note that the learned parameters of the interim model  $T_X$  and  $\pi_H$  are independent of  $a_{RC}$  and, thus, do not include the effect of communications.

**Complete Model.** Communications can affect human behavior either directly (via the policy) or indirectly (by changing the human’s mental states, which in turn affects human’s choice of actions). Both of these effects can be specified in CommPlan. For instance, when the robot requests the human to perform an action, the effect is modeled by the impact of  $a_{RC}$  on policy  $\pi_H$ . In contrast, when the robot requests the human to choose a subgoal, the effect is captured by the impact of  $a_{RC}$  on latent state dynamics  $T_X$ . Thus, mathematically, the objective of this step is to augment the terms of the interim model  $(T_X, \pi_H)$ , where  $s_{RC} \neq \emptyset$  or  $a_{RC} \neq \emptyset$ .

We note that  $T_X$  and  $\pi_H$  are probability distributions (denoted as *p.d.*) over the human’s mental states and actions, respectively. Thus, the effect of  $a_{RC}$  corresponds to the probability of the human changing her behavior based on the robot’s communication. This probability is called *compliance*, which has been previously formalized for the communication type `command` [40]. In the prior art, compliance is considered identical for all communications and task states. However, in practice, the effect of each communication is different and depends on the task context. We extend the formalization of compliance to other communication types and, through the specification of  $s_F = g(s, a_R, a_{RC}, s_{RC})$ , allow for it to vary based on the task state (via  $s$ ) and communication  $a_{RC}$ .

*Example.* We discuss the compliance specification process for the example task before summarizing its generic version. In the example task, the robot can make three types of communication: `command`, `inform`, `ask`. Due to the task structure, the human will be able to follow the `command` (of making a sandwich) only if she currently has access to the ingredients (i.e., the human is at the pantry). Thus, we specify a context-specific model of compliance, i.e., `command` affects human only if  $s_H$  is `pantry`. Further, when the human is at the pantry, she will follow the `command` with probability  $p_{c1}$ . Moreover, by incorporating communication state  $s_{RC}$  as one of the decision factors  $s_A$ , a developer can specify that a message has a temporally-extended influence on human behavior. This is especially important in sequential tasks and to account for communication delay. For the communication type `inform`, the robot shares its intent (subgoal). We model that after learning about the robot’s subgoal, the human will act to improve the fluency of collaboration and not choose the same subgoal with probability  $p_{c2}$ . The probabilities  $p_{c1}$  and  $p_{c2}$  can be different, resulting in a context- and communication- specific compliance model. The `ask` type influences human’s response  $a_{HC}$  but not her actions  $a_H$  and, thus, does not require a compliance model.

*Compliance Specification Process.* Thus, in general, for a communication  $a_{RC}$  that exerts indirect influence on human behavior, the developer needs to specify a *p.d.* over the human’s next mental state  $x'_H$  when the robot makes the communications  $a_{RC}$  for AMM state  $(x_H, s_F)$  and action  $(a_H)$ . We denote this *p.d.* as  $q_c$ , which is used to augment the model as follows:  $T_X(x'_H|x_H, s_F, a_H) = q_c$ . The distribution  $q_c$  mathematically represents the extended formalization of compliance, which is both context- and communication-specific and can be temporally extended. Further, these tuples need only be specified for the communications that influence human’s task-specific behavior. A similar procedure is used for communication with direct influence. However, in this case, the compliance distribution  $q_c$  partially specifies the human’s policy  $\pi_H$ .

**Summary.** Given the compliance specifications and the interim model, the specification of the human action-selection model is complete. When the robot is silent (i.e.,  $s_{RC} = a_{RC} = \emptyset$ ), human behavior is specified by the interim model. When the robot communicates, the manually provided specification is used.

### 6.3 Human Response Model

The human teammate can communicate in addition to performing task-specific actions  $a_H$ . Thus, to communicate effectively, the robot also requires a model of the human’s communication decision-making. In this work, we provide a model for human’s response (i.e., communication types `inform` and `answer`), which is then incorporated into the robot’s decision-making.

In response to the robot’s query (type `ask`), the human may respond correctly, incorrectly, or not respond at all. For instance, if the human’s attention is overloaded with the collaborative task, the robot’s question may go unanswered. We represent this behavior using the property *responsivity*, which is summarized by the probability  $p_r$  of the human responding to a query. In collaborative tasks, we seldom expect the human to provide an incorrect response; however, due to implementation challenges (such as natural language parsing), the robot might receive an incorrect response. Thus, we further model that the sensed response from the human is truthful with probability  $p_t$ . Finally, we consider that the human may share her intention with the robot preemptively (i.e., without being asked) with probability  $p_a$ . For the example task, we use the following values:  $p_r = 0.9$ ,  $p_t = 0.9$ , and  $p_a = 0.01$ .

## 7 PLANNING ROBOT BEHAVIOR

After the model specification process, CommPlan computationally generates the robot behavior without any additional manual effort. First, a POMDP is derived as the robot’s decision-making model. Next, using the decision-making model and an execution-time planner, CommPlan arrives at the robot’s policy  $(\pi_R)$ , for its actions  $(a_R)$  and communications  $(a_{RC})$ .

**Robot’s Decision-Making Model.** Analytically derivation of the robot’s decision-making model (POMDP) is described as follows. The POMDP parameters are denoted using the subscript  $p$ .

