Should Collaborative Robots be Transparent?

Shahabedin Sagheb, Soham Gandhi, and Dylan P. Losey

***Abstract*— Today's robots often assume that their behavior should be transparent. These transparent (e.g., legible, explainable) robots intentionally choose actions that convey their internal state to nearby humans. But while transparent behavior seems beneficial, is it actually *optimal*? In this paper we consider collaborative settings where the human and robot have the same objective, and the human is uncertain about the robot's type (i.e., the robot's internal state). We extend a recursive combination of Bayesian Nash equilibrium and the Bellman equation to solve for optimal robot policies. Interestingly, we discover that it is *not always* optimal for collaborative robots to be transparent; instead, human and robot teams can sometimes achieve higher rewards when the robot is *opaque*. Opaque robots select the same actions regardless of their internal state: because each type of opaque robot behaves in the same way, the human cannot infer the robot's type. Our analysis suggests that opaque behavior becomes optimal when either (a) human-robot interactions have a short time horizon or (b) users are slow to learn from the robot's actions. Across online and in-person user studies with 43 total participants, we find that users reach higher rewards when working with opaque partners, and subjectively rate opaque robots as about equal to transparent robots. See videos of our experiments here: https://youtu.be/u8q1Z7WHUuI**

***Index Terms*— Human-Robot Collaboration, Optimization and Optimal Control, Human-Aware Motion Planning**

I. Introduction

Imagine working with a collaborative robot arm to build a block tower (see Figure 1). You previously taught the robot how to stack different blocks, but now you are not sure what the robot has learned. If the robot learned correctly it will be *capable* of reaching for distant blocks and building a larger tower. But if the robot is still *confused* it will only be able to add the closer, smaller blocks. Whether the robot is capable or confused affects your decisions. If the robot has learned to stack larger blocks you can also add large blocks and build a taller tower; but if the robot only knows how to stack small blocks, you will need to add similarly sized blocks to keep the tower from becoming unstable. Given this uncertainty, it seems intuitive that the collaborative robot should pick up blocks that reveal whether it is capable or confused.

Today's approaches to human-robot interaction often assume that robot behavior should be *transparent* (e.g., legible, explainable, understandable) [1]–[3]. Transparent robots take actions that purposely reveal their internal state. For instance, when reaching for a block on a cluttered table, a transparent robot will exaggerate its trajectory so that nearby humans can predict which block the robot is going to grab [4]–[6]. Transparent motions are beneficial because they convey information to the human, and the human can then leverage this information to better coordinate with the robot [7]–[9].

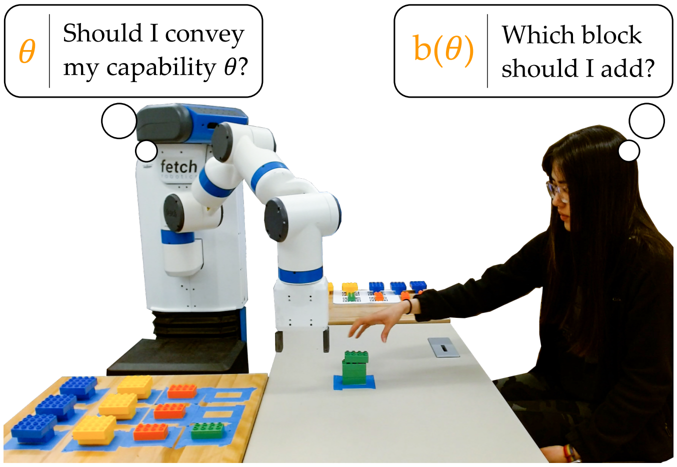

Fig. 1. Collaborative block-stacking task where the human is uncertain about the robot's internal state θ. Transparent robot actions help the human learn θ and decide what blocks to add to the tower. However, we find that the costs of this transparent behavior may outweigh its benefits.

But transparent behavior also comes at a cost. Consider our example of reaching for a block: by exaggerating its trajectory the robot takes longer to reach the block and complete the task. Going one step further, different humans may require different amounts of time to recognize what the robot is trying to convey and update their own behavior.

In this paper we explore whether transparent behavior is always beneficial for human-robot teams. We introduce *opacity* as the opposite of transparency: opaque robots select actions that withhold information from the human. To determine when it is optimal for robots to withhold information and select opaque behaviors, our insight is that:

We can formulate collaborative interactions where the human is uncertain about the robot internal state as a two-player stochastic Bayesian game.

