2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Towards Proactive Safe Human-Robot Collaborations via Data-Efficient Conditional Behavior Prediction

Ravi Pandya*,[1], Zhuoyuan Wang*,[2], Yorie Nakahira[2], Changliu Liu[1]

Abstract— **We focus on the problem of how we can enable a robot to collaborate seamlessly with a human partner, specifically in scenarios where preexisting data is sparse. Much prior work in human-robot collaboration uses *observational* models of humans (i.e. models that treat the robot purely as an *observer*) to choose the robot's behavior, but such models do not account for the influence the robot has on the human's actions, which may lead to inefficient interactions. We instead formulate the problem of optimally choosing a collaborative robot's behavior based on a *conditional* model of the human that depends on the robot's future behavior. First, we propose a novel model-based formulation of conditional behavior prediction that allows the robot to infer the human's intentions based on its future plan in data-sparse environments. We then show how to utilize a conditional model for proactive goal selection and safe trajectory generation around human collaborators. Finally, we use our proposed proactive controller in a collaborative task with real users to show that it can improve users' interactions with a robot collaborator quantitatively and qualitatively.**

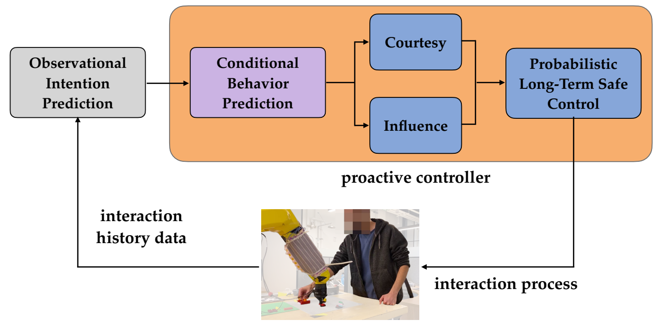

Fig. 1: Overall diagram of the proposed framework, including the CBP model (Sec. IV), courtesy and influence behavior (Sec. V) and long-term safe control (Sec. VI).

I. Introduction

As robots are becoming more common in industrial manufacturing, social, and home environments, it is imperative that they seamlessly collaborate with humans. This would allow us to take advantage of the speed and precision of robots as well as the flexibility of humans for completing tasks (such as in flexible industrial production lines [1], [2]). Instead of passively reacting to the human's behavior, we also want robots to *proactively* collaborate with humans by estimating and influencing their future behavior. One major challenge is the closed-loop nature of the interaction—the robot needs to understand what the human wants to do so it can assist them, but the robot's own actions will also influence the human's actions. Reasoning about this influence is an important factor in both efficient and safe human-robot interaction (HRI). Much existing work in the field of HRI has made use of observational models of humans [3], [4], but generally lacks the ability to predict how the human will respond to the robot's ultimate decision. Game-theoretic approaches do this reasoning, but often assume some hierarchical information structure [5], can be too restrictive to achieve optimal safety-performance tradeoffs in human-machine systems [6], and may be poor predictors of humans [7], [8], [9].

Autonomous driving researchers have taken note of the importance of influence and recently developed predictive models of agents that are conditioned on the ego agent's future plan [10], sometimes called *conditional behavior prediction* (CBP) models. CBP models rely on large datasets of human-human or human-robot interactions to reason about how a robot's actions would change the intentions of humans. Such datasets, however, are not available in many human-robot collaboration domains such as industrial manufacturing and home robotics. In particular, these environments have a variety different task specifications, so it may be infeasible to collect enough data for all tasks of interest. As a result, existing CBP techniques cannot be directly applied.

In this paper, we apply the CBP idea more broadly to HRI across data-sparse domains. We take a step towards enabling robots to similarly reason about the closed-loop collaboration with human partners in a data-efficient manner using a model-based CBP formulation. We then utilize this CBP model for downstream collaborative tasks by introducing an objective that proactively switches between courteous and influential behavior. We additionally show how to utilize a CBP model for long-term safe control in multi-goal human-robot collaborations. The system diagram of the proposed framework is shown in Fig. 1.

Our contributions in this work are the following:

1) Model-based conditional behavior prediction (CBP) in multi-goal human-robot collaborations.
2) An integrated framework for courtesy and influence behavior in robot goal selection utilizing a CBP model.

The merits of the proposed framework are validated using human experiments on multi-goal collaboration tasks.

II. Related Work

Human Goal Prediction: Much prior work has been done that treats humans as rational agents and tries to predict their intentions [3], [4], often using the framework of inverse reinforcement learning [11] and Bayesian inference [12],

*contributed equally to this work.
[1]Authors are with the Robotics Institute at Carnegie Mellon University, Pittsburgh, Pennsylvania, rapandya, cliu6@andrew.cmu.edu
[2]Authors are with the Department of Electrical and Computer Engineering at Carnegie Mellon University, Pittsburgh, Pennsylvania, zhuoyuaw, ynakahir@andrew.cmu.edu

[13]. Generally this line of work treats the robot as a *passive observer* of the human's behavior [14] and has the robot pick a best response to the prediction [15], [16]. Some work has considered the case where both agents have goals in the same space to complete a collaborative task [17], [18], [19], but still do not explicitly condition on the robot's future actions.

Conditional Behavior Prediction: Autonomous agents influencing humans has been considered recently in autonomous driving and pedestrian prediction settings [10], [20]. These CBP approaches have been applied in scenarios where there are existing large datasets that allow large models to be trained [21], [22], [23] (researchers in [10] note that their dataset consists of 18 years of continuous driving data). We are instead interested in general HRI where data may be limited, so we formulate a CBP approach that allows the robot to reason in other data-sparse environments.