*State Space:*  $S_p$  denotes the set of robot’s decision-making states,  $s_p \doteq (s_H, s_R, s_E, s_{RC}, x_H) = (s, s_{RC}, x_H)$ . The POMDP state includes five factors, namely, the three components of the task MMDP state  $s$ , the communication state  $s_{RC}$ , and the latent state of human decision-making  $x_H$ . The inclusion of human’s latent state enables anticipation of human’s actions and generation of adaptive robot behavior, which is an essential feature for fluent collaboration [23].

*Action Space:*  $A_p \doteq A_R \cup A_{RC}$  denotes the set of actions  $a_p$ , where  $A_R$  and  $A_{RC}$  denote the sets of robot’s task-specific actions  $a_R$  and communications  $a_{RC}$ , respectively. In this way, the planning is done jointly for the robot’s actions and communications. Note that the action space does not include human actions and communications, as the robot cannot control them directly.

*Transition Model:* The POMDP state dynamics are Markovian and are denoted by  $T_p(s'_p|s_p, a_p) : S_p \times A_p \times S_p \rightarrow [0, 1]$ . For the factored POMDP state, the transition model is derived as follows,

$$T_p(s'_p|s_p, a_p) \doteq T_{ps}(s'|s_p, a_p)T_{pc}(s'_{RC}|s_p, a_p)T_{px}(x'_H|x_H, s_A) \quad (2)$$

A transition model of  $s$  is available from the task model,  $T(s'|s, a)$ . However, the task model depends on the joint action  $a \doteq (a_H, a_R)$ . In contrast, the robot’s decision-making model requires a transition function that only depends on the robot actions and communications. Thus, using the human model (AMM) and the task model (MMDP), the transition model of the variable  $s$  for robot decision-making is derived analytically as follows,

$$T_{ps}(s'|s_p, a_p) = \sum_{a_H \in A_H} T(s'|s, a_R, a_H) \pi_H(a_H|s_A, x_H) \quad (3)$$

The transition model of human’s mental state  $x_H$  is also available from the AMM; however, it too depends on human actions. Thus, CommPlan analytically obtains  $T_{px}$  as follows,

$$T_{px}(x'_H|x_H, s_A) = \sum_{a_H} T_x(x'_H|x_H, s_A, a_H) \pi_H(a_H|s_A, x_H) \quad (4)$$

Finally,  $T_{pc}$  is the transition model of the robot’s communication state that keeps track of the robot’s communicated intent, latest communication, and time since last communication.

**Reward Model:** The robot’s objective is to maximize the team’s shared reward by effectively utilizing its actions and communications. To achieve this objective, the reward  $R_p(s_p, a_p) : S_p \times A_p \rightarrow \mathbb{R}$  is obtained by adding two components: the task reward  $R_{ps}$  and the communication cost  $R_{pc}$ . Similar to the computation of  $T_{ps}$ , the task reward  $R_{ps}$  is obtained analytically using the human model (AMM) and the task model (MMDP). The model for communication cost is described in Sec. 6.1, i.e.,  $R_{pc} = R_c$ .

**Observation Model:** During the task, the robot has full observability of the MMDP state  $s$  and its own communications  $s_{RC}$ . Hence, the human’s mental state  $x_H$  is the partially observable component of the POMDP state. The observation model of  $x_H$  is specified as the human’s response model, which is described in Sec. 6.3. Thus, through its communication the robot can possibly influence human behavior and receive observations of the latent state. Finally, the time horizon  $t_f$ , and the discount factor  $\gamma$  of the decision-making model (POMDP) are identical to that of the task model (MMDP).

**Robot’s Policy.** Given the decision-making model, the robot’s policy ( $\pi_R$ ) is generated algorithmically by solving the POMDP. To enable communication decision-making for large problems, we utilize a variant of DESPOT, an online POMDP solver [58, 66]. The online solver uses the robot’s observations, actions, and communications to reduce the robot’s uncertainty in the human’s latent state (i.e., explore) and improve the team’s shared reward (i.e., exploit). As summarized in Appendix C, planning is done during interaction and jointly over the robot’s actions and communications.

## 8 EXPERIMENTS

We implement and demonstrate CommPlan using a collaborative robot (Universal Robot 10 with Robotiq gripper) and evaluate it with human participants. To our knowledge, our approach is the first to perform decision-making for multiple communication types while anticipating the human teammate’s behavior and latent states. Hence, we compare it against a domain-specific hand-crafted policy. Through the experiments, we evaluate the following hypotheses: *During the human-robot collaboration, a robot that reasons about*

*its multiple communications results in a shared team reward higher than (H1) a robot that does not communicate, and (H2) a robot that uses a developer-specified communication policy.*

**Methodology.** The human-robot team collaboratively performed the meal preparation task, which was used as the running example, during the experiments. The robot had to reason about a large state space and make decisions within a planning time of 0.3 s. The experimental setup and procedure are detailed in Appendix D.

**Design and Baselines.** We utilize a within-subject design with one independent variable: the approach to communication decision-making. We consider three treatment levels: no-communication policy (denoted as Silent), developer-specified communication policy (denoted as Hand-crafted), and POMDP-based communication policy generated using CommPlan. All three approaches use an identical specification of task model (MMDP), human’s latent state ( $x_H$ ), and interim human model (AMM without communication). However, they differ in their approach to the robot’s decision-making.

For both the baselines, Silent and Hand-crafted, the robot’s action is selected using CommPlan’s POMDP but with an empty communication space, i.e.,  $A_{RC} = S_{RC} = \emptyset$ . For Silent, the robot remains silent during the task. For Hand-crafted, the following domain-specific communication policy is used

- ask when the human’s intent is ambiguous ( $\max Pr(x_H) < 0.35$ );
- inform when the robot has selected its next activity;
- command if the robot anticipates the human and robot activities to interfere and the human is at the pantry; and
- stay silent during the interruption interval.