We develop an algorithmic framework to solve these games and obtain optimal robot policies for each internal state (i.e., for each type of robot). Interestingly, we find that — under some conditions — the optimal policy is the *same* for every robot type and the robot's resulting behavior is opaque to the human. Return to our motivating example in Figure 1. Although we might have expected the capable robot to stack large blocks and the confused robot to stack small blocks, the human-robot team actually has a higher expected reward if both robots always build the smaller tower. Put another way, when the human and robot act optimally the human never learns whether the robot is capable or confused!

Overall, we make the following contributions:

Formalizing Opacity. We capture settings where the human and robot have the same payoff and the human is uncertain about the robot's type as stochastic Bayesian games. Within

The authors are with the Collaborative Robotics Lab (Collab), Dept. of Mechanical Engineering, Virginia Tech, Blacksburg, VA 24061.
Corresponding author's email: shahab@vt.edu

this context we define fully and rationally opaque behavior.

Showing Opacity is Optimal. We extend a recursive combination of Bayesian Nash Equilibrium and the Bellman equation to find optimal robot policies. Our analysis suggests that it is more likely for opaque behavior to be optimal when (a) interactions have a short time horizon and (b) humans are slow to learn from robot actions.

Measuring User Responses to Opaque Robots. We conduct online and in-person user studies where we compare opaque and transparent robots. Across 43 total participants we support our theoretical analysis and show that opaque behavior does lead to higher human rewards. Users prefer opaque robots about equally to transparent partners.

II. Related Work

Transparent Robots. Prior work on human-robot interaction often assumes that robots should be *transparent* (e.g., legible or explainable) [1]–[4]. Transparent robots actively and intentionally reveal their internal state to nearby humans so that these humans understand the robot's intent. Here we specifically focus on transparent robot actions (e.g., motions). Recent research demonstrates that robots can exaggerate their actions to communicate their goal [5], indicate their intended trajectory [6], and express whether or not they are capable of performing a task [10]. When applied to collaborative human-robot teams where both agents are working together to complete a shared task, experiments suggest that transparent robot motion improves overall team performance [7]–[9].

Although transparency is often perceived as a benefit to human-robot interaction, we explore the opposite perspective: is it ever optimal for robots to *hide* their internal state and intentionally *withhold* information from the human?

Deceptive Robots. Robot actions can be deceptive or misleading. For instance, by initially moving towards the wrong goal a robot can convince the human that this goal is what the robot actually wants — even if the robot has another target in mind [11], [12]. We are not interested in explicitly encouraging deceptive actions; instead, we study situations where this behavior emerges *naturally* as part of the robot's optimal policy. Towards this end we will formulate human-robot interaction as a two-player collaborative game [13]–[16]. When solving similar games, recent research has found that it can be optimal for robots to take actions that influence humans [17]–[19] or mislead users [20].

In these prior works the human and robot are *competitors*: the robot has a different objective than the human. For instance, in [17] an autonomous car and human driver are competing to cross the intersection first, and the autonomous car takes misleading actions to influence the human driver to yield. By contrast, in our paper the human and robot are *collaborative*: both agents share the exact same reward function and have no incentive to mislead one another.

III. Problem Formulation

We ask the question: when is it optimal for robots to hide their capabilities from human partners? More specifically, we focus on *collaborative* interactions where both the human and robot share the same reward function (i.e., both agents get the exact same payoff at every timestep). While the human and robot know this reward function, the human is *uncertain* about the robot's *type*. Here type θ captures latent information observed only by the robot. For instance, consider our running example where a human is teaching a robot arm: the robot has either learned the task (type capable) or it has not (type confused). The human does not initially know whether robot is capable or confused, and the robot must decide whether to take actions that reveal its type.

Interaction. Let $s \in \mathcal{S}$ be the system state, let $a_{\mathcal{R}} \in \mathcal{A}_{\mathcal{R}}$ be the robot action, and let $a_{\mathcal{H}} \in \mathcal{A}_{\mathcal{H}}$ be the human action. The system state transitions using deterministic dynamics:

$$s^{t+1} = f(s^t, a^t_{\mathcal{H}}, a^t_{\mathcal{R}}) \tag{1}$$

where both human and robot actions affect the system state. At each timestep the human and robot act simultaneously. Neither agent goes first: both select their actions $a^t_{\mathcal{R}}$ and $a^t_{\mathcal{H}}$ without knowing what action the other agent is taking. The interaction ends after a total of T timesteps.

Robot Type. Throughout the interaction the robot has a fixed type θ. Within our experiments *type* refers to the robot's level of capability (e.g., how quickly the robot can move or how effectively the robot can grasp objects). More generally, θ is latent information known only by the robot. Let there be N possible types of robots. At the start of each interaction θ is sampled from prior $P(\theta)$. The human knows the prior distribution $P(\theta)$ but not the robot's current type θ.