Courtesy and Influence: Prior work on courtesy and influence behavior for robots often relies on human and robot collision avoidance [24], [25] or minimizes the expected cost incurred by a human [26], [27]. Other work on influencing humans usually rely on learning-based models, which are shown to be brittle with real humans [28], [29].

Safe Human-Robot Collaboration (HRC): Safety has long been studied for HRC with techniques like workplace separation [30], rule-based methods [31] and control barrier functions [32], [33]. These techniques ensure safety reactively, so are only safe in the short-term. We are interested in ensuring long-term safety for HRC, so we leverage a long-term safe control strategy for stochastic systems [34] and discuss how to incorporate our CBP model into the safe controller.

III. Problem Formulation

The HRI System. We consider the case of one robot collaborating with one human. The two agents have a shared set of goals Θ, known to both agents a priori. The human has noisy dynamics that depend on both agents' previous states and the human's true goal, denoted by θ_H^*:

$$x_H^{t+1} = m(x_H^t, x_R^t, \theta_H^*) + w^t, \tag{1}$$

where w^t is zero-mean Gaussian noise $w^t \sim \mathcal{N}(0, \Sigma_H)$ with covariance Σ_H. The robot has deterministic dynamics:

$$x_R^{t+1} = f_R(x_R^t, u_R^t), \tag{2}$$

where the robot's control policy u_R^t may arbitrarily depend on the two agents states x_H^t, x_R^t and the set of possible goals Θ. We assume that the robot has access to the form of the human's dynamics function $m(x_H^t, x_R^t, \theta)$, but does not know the actual goal of the human θ_H^* or how it might change over time. The human's dynamics may be estimated offline from data [35] or from online adaptation [36], but our focus is on dealing with uncertainty in the human's goal selection (particularly how it is affected by the robot).

The Robot's Objective. The robot's objective is to allow the team to work together fluently while keeping the human safe. The robot keeps a probabilisitic predictive model of the human $p(X_H \mid X_R = \mathbf{x_R}, \mathbf{o})$, where the future trajectories of the human and robot agents are random variables denoted

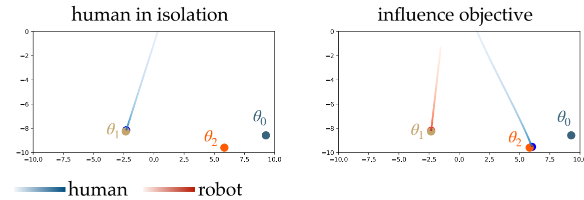

Fig. 2: **Left:** human chooses left goal in isolation. **Right:** robot chooses a goal to successfully influence the human's goal selection by modeling their goal change using model-based CBP.

by X_H and X_R respectively, and $\mathbf{o}$ contains observations of the environment and $\mathbf{x_R} = [x_R^t, ..., x_R^{t+T}]$ is a particular robot future trajectory (similarly $\mathbf{x_H}$ for human). Since the robot's prediction is probabilistic, it needs to stay safe probabilistically. We thus want the probability of collision with the human to be less than a small value ϵ, meaning the robot needs to stay safe with probability $1-\epsilon$. Based on a predefined safety specification, we define the set of safe states $\mathcal{Z}$. Let the joint state $x^t = [x_H^{t,\top}, x_R^{t,\top}]^\top$, so our probabilistic safety constraint is:

$$p(x^t \in \mathcal{Z},\ \forall t = 1, ..., H) \geq 1 - \epsilon, \tag{3}$$

where t is the timestep and H is the trajectory horizon.

Finally, we are given some objective J for the robot to minimize. This objective function could encompass both efficiency of completing the robot's own task and a penalty for affecting the human's intention. We investigate the design of J in Sec. V. Our robot's constrained objective is:

$$\min_{\mathbf{x_R}} \quad \mathbb{E}_{\mathbf{x_H} \sim p(X_H \mid X_R = \mathbf{x_R}, \mathbf{o})} \sum_{t=1}^{H} J(x_R^t, x_H^t) \tag{4a}$$

$$\text{s.t.} \quad p(x^t \in \mathcal{Z},\ \forall t = 1, ..., H) \geq 1 - \epsilon. \tag{4b}$$

Simulation Environment. In all simulations, the human and the robot are double-integrator LTI systems:

$$x_H^{t+1} = m(x_H^t, x_R^t, \theta_H^*) + w^t = A x_H^t + B u_H^t + w^t \tag{5}$$

$$x_R^{t+1} = f_R(x_R^t, u_R^t) = A x_R^t + B u_R^t \tag{6}$$

where $x_{\{H,R\}} = [p_x, v_x, p_y, v_y]^\top$ and $w^t \sim \mathcal{N}(0, \Sigma_H)$. The human's control is inspired by social force models of humans [37], where they are attracted to their goal θ_H^* and repelled from the robot:

$$u_H^t = K(\theta_H^* - x_H^t) + \frac{\gamma}{d^2}(C_H x_H^t - C_R x_R^t). \tag{7}$$

Here, K is the control gain computed from LQR, C_H and C_R are matrices that select the (x, y) positions of agents' states, $d = ||C_R x_R^t - C_H x_H^t||_2$ is the distance between the two agents, and γ controls how strongly the human is repelled from the robot. The two agents are trying to reach goal states from a set of three fixed goals without colliding with each other. The simulation environment can be seen in Fig. 2.

IV. Model-Based CBP

Key Insight. As previously noted, prior work using observational inference over a human's goal does not account for potential effects of a robot collaborator, since the robot is treated as an external observer. Our key insight is that we can *decouple* the human's *prior* goal θ_H^{prior} (from observational inference) from their *posterior* goal they will ultimately choose around the robot, θ_H^{post}. This allows us to formulate a Bayesian-inference-based CBP model that can be used for HRC without the need for large datasets.