For both Hand-crafted and CommPlan, the communication capability of the robot ( $A_{RC}$ ) is identical. However, while using CommPlan, the robot can communicate any of its messages  $a_{RC}$  at any point during the task. It can even repeat messages if it chooses to. Thus, it requires the capability to answer “if, when, and what to communicate” and utilize its communications effectively. Hence, for CommPlan, both the action and communication are selected using the POMDP model and solver described in Sec. 7. This POMDP model with communication, used by CommPlan, has  $\approx 31$  million states  $s_p$ , 13 actions  $a_p$ , and a timestep (planning time) of 0.3 s.

**Results.** We next report the results of the study, summarized in Table 1, for interaction with 15 participants (5 male, 10 female, median age: 26 years). Six participants reported prior experience with robots. Before discussing the quantitative results, we qualitatively describe the interaction observed in the experiments. A video from the experiments is available at <http://tiny.cc/CommPlan>

**Qualitative Behavior with CommPlan.** During the collaborative task, the robot needs to make a slew of action and communication decisions. Specifically, to complete the meal preparation task safely and efficiently, the robot’s policy is needed to determine

- which of the four cups should be filled next;
- whether to wait to ensure safety or to move to complete the task;
- if it chooses to move, its trajectory to reach the cup;
- whether to use its communication modality; and
- which communication message to convey.

While using CommPlan, the robot makes all of these decisions jointly and, as the experiments demonstrate, effectively.

	Shared Reward	Task Time (s)	Number of Robot Comm.
Silent	-292.2 ± 24.4	95.4 ± 4.0	0 ± 0
Hand-crafted	-312.2 ± 19.5	98.7 ± 3.8	4.4 ± 0.1
CommPlan	<b>-236.4 ± 11.2</b>	<b>88.4 ± 2.8</b>	5.4 ± 0.3

**Table 1: Summary statistics from the experiments** ( $N = 15$ ).

For example, in the supplementary video, the robot communicates at the start of the interaction and requests human to make the next sandwich at location 2 (L2). This decision is made to ensure safety and efficiency, as the robot anticipates a motion conflict. Specifically, by weighing the cost of communication ( $-R_{pc}$ ) against the reward of completing the task ( $R_{ps}$ ), the robot deems that it is better to use communication to influence the human’s decisions. Further, after the human goes to location 2, the robot does not unnecessarily repeat this communication. This decision, too, is made algorithmically since the added benefit of repeated communication on the task reward does not outweigh its costs.

Similarly, the robot does not communicate later during the interaction (e.g., when the human goes to location 1), when it has a strong belief of where the human is making a sandwich and communication is not needed to deconflict the motion. This observed behavior is possible due to the nonlinear cost model, the predictive human model, and the online solver. Since the approach is probabilistic and model-based, rarely, the robot might make incorrect decisions. For instance, while filling the cup at location 4 in the video, the robot informs the human about its intent twice. This repetition is redundant. However, as discussed next, our approach (on average) results in useful decisions.

**Quantitative Performance.** We first compare the objective criterion stated in the problem definition, i.e., the cumulative shared reward  $R(s, a)$ . CommPlan, averaged across the participants, accrues a substantially higher cumulative reward than both the baselines. The effect of the communication decision-making approach is statistically significant ( $p = 0.01$ ) as evaluated by the Friedman test, which rendered a Chi-square value of 9.09. Next, we conduct pairwise comparisons using the Wilcoxon signed-rank test. Our approach, CommPlan, results in statistically significant differences when compared with both Silent ( $p < 0.05$ ) and Hand-Crafted ( $p < 0.01$ ) communication policies. Thus, the experiments provide evidence to support both the hypotheses  $H1$  and  $H2$ .

Similar trends are observed for task completion times ( $p = 0.11$ ). No statistically significant differences are found in the subjective measures, possibly due to the difference in the task completion times across treatments not being perceptible for the participants.

**Lessons Learned.** As reflected in the objective measures and the supporting evidence for hypothesis 1 ( $H1$ ), the experiments confirm the utility of communications for HRI. Specifically, by effectively utilizing robot communications, the POMDP-based CommPlan framework results in higher cumulative reward and lower task completion times as compared to the Silent policy.

However, as observed by the inferior performance of the Hand-crafted policy, the presence of a communication modality alone is insufficient to improve collaboration. Instead, the purposeful use of communication modality is essential to realize its benefits. Thus,

the experiments also provide evidence supporting the importance of our research question. Despite making only one more communication (on average) than the Hand-crafted policy, CommPlan accrues substantially higher reward. Further, the policy is algorithmically generated via a general framework that is applicable to other collaborative tasks and modalities.

This promising performance indicates that CommPlan is ripe for applications in other domains. The experiments also provide guidance for deploying CommPlan in these novel situations. For instance, we observe that, at times, CommPlan can choose to repeat the robot’s communication. This redundant communication may be undesirable in specific tasks and contexts. CommPlan allows the developer to de-emphasize such repetitive communications by tuning the cost model (e.g., interruption interval). However, as the communication options available to the robot increase, we qualitatively observe that specifying the cost model can be challenging. Thus, to facilitate the tuning process, algorithmic approaches that can shape the cost model will be particularly useful.