Belief. The human updates their estimate of the robot's type based on the robot's behavior. Let $b^{t+1}(\theta) = P(\theta \mid s^{0:t}, a^{0:t}_{\mathcal{R}})$ be the probability that the robot is type θ given that we have visited states $s^{0:t}$ and the robot has taken actions $a^{0:t}_{\mathcal{R}}$. Similar to prior work [21], we assume that the human updates this belief using Bayesian inference:

$$b^{t+1}(\theta) \propto b^t(\theta) \cdot P(a^t_{\mathcal{R}} \mid s^t, \theta) \tag{2}$$

where $b^0(\theta) = P(\theta)$ is the known prior over the robot's type. The likelihood function $P(a_{\mathcal{R}} \mid s, \theta)$ expresses — from the human's perspective — how likely the robot is to take action $a_{\mathcal{R}}$ at state s given that the robot is type θ. In the following sections we will explore different models for $P(a_{\mathcal{R}} \mid s, \theta)$.

Reward. At each timestep the collaborative human and robot receive the same reward $r(s)$. Both agents know this reward function and know that the other agent shares the same reward. Summing reward at each timestep gives the total reward across the entire interaction: $\sum_{t=0}^{T} r(s^t)$.

Stochastic Bayesian Game. The human and robot want to maximize their reward across the interaction. If we assume that the human is a rational agent, this problem is an instance of a two-player stochastic Bayesian game where the human is uncertain about the robot's type [16]. We highlight two non-standard aspects of our formulation. First, our game is entirely collaborative: the agents always receive the same payoffs. Second, the human updates their belief according to Equation (2), and different humans may leverage different likelihood functions when updating this belief.

Policies. Within a stochastic Bayesian game the human and robot each have policies. The human's policy $a_{\mathcal{H}} = \pi_{\mathcal{H}}(s, b)$ maps the current state and belief to actions, and the robot's policy $a_{\mathcal{R}} = \pi_{\mathcal{R}}(s, b, \theta)$ maps the state, belief, and type θ to robot actions. Remember that the robot knows its type θ. In practice, robots of different types may take different actions given the same state and belief: e.g., at a given (s, b) a capable robot could reach for the farther block while the confused robot might move for the closer block.

IV. Should Collaborative Robots be Opaque?

It seems intuitive to think collaborative robots (i.e., robots that share our reward function) will be transparent. Returning to our example, we might expect the capable robot to take actions that reveal it can reach the far block, and the confused robot to take actions that demonstrate it is limited to the close block. However, in this section we theoretically prove that transparency is not always optimal. We start by introducing formal definitions for *opacity* when the robot is interacting with rational and irrational humans (Section IV-A). Next, we derive a game-theoretic approach for finding *optimal* robot behavior in collaborative games (Section IV-B). Finally, we provide an example to prove that — within our problem setting — there exist cases where it is optimal for robots to be opaque and hide their capabilities (Section IV-C).

A. Formalizing Robot Opacity

What do we mean when we say a robot is opaque? Here we will introduce two definitions for opacity within stochastic Bayesian games where the human is uncertain about the robot's type. Our first definition makes no assumptions about the human's policy, and applies when the robot is interacting with *any* human partner. Our second definition assumes that the human acts *rationally*: i.e., the human chooses actions to maximize the expected cumulative reward. For both definitions we reason about opacity in terms of the human's belief at the end of the interaction, b^T. Intuitively, a transparent system should cause the human to reach different beliefs when interacting with different robot types, while an opaque system should cause humans to converge to the same final belief regardless of the robot's actual type θ.

Fully Opaque. Let (s^0, b^0) be the initial system state and prior over the robot type. We say (s^0, b^0) is fully opaque if — no matter which type $\theta \in \Theta$ the robot actually is — the human's final belief b^T is identical.

As a simple example, consider a robot where $\pi_{\mathcal{R}}(s, b, \theta)$ is always zero; i.e., a robot that never takes any action. The human cannot distinguish what type of robot they are working with regardless of what policy $\pi_{\mathcal{H}}$ the human chooses. Accordingly, the human's final belief b^T after interacting with the robot for T timesteps is the same when $\theta =$ capable and when $\theta =$ confused. We claim that this robot is *fully opaque* because, no matter what the human does, they always reach the same understanding of the robot's type.