Derivation. Observational Bayesian inference gives the robot access to

$$b_R^t(\theta_H^{prior}) := p(\theta_H^{prior} \mid x_H^{0:t}, x_R^{0:t}, u_H^{0:t}). \tag{8}$$

Given the simulated human's LQR controller (Sec. III), we follow [38] to compute an exact form of this belief.

We want the robot to estimate the distribution of θ_H^{post} conditioned on the robot's future plan: $b_R^t(\theta_H^{post} \mid x_R^{t+1:T}) = p(\theta_H^{post} \mid x_H^{0:t}, x_R^{0:t}, u_H^{0:t}, x_R^{t+1:T})$. However, we know that it may be computationally intractable to integrate over the space of future trajectories. We instead consider the robot's goal θ_R as a proxy for the robot's full plan:

$$b_R^t(\theta_H^{post} \mid \theta_R) := p(\theta_H^{post} \mid x_H^{0:t}, x_R^{0:t}, u_H^{0:t}, \theta_R). \tag{9}$$

We note that one could include θ_R in the human's reward function in observational Bayesian inference directly. However, decomposing the inference as we propose frees the robot to convert any nominal intention predictor (even learning-based) to a conditional one.

For ease of notation, we denote the robot's *observation* $(x_H^{0:t}, x_R^{0:t}, u_H^{0:t})$ as $\mathbf{o^{0:t}}$. To compute the conditional distribution, we can integrate out the variable θ_H^{prior} from the joint distribution $p(\theta_H^{post}, \theta_H^{prior} \mid \mathbf{o}^{0:t}, \theta_R)$ to get $p(\theta_H^{post} \mid \mathbf{o^{0:t}}, \theta_R) = \int_{\theta_H^{prior}} p(\theta_H^{post} \mid \theta_H^{prior}, \mathbf{o^{0:t}}, \theta_R) p(\theta_H^{prior} \mid \mathbf{o^{0:t}}, \theta_R)$. We know that the human's *prior* goal selection θ_H^{prior} will be independent of the robot's *future* goal θ_R, so we can simplify this equation as $\int_{\theta_H^{prior}} p(\theta_H^{post} \mid \theta_H^{prior}, \mathbf{o^{0:t}}, \theta_R) p(\theta_H^{prior} \mid \mathbf{o^{0:t}})$. Since we have a discrete set of goal locations, the final formula is

$$\begin{aligned} &p(\theta_H^{post} \mid \mathbf{o^{0:t}}, \theta_R) \\ &= \sum_{\theta_H^{prior}} p(\theta_H^{post} \mid \theta_H^{prior}, \mathbf{o^{0:t}}, \theta_R) p(\theta_H^{prior} \mid \mathbf{o^{0:t}}). \end{aligned} \tag{10}$$

Note that the second term is exactly what comes out of observational Bayesian goal inference. Now we need a way to compute $p(\theta_H^{post} \mid \theta_H^{prior}, \mathbf{o^{0:t}}, \theta_R)$. Inspired by prior work [39], [12], we use a softmax distribution with well-designed features as a strong prior so we can utilize this formulation in data-sparse environments:

$$p(\theta_H^{post} \mid \theta_H^{prior}, \mathbf{o^{0:t}}, \theta_R) = \frac{e^{-\beta_{cbp} s(\theta_H^{post}, \theta_R; \mathbf{o^{0:t}}, \theta_H^{prior})}}{\sum_{\theta_H} e^{-\beta_{cbp} s(\theta_H, \theta_R; \mathbf{o^{0:t}}, \theta_H^{prior})}}, \tag{11}$$

where β_{cbp} is an additional inverse temperature parameter for this distribution and $s(\cdot)$ is a score function that encodes how much robot's goal selection influences the human's goal.

V. The Robot's Multi-Stage Objective

Design Goal. Equipped with a CBP model, we wish to design an objective $J(\cdot)$ that will enable efficient collaborations with a human partner. We first note that many collaborative

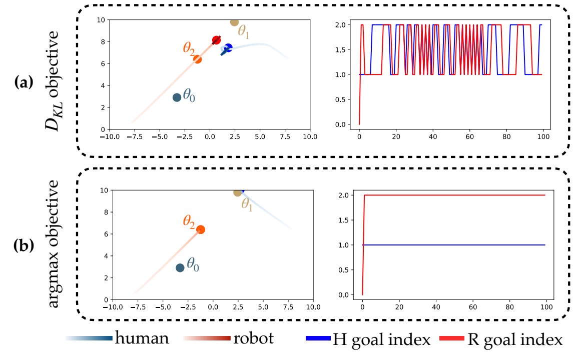

Fig. 3: Visualization of interaction and goal selection with **(a)** KL-divergence cost function **(b)** argmax belief cost function. The KL-divergence cost results in chattering of both agents' goals while the argmax belief results in a stable interaction.

tasks can be solved in a leader-follower manner [40], [41] where one agent decides on a general strategy and the other follows along. When interacting with humans, we want robots to be *courteous* so as to not interfere with the human's intended strategy. However, there may be times where the human is uncertain about what strategy to choose. At such times, we want the robot to *proactively influence* the human to choose an efficient strategy. We introduce a switching controller that estimates whether the human is hesitating then utilizes its CBP model to be either proactive or courteous.