In addition to the cost model, the developer also needs to specify modules for task and human behavior. Through our experience of implementing the meal preparation task, we observe that specifying the task model was straightforward (since the task objective is quantifiable and depends on observable features, with known dynamics). However, concurrent research [46, 47, 50] underscores that effectively specifying the reward function for general tasks remains an open problem. On a related note, the parameters of the response model were specified heuristically in our implementation and kept uniform across all participants. However, in general, these parameters will not be identical for all participants. Thus, approaches that can estimate these parameters prior to task execution will be especially useful in personalizing robot behavior.

Finally, in addition to the communications implemented in the study, the participants indicated a preference for “the communication back and forth” through the open-ended survey. These responses highlight the need for further investigation into expanding robot’s vocabulary. Critical future directions include the development of (a) joint approaches for symbol grounding and communication decision-making, (b) algorithms for decision-making with incomplete task models, and (c) models of communication cost.

## 9 CONCLUSION

We provide CommPlan, a framework that enables robots to decide *if, when, and what* to communicate while performing sequential tasks with humans. Informed by the challenges of real-world HRI, CommPlan’s model specification process is both modular and hybrid. CommPlan utilizes an AMM to represent the human’s mental states and uses a POMDP to reason about uncertainty in the robot’s decision-making. Further, we introduce *responsivity* and extend the formalization of *compliance* to capture temporal, context- and communication- dependent variations in human behavior. CommPlan’s ability to jointly reason about actions and communications is demonstrated on a physical collaborative robot. Through experiments with human participants, we confirm the ability of CommPlan to realize effective communications and enable fluent human-robot collaboration, including for tasks with large state spaces, multiple communication options, and short planning times.

## REFERENCES

- [1] Henny Admoni and Brian Scassellati. 2017. Social Eye Gaze in Human-Robot Interaction: A Review. *Journal of Human-Robot Interaction (JHRI)* 6, 1 (2017), 25–63.
- [2] Sean Andrist, Micheline Ziadee, Halim Boukaram, Bilge Mutlu, and Majd Sakr. 2015. Effects of Culture on the Credibility of Robot Speech: A Comparison between English and Arabic. In *International Conference on Human-Robot Interaction (HRI)*. ACM, 157–164.
- [3] Jacob Arkin, Rohan Paul, Daehyung Park, Subhro Roy, Nicholas Roy, and Thomas M. Howard. 2018. Real-Time Human-Robot Communication for Manipulation Tasks in Partially Observed Environments. In *International Symposium on Experimental Robotics (ISER)*.
- [4] Kent Bach and Robert M. Harnish. 1979. Linguistic Communication and Speech Acts. (1979).
- [5] Kim Baraka and Manuela M. Veloso. 2018. Mobile Service Robot State Revealing through Expressive Lights: Formalism, Design, and Evaluation. *International Journal of Social Robotics (IJSR)* 10, 1 (2018), 65–92.
- [6] Michael D. Basil. 2012. Multiple Resource Theory. *Encyclopedia of the Sciences of Learning* (2012), 2384–2385.
- [7] Craig Boutilier. 1996. Planning, Learning and Coordination in Multi-Agent Decision Processes. In *Conference on Theoretical Aspects of Rationality and Knowledge*. Morgan Kaufmann Publishers Inc., 195–210.
- [8] Frank Broz, Illah Nourbakhsh, and Reid Simmons. 2013. Planning for Human-Robot Interaction in Socially Situated Tasks. *International Journal of Social Robotics (IJSR)* 5, 2 (2013), 193–214.
- [9] Abhizna Butchibabu, Christopher Sparano-Huiban, Liz Sonenberg, and Julie Shah. 2016. Implicit Coordination Strategies for Effective Team Communication. *Human Factors* 58, 4 (2016), 595–610.
- [10] Crystal Chao and Andrea Thomaz. 2016. Timed Petri Nets for Fluent Turn-Taking over Multimodal Interaction Resources in Human-Robot Collaboration. *International Journal of Robotics Research (IJRR)* 35, 11 (2016), 1330–1353.
- [11] Sonia Chernova and Andrea L. Thomaz. 2014. Robot Learning from Human Teachers. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 8, 3 (2014), 1–121.
- [12] Herbert H. Clark, Susan E. Brennan, et al. 1991. Grounding in Communication. *Perspectives on Socially Shared Cognition* 13, 1991 (1991), 127–149.
- [13] J. Michael Crant. 2000. Proactive Behavior in Organizations. *Journal of Management* 26, 3 (2000), 435–462.
- [14] Sandra Devin and Rachid Alami. 2016. An Implemented Theory of Mind to Improve Human-Robot Shared Plans Execution. In *International Conference on Human-Robot Interaction (HRI)*. ACM, 319–326.
- [15] Sandra Devin, Aurélie Clodic, and Rachid Alami. 2017. About Decisions during Human-Robot Shared Plan Achievement: Who should Act and How?. In *International Conference on Social Robotics (ICSR)*. Springer, 453–463.
- [16] Barbara Di Eugenio, Pamela W. Jordan, Richmond H. Thomason, and Johanna D. Moore. 2000. The Agreement Process: An Empirical Investigation of Human-Human Computer-Mediated Collaborative Dialogs. *International Journal of Human-Computer Studies* 53, 6 (2000), 1017–1076.
- [17] Myron A. Diftler, J.S. Mehling, Muhammad E. Abdallah, Nicolaus A. Radford, Lyndon B. Bridgewater, Adam M. Sanders, Roger Scott Askew, D. Marty Linn, John D. Yamokoski, F.A. Permenter, et al. 2011. Robonaut 2—the First Humanoid Robot in Space. In *International Conference on Robotics and Automation (ICRA)*. IEEE, 2178–2183.
- [18] Anca D. Dragan, Kenton C.T. Lee, and Siddhartha S. Srinivasa. 2013. Legibility and Predictability of Robot Motion. In *International Conference on Human-Robot Interaction (HRI)*. ACM, 301–308.
- [19] Michelangelo Fiore, Aurélie Clodic, and Rachid Alami. 2016. On Planning and Task Achievement Modalities for Human-Robot Collaboration. In *Experimental Robotics*. Springer, 293–306.
- [20] Kerstin Sophie Haring, David Silvera-Tawil, Yoshio Matsumoto, Mari Velonaki, and Katsumi Watanabe. 2014. Perception of an Android Robot in Japan and Australia: A Cross-Cultural Comparison. In *International Conference on Social Robotics (ICSR)*. Springer, 166–175.
- [21] Jun Hatori, Yuta Kikuchi, Sosuke Kobayashi, Kuniyuki Takahashi, Yuta Tsuboi, Yuya Unno, Wilson Ko, and Jethro Tan. 2018. Interactively Picking Real-World Objects with Unconstrained Spoken Language Instructions. In *International Conference on Robotics and Automation (ICRA)*. IEEE, 3774–3781.
- [22] Guy Hoffman. 2019. Evaluating Fluency in Human-Robot Collaboration. *IEEE Transactions on Human-Machine Systems (THMS)* 49, 3 (2019), 209–218.
- [23] Guy Hoffman and Cynthia Breazeal. 2007. Effects of Anticipatory Action on Human-Robot Teamwork Efficiency, Fluency, and Perception of Team. In *International Conference on Human-Robot Interaction (HRI)*. ACM, 1–8.
- [24] Eric Horvitz and Johnson Apacible. 2003. Learning and Reasoning about Interruption. In *International Conference on Multimodal Interfaces (ICMI)*. ACM, 20–27.
- [25] Mithun George Jacob, Yu-Ting Li, George A. Akingba, and Juan P. Wachs. 2013. Collaboration with a Robotic Scrub Nurse. *Commun. ACM* 56, 5 (2013), 68–75.
- [26] Matthew Johnson, Catholijn Jonker, Birna Van Riemsdijk, Paul J. Feltovich, and Jeffrey M. Bradshaw. 2009. Joint Activity Testbed: Blocks World for Teams (BW4T). In *International Workshop on Engineering Societies in the Agents World*. Springer, 254–256.
- [27] Tobias Kaupp, Alexei Makarenko, and Hugh Durrant-Whyte. 2010. Human-Robot Communication for Collaborative Decision Making — A Probabilistic Approach. *Robotics and Autonomous Systems* 58, 5 (2010), 444 – 456.
- [28] Ross A. Knepper, Christoforos I. Mavrogiannis, Julia Proft, and Claire Liang. 2017. Implicit Communication in a Joint Action. In *International Conference on Human-Robot Interaction (HRI)*. ACM, 283–292.
- [29] Thibault Kruse, Patrizia Basili, Stefan Glasauer, and Alexandra Kirsch. 2012. Legible Robot Navigation in the Proximity of Moving Humans. In *Advanced Robotics and its Social Impacts (ARSO)*. IEEE, 83–88.
- [30] Paul Lamere, Philip Kwok, Evandro Gouvêa, Bhiksha Raj, Rita Singh, William Walker, Manfred Warmuth, and Peter Wolf. 2003. The CMU SPHINX-4 Speech Recognition System.
- [31] Przemyslaw A. Lasota and Julie A. Shah. 2017. A Multiple-Predictor Approach to Human Motion Prediction. In *International Conference on Robotics and Automation (ICRA)*. IEEE, 2300–2307.
- [32] M. Leonetti, L. Iocchi, Anthony G. Cohn, and Daniele Nardi. 2019. Adaptive Human-Aware Task Planning. In *ICAPS Workshop on Planning and Robotics (PlanRob)*.
- [33] Jing Li, Zhong Shen, Wen Yu Tian Xu, Walter Yu Hang Lam, Richard Tai Chiu Hsung, Edmond Ho Nang Pow, Kazuhiro Kosuge, and Zheng Wang. 2019. A Compact Dental Robotic System Using Soft Bracing Technique. *IEEE Robotics and Automation Letters (RA-L)* 4, 2 (2019), 1271–1278.
- [34] Sirui Li, Weixing Sun, and Tim Miller. 2015. Communication in Human-Agent Teams for Tasks with Joint Action. In *Coordination, Organizations, Institutions, and Norms in Agent Systems (COIN)*. Springer, 224–241.
- [35] Jean MacMillan, Elliot E. Entin, and Daniel Serfaty. 2004. Communication Overhead: The Hidden Cost of Team Cognition. *Team Cognition: Process and Performance at the Inter and Intra-Individual Level* (2004).
- [36] Cynthia Matuszek, Evan Herbst, Luke Zettlemoyer, and Dieter Fox. 2013. Learning to Parse Natural Language Commands to a Robot Control System. In *International Symposium on Experimental Robotics (ISER)*. 403–415.
- [37] Nikolaos Mavridis. 2015. A Review of Verbal and Non-Verbal Human-Robot Interactive Communication. *Robotics and Autonomous Systems* 63 (2015), 22–35.
- [38] Nikolaos Mavridis and Haiwei Dong. 2012. To Ask or To Sense? Planning to Integrate Speech and Sensori-motor Acts. In *International Congress on Ultra Modern Telecommunications and Control Systems*. IEEE, 227–233.
- [39] Nikolaos Mavridis, Marina-Selini Katsaiti, Silvia Naef, Abdullah Falasi, Abdulrahman Nuaimi, Hamdan Araifi, and Ahmed Kitbi. 2012. Opinions and Attitudes toward Humanoid Robots in the Middle East. *AI & society* 27, 4 (2012), 517–534.
- [40] Stefanos Nikolaidis, Minae Kwon, Jodi Forlizzi, and Siddhartha Srinivasa. 2018. Planning with Verbal Communication for Human-Robot Collaboration. *ACM Transactions on Human-Robot Interaction (THRI)* 7, 3 (2018), 22.
- [41] Stefanos Nikolaidis and Julie Shah. 2013. Human-Robot Cross-Training: Computational Formulation, Modeling and Evaluation of a Human Team Training Strategy. In *International Conference on Human-Robot Interaction (HRI)*. ACM, 33–40.
- [42] Frans A. Oliehoek, Christopher Amato, et al. 2016. *A Concise Introduction to Decentralized POMDPs*. Vol. 1. Springer.
- [43] Steven Pinker, Martin A. Nowak, and James J. Lee. 2008. The Logic of Indirect Speech. *Proceedings of the National Academy of Sciences* 105, 3 (2008), 833–838.
- [44] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y. Ng. 2009. ROS: An Open-source Robot Operating System. In *ICRA Workshop on Open Source Software*, Vol. 3, 5.
- [45] Aditi Ramachandran, Chien-Ming Huang, and Brian Scassellati. 2017. Give Me a Break!: Personalized Timing Strategies to Promote Learning in Robot-Child Tutoring. In *International Conference on Human-Robot Interaction (HRI)*. ACM, 146–155.
- [46] E. Ratner, D. Hadfield-Mennell, and A. Dragan. 2018. Simplifying Reward Design through Divide-and-Conquer. In *Robotics: Science and Systems (RSS)*.
- [47] Dorsa Sadigh, Anca D. Dragan, S. Shankar Sastry, and Sanjit A. Seshia. 2017. Active Preference-Based Learning of Reward Functions. In *Robotics: Science and Systems (RSS)*.
- [48] Kristin E. Schaefer, Edward R. Straub, Jessie Y.C. Chen, Joe Putney, and Arthur W. Evans III. 2017. Communicating Intent to Develop Shared Situation Awareness and Engender Trust in Human-Agent Teams. *Cognitive Systems Research* 46 (2017), 26–39.
- [49] Antonio Sgorbissa, Irena Papadopoulou, Barbara Bruno, Christina Koulouglioti, and Carmine Recchiuto. 2018. Encoding Guidelines for a Culturally Competent Robot for Elderly Care. In *International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 1988–1995.
- [50] Ankit Shah, Pritish Kamath, Shen Li, and Julie A. Shah. 2018. Bayesian Inference of Temporal Task Specifications from Demonstrations. In *Advances in Neural Information Processing Systems (NeurIPS)*. 3804–3813.