Detecting Human Uncertainty. We aim to equip the robot with the ability to detect when the human is uncertain of their goal and subsequently influence them towards an efficient strategy. To do this, we turn to prior work on detecting model misspecificsation in Bayesian models [15], [42]. We jointly keep a belief over the human's prior intention θ_H^{prior} and the robot's model-confidence parameter β_R:

$$b_R^{t+1}(\theta_H^{prior}, \beta_R) = \frac{p(u_H^t \mid \mathbf{o^{0:t}}; \theta_H^{prior}, \beta_R) b_R^t(\theta_H^{prior}, \beta_R)}{\int_{\theta,\beta} p(u_H^t \mid \mathbf{o^{0:t}}; \theta, \beta) b_R^t(\theta, \beta) d\theta d\beta}. \tag{12}$$

We follow prior work [42] to detect when the model confidence is low: $\forall \theta \in \Theta, \text{argmax}_\beta b_R(\beta \mid \theta) < \delta$. If all hypotheses have the most probability mass on βs that are smaller than a threshhold δ, the robot will raise a flag saying that the human is uncertain about their intention. Since we assume both agents know the set of goals, the most likely cause of low model confidence is that the human has not decided on their goal.

Following Behavior: Courtesy Objective. Intuitively, avoiding changing the human's intention might be formulated as minimizing the KL-divergence between the robot's nominal belief of the human and the conditional belief of the human, which would mean that the robot's future actions have no effect on the human's intention:

$$J(\theta_R) = D_{\text{KL}}\left(b_R^{t+1}(\theta_H^{post} \mid \theta_R) \parallel b_R^t(\theta_H^{prior})\right). \tag{13}$$

We actually find that this approach results in chattering of the robot's goal (Fig. 3(a)). Since the robot is trying to keep the *belief distribution* close to the prior belief, it changes its goal to keep its belief from changing over time. Instead, we consider maximizing the probability that the human keeps the same intention before and after considering the robot's action. The objective function to *minimize* is:

$$\begin{aligned} J_c(\theta_R) &= -b_R^{t+1}(\theta_H^{post} = \hat{\theta}_H \mid \theta_R), \\ \text{where } \hat{\theta}_H &= \underset{\theta \in \Theta}{\text{argmax}}\ b_R^t(\theta_H^{prior} = \theta). \end{aligned} \tag{14}$$

Using this courtesy objective results in stable goal selection for the robot, seen in Fig. 3(b).

Leading Behavior: Influence Objective. When the human is uncertain of which goal to choose, the robot has a chance to influence them. We want the interaction to be efficient, so the robot can pick a goal for the human that it thinks is best for the team and try to influence the human to move towards it. The cost function is the total distance that the team would travel:

$$\begin{aligned} J_i(\theta_R) &= ||x_R^t - \theta_R|| + ||x_H^t - \hat{\theta}_H||, \\ \text{where } \hat{\theta}_H &= \underset{\theta \in \Theta}{\text{argmax}}\ b_R^{t+1}(\theta_H^{post} = \theta \mid \theta_R). \end{aligned} \tag{15}$$

Using this objective function and the CBP model, the robot influences the simulated human to select a different goal than they would have in isolation (Fig. 2).

Overall Objective. The robot's objective switches modes between *influencing* the human and being *courteous* to them based on whether the robot believes the human is unsure about what to do:

$$J(\cdot) = \begin{cases} J_i(\cdot), & \text{if } \forall \theta \in \Theta, \text{argmax}_\beta b_R(\beta \mid \theta) < \delta, \\ J_c(\cdot), & \text{otherwise.} \end{cases} \tag{16}$$

VI. Safe Trajectory Generation

We still need the robot to stay safe in the long term with respect to the actual *conditional* model of the human, since this will be the distribution that the human's goal selection ultimately takes. One major challenge is that the uncertainty in the human's intention hinders the direct application of short-term-based safe control methods to ensure long-term safety [43], [44]. We instead describe a method for allowing an existing *long-term* probabilistic safe controller [34] to sample directly from the CBP model to compute safe trajectories. We define safety as keeping a minimum distance d_{min} between human and robot. Mathematically, we can define a barrier function

$$\phi_{d_{\text{min}}}(x) = ||C_R x_R - C_H x_H||_2 - d_{\text{min}} \tag{17}$$

The safe set of states for the system is then $\mathcal{Z} = \{x : \phi_{d_{\text{min}}}(x) \geq 0\}$. To use the CBP model (Sec. IV) for safe control, the robot needs to know how its low-level action will affect the human's intention.

This can be done by letting the robot keep a *mental model* of the human's inference over the robot's goal $\hat{b}_H^t(\theta_R)$ (a common paradigm in the literature [39], [45]). The mental model is essentially an observational Bayesian inference model of the *robot's* goal. The mental model can be simulated forward and used to keep an overall belief over the human's posterior goal θ_H^{post}: $p(\theta_H^{post}) = \hat{b}_H^t(\theta_R) b_R^t(\theta_H^{post} \mid \theta_R)$. The robot can sample from $p(\theta_H^{post})$ to compute the probability of safety, which is a distribution that incorporates the proposed CBP model. We equip a probabilistic long-term safe controller from the literature [34] with this distribution $p(\theta_H^{post})$ to compute safe trajectories for the robot by sampling from the uncertain human intention instead of just from a stochastic dynamics model as described in the original paper.

VII. Simulation Environment

A. Score Function

For our two-agent goal-reaching tasks, we find the following score function to be expressive enough to enable proactive robot goal selection:

$$\begin{aligned} s(\theta_H, \theta_R; \mathbf{o^{0:t}}, \theta_H^{prior}) &= w_1||x_H^t - \theta_H|| \\ &- w_2||\theta_H - \theta_R|| + w_3||\theta_H - \theta_H^{prior}||, \end{aligned} \tag{18}$$

where $w_1, w_2, w_3 \in \mathbb{R}$ are hyperparameters[1]. This score function captures the idea that the human is more likely to choose a goal close to their current state (term 1), farther from the robot's goal (term 2), and prefer to not changing their current goal (term 3). A different score function would need to be chosen for different tasks, but in general should encode how likely the human is to change their goal from θ_H^{prior} to θ_H given that the robot chooses θ_R.