- [51] Julie Shah, James Wiken, Brian Williams, and Cynthia Breazeal. 2011. Improved Human-Robot Team Performance using Chaski, a Human-Inspired Plan Execution System. In *International Conference on Human-Robot Interaction (HRI)*. ACM, 29–36.
- [52] Aaron St. Clair and Maja Mataric. 2015. How Robot Verbal Feedback can Improve Team Performance in Human-Robot Task Collaborations. In *International Conference on Human-Robot Interaction (HRI)*. ACM, 213–220.
- [53] Daniel Szafrir, Bilge Mutlu, and Terrence Fong. 2014. Communication of Intent in Assistive Free Flyers. In *International Conference on Human-Robot Interaction (HRI)*. ACM, 358–365.
- [54] Stefanie Tellex, Thomas Kollar, Steven Dickerson, Matthew R. Walter, Ashis Gopal Banerjee, Seth Teller, and Nicholas Roy. 2011. Approaching the Symbol Grounding Problem with Probabilistic Graphical Models. *AI magazine* 32, 4 (2011), 64–76.
- [55] Stefanie Tellex, Pratiksha Thaker, Robin Deits, Dimitar Simeonov, Thomas Kollar, and Nicholas Roy. 2012. Toward Information Theoretic Human-Robot Dialog. In *Robotics: Science and Systems (RSS)*.
- [56] Andrea Thomaz, Guy Hoffman, Maya Cakmak, et al. 2016. Computational Human-Robot Interaction. *Foundations and Trends® in Robotics* 4, 2-3 (2016), 105–223.
- [57] Michael Tomasello. 2010. *Origins of Human Communication*. MIT Press.
- [58] Vaibhav V. Unhelkar\*, Shen Li\*, and Julie A. Shah. 2019. Semi-Supervised Learning of Decision-Making Models for Human-Robot Collaboration. In *Conference on Robot Learning (CoRL)*. Asterisks denote equal contribution.
- [59] Vaibhav V. Unhelkar and Julie A. Shah. 2019. Learning Models of Sequential Decision-Making with Partial Specification of Agent Behavior. In *AAAI Conference on Artificial Intelligence*.
- [60] Vaibhav V. Unhelkar, X. Jessie Yang, and Julie A. Shah. 2017. Challenges for Communication Decision-Making in Sequential Human-Robot Collaborative Tasks. In *Workshop on Mathematical Models, Algorithms, and Human-Robot Interaction at RSS*.
- [61] Ning Wang, David V. Pynadath, and Susan G. Hill. 2016. Trust Calibration within a Human-Robot Team: Comparing Automatically Generated Explanations. In *International Conference on Human-Robot Interaction (HRI)*. ACM, 109–116.
- [62] Changyun Wei, Koen V. Hindriks, and Catholijn M. Jonker. 2014. The Role of Communication in Coordination Protocols for Cooperative Robot Teams. In *International Conference on Agents and Artificial Intelligence*. 28–39.
- [63] David Whitney, Eric Rosen, James MacGlashan, Lawson L.S. Wong, and Stefanie Tellex. 2017. Reducing Errors in Object-Fetching Interactions through Social Feedback. In *International Conference on Robotics and Automation (ICRA)*. IEEE.
- [64] Tom Williams, Matthew Bussing, Sebastian Cabrol, Elizabeth Boyle, and Nhan Tran. 2019. Mixed Reality Deictic Gesture for Multi-Modal Robot Communication. In *International Conference on Human-Robot Interaction (HRI)*. ACM, 191–201.
- [65] Ping Xuan, Victor Lesser, and Shlomo Zilberstein. 2004. Modeling Cooperative Multiagent Problem Solving as Decentralized Decision Processes. *Submitted for Publication to Autonomous Agents and Multi-Agent Systems* (2004).
- [66] Nan Ye, Adhiraj Somani, David Hsu, and Wee Sun Lee. 2017. DESPOT: Online POMDP Planning with Regularization. *Journal of Artificial Intelligence Research (JAIR)* 58 (2017), 231–266.
- [67] Andrea Maria Zanchettin, Nicola Maria Ceriani, Paolo Rocco, Hao Ding, and Björn Matthias. 2015. Safety in Human-Robot Collaborative Manufacturing Environments: Metrics and Control. *IEEE Transactions on Automation Science and Engineering (T-ASE)* 13, 2 (2015), 882–893.
- [68] Brian D. Ziebart, Andrew L. Maas, J. Andrew Bagnell, and Anind K. Dey. 2008. Maximum Entropy Inverse Reinforcement Learning. In *AAAI Conference on Artificial Intelligence*, Vol. 8. 1433–1438.