B. Simulated Human Models

Uncertain Human: The first simulated human infers the robot's goal online (a common paradigm from literature [39], [45], [46]) by keeping a belief $b_H^t(\theta)$ over the robot's goal. It chooses the closest goal that is not the same as the robot's most likely goal once its belief is sufficiently low-entropy ($\max_\theta b_H^t(\theta) \geq 0.4$):

$$\theta_H^* = \underset{\theta \in \Theta}{\text{argmin}}\ ||x_H^t - \theta||, \ \ \theta \neq \underset{\theta' \in \Theta}{\text{argmax}}\ b_H^t(\theta'). \tag{19}$$

Stubborn Human: The second simulated human is a stubborn human that will never change their goal (also a mode of human behavior studied in prior literature [47], [48]). This human chooses its goal to be the goal closest to itself

$$\theta_H^* = \underset{\theta \in \Theta}{\text{argmin}}\ ||C_H x_H - \theta||. \tag{20}$$

C. Baseline Robot Controllers

Naive Robot: This robot simply chooses the closest goal:

$$\theta_R^* = \underset{\theta \in \Theta}{\text{argmin}}\ ||C_R x_R - \theta|| \tag{21}$$

Reactive Robot: For a less naive baseline, we use a reactive goal selection method based on observational Bayesian inference, which picks a goal that is close to the robot but is not the inferred goal of the human:

$$\theta_R^* = \underset{\theta \in \Theta}{\text{argmin}}\ ||C_R x_R - \theta||, \ \ \theta \neq \underset{\theta' \in \Theta}{\text{argmax}}\ b_R^t(\theta') \tag{22}$$

Proactive-NN Robot: For a strong baseline, we implement a learning-based conditional behavior prediction model that has a similar structure to other learning-based CBP models

[1]The weights or the score function itself could be learned from data or adapted online, an exploration of which we leave to future work.

	% rollouts H changed θ	# times H changed θ
Naive	61%	0.8 ± 0.8
Reactive	42%	0.7 ± 1.1
Proactive-NN	**17%**	$\mathbf{0.27 \pm 0.7}$
Proactive-Model	*18%*	$\mathit{0.31 \pm 0.8}$

TABLE I: Average courtesy metrics in simulation (mean $\pm$ SD).

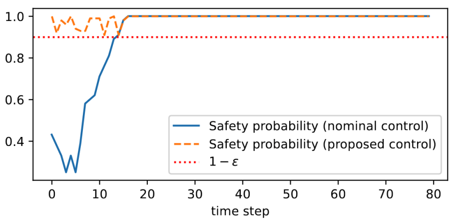

Fig. 4: Long-term safety probabilities in simulation.

from the literature [10], [20]. It takes in the agents' trajectory history and the robot's future plan and outputs the probability of the human reaching each goal. We construct a dataset of 3.5 million data points and use the same proactive goal-selection method (16) for this approach.

VIII. Simulation Results

We tested the proposed controller against a simulated human that is randomly selected to be either "uncertain" or "stubborn," (Sec. VII-B), unknown to the robot, for 100 simulated games. Each game lasts for 30 seconds. For the proactive-model robot, the weights in (18) are fixed at $[w_1, w_2, w_3] = [2, 0.9, 2]$ for both human models.

Proactive models are courteous. First, we test how courteous each robot agent is around the uncertain human (Table I). We measure the percentage of trajectories that the human changed their initial goal and the average number of times per trajectory that the human changed their goal. The proactive-NN controller performs the best[2] with our proposed proactive-model controller barely behind it. The proactive-model controller is able to perform just as well as the learning-based baseline in a data-sparse environment by just doing online inference.

Proactive models are flexible. We measure the efficiency of the interactions by counting the total number of goals reached by the team (Table II) over 100 initial conditions. The proactive-NN model performs the best on average with our proactive-model controller just behind it, showing us that the proactive models are flexible enough to efficiently interact with very different kinds of humans without prior knowledge. This also shows that the proposed controller is able to perform well in a data-sparse environment. We do,

[2]The NN-based controller is likely not near 0 because of distribution shift between training and testing.

	stubborn human	uncertain human	overall
Naive	3.2 ± 3.1	$\mathbf{7.68 \pm 4.0}$	5.44 ± 4.4
Reactive	6.4 ± 3.0	4.76 ± 2.1	5.6 ± 2.7
Proactive-NN	$\mathbf{9.1 \pm 3.9}$	6.28 ± 3.8	$\mathbf{7.7 \pm 4.1}$
Proactive-Model	$\mathit{8.6 \pm 4.9}$	6.4 ± 3.5	$\mathit{7.5 \pm 4.0}$

TABLE II: Average number of goals reached in simulation (mean $\pm$ SD).

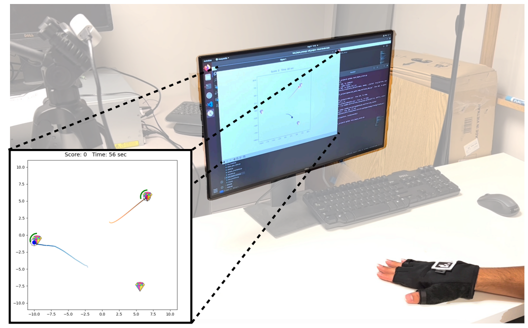

Fig. 5: Shows the setup for the user study where participants control an on-screen avatar directly with their hand to try and collect diamonds in collaboration with different robots.

however, see that the naive controller works best with the uncertain human. Because the naive controller will immediately commit to an action, it's very easy for the simulated human to respond immediately, whereas the other controllers take longer because they first estimate the human's intent.

Long-term safety can be assured. Fig. 4 shows the safety probability under time horizon $H = 20$ and risk tolerance $\epsilon = 0.1$ with and without the long-term safe controller while the robot selects goals with the proactive-model controller. We can see that with safety the modification, long-term safety probability is maintained over 90% as desired while the nominal goal-seeking controller fails to do so.

IX. User Study

We ran an IRB-approved study with real users interacting with our 2D environment. The study lasted 20 minutes and participants were paid \$5.

Experimental Setup: Users control an avatar on a computer screen with their hand. Their hand position is measured by a camera and mapped onto the 2D space (Fig. 5). The task is for the user to collect diamonds simultaneously with the (virtual) robot in the allotted time without colliding. Each participant interacted with all three robot types in a randomized order (four 45-second games per robot).

Independent Variables: We ran a within-subjects study and manipulated the *robot type* with three levels: *naive* (Sec. VII-C), *reactive* (Sec. VII-C) and *proactive* (Sec. V). We do not include the proactive-NN robot due to the lack of a large dataset of real people doing this task.

Objective Measures: We cannot directly measure if the human's intention changed over time, so we instead measure the *hesitation time*, or the amount of time the human waits before reaching for a goal, as a proxy for how much mental energy participants exert to choose their goal. We additionally measure the team's *efficiency* as the number of goals reached.

Subjective Measures: We ask participants a series of 5-point Lkert scale [49] questions about their experience with each robot and finally ask for a ranking of all robots. Here, we show the questions most relevant to the robot controller, but users were also asked general questions about their overall satisfaction and strategies on the task.

Hypothesis H1: *Participants interacting with the proactive robot will hesitate less and score more points than when*

	# goals reached	avg. hesitation time (*sec*)
Naive	$\mathbf{18.9 \pm 2.1}$	0.95 ± 0.34
Reactive	15.2 ± 1.6	1.2 ± 0.55
Proactive-Model	17.1 ± 2.3	$\mathbf{0.87 \pm 0.37}$

TABLE III: Objective results for user study (mean ± SD).

	Naive	Reactive	Proactive
Accounted: "[Robot] accounted for the [goal] I wanted to pick when it was picking a [goal]."	2.1	**4.0**	3.7
Changed: "I often changed which [goal] I picked initially because of [Robot]."	3.7	**1.7**	2.7
% ranked 1st (forced choice)	19%	33%	**47%**

TABLE IV: Responses to subjective survey questions (5: Strongly Agree, 1: Strongly Disagree).

interacting with the reactive robot.

Hypothesis H2: *Participants will feel that the proactive robot was easier to work with because it better accounted for their intentions.*

Participants: We recruited 22 participants (largely with technical backgrounds) from the campus community. One user's data was not included due to technical difficulties.

Quantitative Results: Running one-way repeated measures ANOVAs tell us the robot type had a statistically significant effect on the number of goals reached by the team (Table III, $F(2, 40) = 67.1, p < 0.0001$) and on users' hesitation time ($F(2, 40) = 14.7, p < 0.0001$). A post-hoc Bonferroni test on the score tells us that the differences in goals reached between all three pairs are significant ($p < 0.001$ for all). Users performed the best with the Naive robot, second best with the Proactive robot and worst with the Reactive robot. This partially supports **H1**, though we did not expect users to perform so well with the naive robot—users were ultimately capable of being followers, even if they did not necessarily enjoy it. A post-hoc Bonferroni test on the hesitation time tells us that users hesitated significantly less with the proactive robot than the reactive robot ($p = 0.0004$) and less with the naive robot than the reactive robot ($p = 0.004$). This result also supports **H1**, and we see a surprisingly positive result with the naive robot, likely meaning users were able to quickly react to the naive robot's goal selection. Throughout the interaction, users could likely easily predict the naive robot's actions, telling us that predictability is also important in designing efficient collaborative robots.

We also empirically measured the safety of the interactions to check that they line up with the theoretical guarantees. Across interactions with all 21 users, there were 13 collisions out of $45\text{s} \times 10\text{Hz} \times 4$ trials $\times$ 21 users $= 37800$ timesteps, for an empirical safety probability of 99.96%.

Qualitative Results: To test **H2**, we ran a repeated measures ANOVA on the effect of the robot type on the survey question "The robot was easy to collaborate with," but found no significant differences ($F(2, 40) = 2.9, p = 0.06$), since users rated all three as easy to collaborate with. We did, however, find significant differences between the robots on the "Accounted" (Table IV, $F(2, 40) = 12.4, p < 0.0001$) and "Changed" questions ($F(2, 40) = 18.5, p < 0.0001$). A post-hoc Bonferroni test tells us that, relative to the naive robot, participants thought the proactive ($p = 0.0006$) and reactive ($p = 0.0005$) robots accounted more for their goal selection. This seems to support **H2**, although participants tended to think the proactive robot was about as responsive to their goal selection as the reactive robot. A post-hoc Bonferroni test tells us that participants felt that they changed their goals around the naive robot, while they did not change their goals around the proactive robot ($p = 0.01$) or around the reactive robot ($p < 0.0001$), also supporting **H2**. They also felt that they changed their goal less with the reactive robot than the proactive robot ($p = 0.002$).