## A LIST OF NOTATIONS

$H$	human, used as subscript	$s_E$	task/environment variables in the task state $s$
$R$	robot, used as subscript	$s_R$	robot-specific variables in the task state $s$
'	next (state), used as superscript	$s_{RC}$	robot's communication state, provides memory for $s_{RI}$ , $a_{RC}$ , and $t_{RC}$
AMM	Agent Markov Model	$s_{RI}$	robot's communicated intent
MDP	Markov Decision Process	$S$	state space of the task MMDP
MMDP	Multi-Agent Markov Decision Process	$S_H$	set of all $s_H$
POMDP	Partially Observable Markov Decision Process	$S_E$	set of all $s_E$
		$S_p$	set of all $s_p$
		$S_R$	set of all $s_R$
		$S_{RC}$	set of all $s_{RC}$
CommPlan	POMDP-based communication policy, our decision-making approach	$R$	reward function of the task MMDP, when not used as subscript
Silent	no-communication policy	$R_c$	cost (negative reward) of communication
Hand-crafted	developer-specified communication policy	$R_{pc}$	cost (negative reward) of communication that depends only on the POMDP action $a_p$
$a$	joint human-robot action, action of the task model (MMDP)	$R_{ps}$	task reward that depends only on the POMDP action $a_p$
$a_H$	human's action	$T$	transition function of task MMDP, which depends on the joint action $a$
$a_{HC}$	human's communication	$T_H$	transition function of $s_H$ , which depends on the joint action $a$
$a_p$	action of the robot's decision-making model (POMDP)	$T_E$	transition function of $s_E$ , which depends on the joint action $a$
$a_R$	robot's action	$T_F$	transition function of $s_F$ , which depends on the human action $a_H$
$a_{RC}$	robot's communication	$T_p$	transition function of $s_p$ that depends only on the POMDP action $a_p$
$A$	action space of the task MMDP	$T_{pc}$	transition function of $s_{RC}$ that depends only on the POMDP action $a_p$
$A_H$	set of all $a_H$	$T_{ps}$	transition function of $s$ that depends only on the POMDP action $a_p$
$A_{HC}$	set of all $a_{RC}$	$T_{px}$	transition function of $x_H$ that depends only on the POMDP action $a_p$
$A_p$	set of all $a_p$	$T_R$	transition function of $s_R$ , which depends on the joint action $a$
$A_R$	set of all $a_R$	$T_X$	transition function of $x_H$ , which depends on the human action $a_H$
$A_{RC}$	set of all $a_{RC}$	$x_H$	human's mental state, also human's latent decision factor in the AMM
$g$	an arbitrary nonlinear function	$X_H$	set of all $x_H$
$h$	time horizon of task MMDP	$\gamma$	discount factors of task MMDP
$h_{RC}$	duration of interruption interval	$\pi_H$	human's policy
$t_{RC}$	time since last communication	$\pi_R$	robot's policy
$p_a$	probability of the human teammate's anticipatory information sharing	$\rho_a$	minimum cost of communication
$p_c$	compliance, probability of the human complying to the robot's communication	$\rho_b$	maximum cost of communication
$p_r$	responsivity, probability of the human answering robot's query	$\emptyset$	silence or no communication
$p_t$	probability of human's utterance being true		
$q_c$	compliance probability distribution for a specific $(s_F, a_H, x_H)$ -tuple		
$s$	state of the task MMDP, $s \doteq (s_H, s_R, s_E)$		
$s_p$	state of the robot's decision-making model (POMDP), $s_p \doteq (s, s_{RC}, x_H)$		
$s_H$	human-specific variables in the task state $s$		
$s_F$	human's observable decision factors in the AMM, $s_F \doteq g(s, a_R, s_{RC}, a_{RC})$		