We additionally look at users' rankings of the three robots (Table IV) to test **H2**. While not a majority, a plurality of participants (47%) ranked the proactive robot as the most preferred interaction partner when asked to rank all three robots, which supports **H2**. Treating the participants' rankings as votes for the robots and running an instant-runoff-voting election shows that the proactive robot is the winner of the "election."

X. Conclusion and Future Work

We proposed a Bayesian-inference-based formulation of conditional behavior prediction that can allow a robot to reason about the affect its future actions will have on a human in collaborative multi-goal settings. The CBP formulation was based on the insight that we can decouple the human's prior goal selection θ_H^{prior} from their posterior goal selection θ_H^{post} that accounts for the robot's actions. We additionally discussed how to utilize a CBP model for long-term safety as well as for switching between courteous and influential behavior. In simulations, we found that our proposed controller allows the robot to be efficient at interacting with different kinds of humans (even without tuning the score function weights) and that it performs very similarly to a learning-based CBP model without the need for a large dataset. We tested our proposed controller in a user study and found that although users tended to perform the best with the naive baseline robot, users tended to *prefer* interacting with our proactive controller. Future work may try to improve the model-based CBP approach by adapting the score function weights online, which would keep the data-efficiency benefits while potentially increasing the robot's efficiency at acting with different kinds of real human partners.

Acknowledgments

This material is based upon work supported by the National Science Foundation (NSF) Graduate Research Fellowship under Grant No. DGE1745016 and DGE2140739 and additionally under Grant No. 2144489 as well as the Manufacturing Futures Institute, Carnegie Mellon University, through a grant from the Richard King Mellon Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

[1] J. Krüger, T. K. Lien, and A. Verl, "Cooperation of human and machines in assembly lines," *CIRP annals*, vol. 58, no. 2, pp. 628–646, 2009.

[2] V. Villani, F. Pini, F. Leali, and C. Secchi, "Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications," *Mechatronics*, vol. 55, pp. 248–266, 2018.

[3] C. L. Baker, J. B. Tenenbaum, and R. R. Saxe, "Goal inference as inverse planning," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 29, no. 29, 2007.

[4] C. L. Baker, R. Saxe, and J. B. Tenenbaum, "Action understanding as inverse planning," *Cognition*, vol. 113, no. 3, pp. 329–349, 2009.

[5] J. F. Fisac, E. Bronstein, E. Stefansson, D. Sadigh, S. S. Sastry, and A. D. Dragan, "Hierarchical game-theoretic planning for autonomous vehicles," in *2019 International conference on robotics and automation (ICRA)*. IEEE, 2019, pp. 9590–9596.

[6] Z. Zhang, A.-S. Maitham, H. Jing, H. Shirado, and Y. Nakahira, "Rethinking safe control in the presence of self-seeking humans," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 12, 2023, pp. 15 331–15 339.

[7] F. Bornhorst, A. Ichino, O. Kirchkamp, K. H. Schlag, and E. Winter, "How do people play a repeated trust game?: Experimental evidence," *None*, 2004.

[8] L. Bruttel and U. Kamecke, "Infinity in the lab. how do people play repeated games?" *Theory and Decision*, vol. 72, pp. 205–219, 2012.

[9] S. Gächter, "Behavioral game theory," *Blackwell handbook of judgment and decision making*, pp. 485–503, 2004.

[10] E. Tolstaya, R. Mahjourian, C. Downey, B. Vadarajan, B. Sapp, and D. Anguelov, "Identifying driver interactions via conditional behavior prediction," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 3473–3479.

[11] A. Y. Ng, S. Russell *et al.*, "Algorithms for inverse reinforcement learning." in *Icml*, vol. 1, 2000, p. 2.

[12] A. Bestick, R. Pandya, R. Bajcsy, and A. D. Dragan, "Learning human ergonomic preferences for handovers," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 3257–3264.

[13] D. P. Losey and M. K. O'Malley, "Enabling robots to infer how end-users teach and learn through human-robot interaction," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1956–1963, 2019.

[14] D. Vasquez, T. Fraichard, O. Aycard, and C. Laugier, "Intentional motion on-line learning and prediction," *Machine Vision and Applications*, vol. 19, no. 5-6, pp. 411–425, 2008.

[15] J. F. Fisac, A. Bajcsy, S. L. Herbert, D. Fridovich-Keil, S. Wang, C. J. Tomlin, and A. D. Dragan, "Probabilistically safe robot planning with confidence-based human predictions," *arXiv preprint arXiv:1806.00109*, 2018.

[16] J. Mainprice and D. Berenson, "Human-robot collaborative manipulation planning using early prediction of human motion," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 299–306.

[17] S. Pellegrinelli, H. Admoni, S. Javdani, and S. Srinivasa, "Human-robot shared workspace collaboration via hindsight optimization," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 831–838.

[18] S. J. Levine and B. C. Williams, "Watching and acting together: Concurrent plan recognition and adaptation for human-robot teams," *Journal of Artificial Intelligence Research*, vol. 63, pp. 281–359, 2018.

[19] M. C. Gombolay, R. A. Gutierrez, S. G. Clarke, G. F. Sturla, and J. A. Shah, "Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams," *Autonomous Robots*, vol. 39, pp. 293–312, 2015.

[20] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*. Springer, 2020, pp. 683–700.

[21] N. Rhinehart, R. McAllister, K. Kitani, and S. Levine, "Precog: Prediction conditioned on goals in visual multi-agent settings," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 2821–2830.

[22] E. Schmerling, K. Leung, W. Vollprecht, and M. Pavone, "Multimodal probabilistic model-based planning for human-robot interaction," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 3399–3406.

[23] C. Tang, W. Zhan, and M. Tomizuka, "Interventional behavior prediction: Avoiding overly confident anticipation in interactive prediction," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 11 409–11 415.

[24] H. Yu, Z. Kuang, F. Sun, J. Yuan, M. Liu, and M. Xie, "Path planning for mobile robots with courteous behaviors in domestic environments," in *2021 40th Chinese Control Conference (CCC)*. IEEE, 2021, pp. 3894–3899.

[25] D. Sadigh, S. Sastry, S. A. Seshia, and A. D. Dragan, "Planning for autonomous cars that leverage effects on human actions." in *Robotics: Science and systems*, vol. 2. Ann Arbor, MI, USA, 2016, pp. 1–9.

[26] L. Sun, W. Zhan, M. Tomizuka, and A. D. Dragan, "Courteous autonomous cars," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 663–670.

[27] A. Bestick, R. Bajcsy, and A. D. Dragan, "Implicitly assisting humans to choose good grasps in robot to human handovers," in *2016 International Symposium on Experimental Robotics*. Springer, 2017, pp. 341–354.

[28] A. Xie, D. Losey, R. Tolsma, C. Finn, and D. Sadigh, "Learning latent representations to influence multi-agent interaction," in *Conference on robot learning*. PMLR, 2021, pp. 575–588.

[29] W. Z. Wang, A. Shih, A. Xie, and D. Sadigh, "Influencing towards stable multi-agent interactions," in *Conference on robot learning*. PMLR, 2022, pp. 1132–1143.

[30] S. Robla-Gómez, V. M. Becerra, J. R. Llata, E. Gonzalez-Sarabia, C. Torre-Ferrero, and J. Perez-Oria, "Working together: A review on safe human-robot collaboration in industrial environments," *Ieee Access*, vol. 5, pp. 26 754–26 773, 2017.

[31] A. M. Zanchettin, N. M. Ceriani, P. Rocco, H. Ding, and B. Matthias, "Safety in human-robot collaborative manufacturing environments: Metrics and control," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 882–893, 2015.

[32] F. Ferraguti, C. T. Landi, S. Costi, M. Bonfè, S. Farsoni, C. Secchi, and C. Fantuzzi, "Safety barrier functions and multi-camera tracking for human–robot shared environment," *Robotics and Autonomous Systems*, vol. 124, p. 103388, 2020.

[33] C. Liu and M. Tomizuka, "Safe exploration: Addressing various uncertainty levels in human robot interactions," in *2015 American Control Conference (ACC)*. IEEE, 2015, pp. 465–470.

[34] Z. Wang, H. Jing, C. Kurniawan, A. Chern, and Y. Nakahira, "Myopically verifiable probabilistic certificate for long-term safety," in *2022 American Control Conference (ACC)*. IEEE, 2022, pp. 4894–4900.

[35] R. Liu and C. Liu, "Human motion prediction using adaptable recurrent neural networks and inverse kinematics," *IEEE Control Systems Letters*, vol. 5, no. 5, pp. 1651–1656, 2020.

[36] R. Pandya and C. Liu, "Safe and efficient exploration of human models during human-robot interaction," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 6708–6715.

[37] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," *Physical review E*, vol. 51, no. 5, p. 4282, 1995.

[38] R. Tian, M. Tomizuka, A. D. Dragan, and A. Bajcsy, "Towards modeling and influencing the dynamics of human learning," in *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 2023, pp. 350–358.

[39] A. D. Dragan, K. C. Lee, and S. S. Srinivasa, "Legibility and predictability of robot motion," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2013, pp. 301–308.

[40] E. M. Van Zoelen, E. I. Barakova, and M. Rauterberg, "Adaptive leader-follower behavior in human-robot collaboration," in *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2020, pp. 1259–1265.

[41] C. Messeri, A. M. Zanchettin, P. Rocco, E. Gianotti, A. Chirico, S. Magoni, and A. Gaggioli, "On the effects of leader-follower roles in dyadic human-robot synchronisation," *IEEE Transactions on Cognitive and Developmental Systems*, 2020.

[42] A. Bobu, A. Bajcsy, J. F. Fisac, S. Deglurkar, and A. D. Dragan, "Quantifying hypothesis space misspecification in learning from human–robot demonstrations and physical corrections," *IEEE Transactions on Robotics*, vol. 36, no. 3, pp. 835–854, 2020.

[43] S. He, W. Zhao, C. Hu, Y. Zhu, and C. Liu, "A hierarchical long short term safety framework for efficient robot manipulation under uncertainty," *Robotics and Computer-Integrated Manufacturing*, vol. 82, p. 102522, 2023.

[44] T. Wei and C. Liu, "Safe control algorithms using energy functions: A unified framework, benchmark, and new directions," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 238–243.
[45] S. H. Huang, I. Huang, R. Pandya, and A. D. Dragan, "Nonverbal robot feedback for human teachers," *arXiv preprint arXiv:1911.02320*, 2019.
[46] S. H. Huang, D. Held, P. Abbeel, and A. D. Dragan, "Enabling robots to communicate their objectives," *Autonomous Robots*, vol. 43, pp. 309–326, 2019.
[47] S. Nikolaidis, D. Hsu, and S. Srinivasa, "Human-robot mutual adaptation in collaborative tasks: Models and experiments," *The International Journal of Robotics Research*, vol. 36, no. 5-7, pp. 618–634, 2017.
[48] S. Devin, A. Clodic, and R. Alami, "About decisions during human-robot shared plan achievement: Who should act and how?" in *Social Robotics: 9th International Conference, ICSR 2017, Tsukuba, Japan, November 22-24, 2017, Proceedings 9*. Springer, 2017, pp. 453–463.
[49] R. Likert, "A technique for the measurement of attitudes." *Archives of psychology*, 1932.