Section	Description
3	The developer specifies the collaborative task as an MMDP, namely, the parameters $(S, A, T, R, \gamma)$ .
4	The developer specifies robot’s communication capability and symbol grounding, namely, the parameters $(A_{RC}, A_{HC})$ .
6.1	The developer specifies parameters for the model of communication cost, namely, the parameters $(\rho_a, \rho_b, h_{RC})$ .
6.2	The developer specifies the features that impact human’s action-selection, i.e., the decision factors of the AMM $(X_H, S_F)$ .
6.2	The developer collects and annotates data of human’s behavior. An interim model of human’s action-selection behavior is learned using the data and algorithms for Bayesian AMM learning.
6.2	The developer specifies compliance distributions $q_c$ for robot communications $a_{RC}$ . The complete model of human’s action-selection behavior is analytically arrived using the interim model and compliance distributions.
6.3	The developer specifies parameters of human’s response model, namely, the parameters $p_a, p_r, p_t$ .
7	The model of robot’s decision-making (POMDP) is arrived at analytically using the previously specified models.

**Table 2: Steps involved in specifying robot’s decision-making model. Note that some steps are manual (requiring developer input) while others are automated.**

## B SUMMARY: SPECIFYING ROBOT’S DECISION-MAKING MODEL

The CommPlan framework is designed with the aim of facilitating robot developers (i.e., person/s tasked with specifying the robot’s policy) in achieving effective robot communication and, consequently, fluent human-robot collaboration. Effective robot decisions and fluent collaboration depend on a variety of factors, including the task objective, human behavior, and communication modality. Thus, in practice, specifying an effective robot policy is challenging and often time-intensive.

In order to address this challenge of specifying the robot’s policy, CommPlan adopts a model-based planning paradigm, where

- (Sections 3-6) the developer specifies component models for task, human behavior, and communication;
- (Section 7) using these component models, the robot’s decision-making model is arrived analytically; and
- (Section 7) using the decision-making model, the robot’s policy is generated algorithmically.

Further, the model specification process is modular and hybrid. As collecting communication-based interaction data is difficult in practice, the specification process enables merging modules learned from data with those obtained from the developer’s domain expertise. We summarize the steps involved in specification of the robot’s decision-making model in Table 2.

## C SUMMARY: DECISION-MAKING FOR ROBOT ACTIONS AND COMMUNICATIONS

As described in Sec. 7, given the decision-making model, CommPlan generates the robot’s policy using an online POMDP solver [58, 66]. Specifically, at each time step during the execution of the human-robot collaborative task, the robot

- (sense) receives observation;
- (belief update) uses the observation to update its belief regarding the human’s latent state;
- (plan) uses the POMDP model and updated belief to select its action and communication; and
- (act) executes its action and communication.

## D EXPERIMENTS: SETUP, MATERIALS, AND PROCEDURE

As detailed in Sec. 8, we utilize the meal preparation task for the experiments with human subjects. Figure 3 provides a still from the experiments, wherein the human-robot team is working in the shared workspace. A video of the team performing the example task is available as part of the supplementary material and also at the following link: <http://tiny.cc/CommPlan>. Here, we summarize the setup, materials, and procedure for the experiments.

**Setup and Materials.** Briefly, during the meal preparation task, the robot needs to fill the cups with juice, while the human prepares sandwiches. The human-robot team needs to coordinate their motion and activities to complete the task safely and efficiently in a shared workspace. In the experiments, juice was simulated with candy and the human used toy wooden ingredients.

The robot communicated using a speaker. It used off-the-shelf speech-to-text software for detecting human’s response [30]. The robot’s vocabulary was limited to its communication space, which is specified in Sec. 4. However, the robot could communicate any of the following messages at any point during the task,

- Please make the next sandwich at location 1.
- Please make the next sandwich at location 2.
- I am going to location 3.
- I am going to location 4.
- I am going back to refill juice (location 6).
- Where are you going?

In order to facilitate symbol grounding, numbered labels were placed next to the key landmarks (L1 to L5) in the shared workspace. The human and robot used the location numbers to refer to the landmarks while communicating.

A motion-capture system was used for sensing the position of human’s hand (i.e.,  $s_H$ ). In order to ensure safety and prevent collisions during task execution, a safety stop was implemented. Due to the safety stop, the robot is stopped if it is within a safety radius ( $= 0.1$  m) of the human or if the human is occluded. Further, once the stop is triggered, the robot remains static until the human leaves the safety radius. Finally, for all the baselines, the robot’s sensors and actuators were controlled using the Robot Operating System [44].



**Figure 3: A human-robot team performing a shared workspace task, namely, meal preparation in a kitchen.**

**Procedure.** Participants for the experiments were recruited on the MIT campus. The experiment protocol was approved by the institutional review board. After obtaining informed consent, participants were briefed about the collaborative robot, the safety systems in place, and the meal preparation task.

Each participant completed the task six times (two consecutive repetitions for each treatment: Silent, Hand-crafted, and CommPlan) with the robot. The order of treatments was randomized across

participants. In addition, to alleviate novelty effects, the participants performed a training trial at the start of the experiment session.

During each task trial, the robot indicated the start and end of the experiment by communicating a pre-scripted message. The participants were administered a pre-experiment demographic survey, three identical questionnaires during the experiment (one after each condition), and an open-ended survey at the end of the experiment. The dependent measures included objective measures of task performance and subjective measures of collaborative fluency [22].