Rationally Opaque. Let (s^0, b^0) be the initial system state and prior, and assume the human takes actions to maximize their expected total reward in the stochastic Bayesian game (i.e., the human acts optimally). We say (s^0, b^0) is rationally opaque if — no matter which type $\theta \in \Theta$ the robot actually is — the rational human's final belief b^T is identical.

The difference between *fully opaque* and *rationally opaque* is our assumption about the human's policy. In both cases the robot does not reveal any information about its type over the course of interaction. But for rationally opaque the human is *constrained* to always take optimal actions, while for fully opaque the human is free to choose any policy. As we will show, a robot starting at (s^0, b^0) may be rationally opaque *but not* fully opaque (i.e., an irrational human could take random actions that cause the robot to reveal its type).

B. Identifying Optimal Behavior

Now that we have defined opacity, we want to determine if it is ever *optimal* for robots to be opaque (i.e., to withhold their type). Remember that the human and robot are agents in a two-player stochastic Bayesian game. In this subsection we will present an algorithm for finding optimal human and robot policies $\pi_{\mathcal{H}}$ and $\pi_{\mathcal{R}}$ within our game-theoretic setting. Next, in Section IV-C we will apply this algorithm to specific scenarios and demonstrate that the robot's optimal behavior in these settings is fully opaque or rationally opaque.

Augmented State. The human's policy $\pi_{\mathcal{H}}(s, b)$ and robot's policy $\pi_{\mathcal{R}}(s, b, \theta)$ depend on the system state s and the human's belief b. The state transitions according to Equation (1) and the belief transitions according to Equation (2). Here we combine these two equations into a single dynamics:

$$\left(s^{t+1}, b^{t+1}\right) = F\left(s^t, b^t, a^t_{\mathcal{H}}, a^t_{\mathcal{R}}\right) \tag{3}$$

where (s, b) is the *augmented state* and F is the augmented dynamics under which the state and belief evolve. In practice, F is not a new equation: we are just evaluating Equation (1) and Equation (2) to find the next state-belief pair (s, b) given the human takes action $a_{\mathcal{H}}$ and the robot takes action $a_{\mathcal{R}}$.

Harsanyi-Bellman *Ad Hoc* Coordination. Recent research finds optimal policies for stochastic Bayesian games through a recursive combination of Bayesian Nash equilibrium and the Bellman equation. This method is referred to as Harsanyi-Bellman *Ad Hoc* Coordination (HBA) [22]. Define $V(s, b)$ as the *value* of the augmented state (s, b), i.e., the total reward of starting in (s, b) and acting optimally thereafter. Because both of the agents in our setting have common payoffs, and because there is only uncertainty about the robot's type, HBA applied to our context becomes:

$$V\big(s, b\big) = r\big(s\big) + \underbrace{\max_a \sum_{i=1}^{N} b(\theta_i) \cdot V\Big(F\big(s, b, a_{\mathcal{H}}, a_{\mathcal{R},i}\big)\Big)}_{\text{Bayesian Nash Equilibrium at }(s,b)} \tag{4}$$

where $a_{\mathcal{R},i}$ is the action assigned to the i-th type of robot, and $a = (a_{\mathcal{H}}, a_{\mathcal{R},1}, a_{\mathcal{R},2}, \ldots, a_{\mathcal{R},N})$ includes the human's action and an action for each of the N types of robots. We will refer to the action a that maximizes the underlined portion of Equation (4) as $a^* = (a^*_{\mathcal{H}}, a^*_{\mathcal{R},1}, a^*_{\mathcal{R},2}, \ldots, a^*_{\mathcal{R},N})$.

Overall, Equation (4) is an instance of the Bellman equation [23]. The value $V(s, b)$ is equal to the immediate reward at s plus the maximum expected value of the next augmented state. This expectation is taken over the human's belief in the robot's type: remember that $b(\theta_i)$ is the probability — from the human's perspective — that the robot is type i. Within the underlined portion of Equation (4) we find the Bayesian Nash Equilibrium at the current augmented state (s, b). Specifically, we find an action $a^* = (a^*_{\mathcal{H}}, a^*_{\mathcal{R},1}, a^*_{\mathcal{R},2}, \ldots, a^*_{\mathcal{R},N})$ for the human and each type of robot such that a^* maximizes the next value given the human's current belief b. Intuitively, the human is not sure which robot type they are dealing with: each type may take a different action $a_{\mathcal{R}}$, and the rational human identifies an action $a_{\mathcal{H}}$ that will maximize the expected value across robot types. In practice we can solve Equation (4) by either (a) discretizing the states and actions and applying classical value iteration approaches [23], or by (b) leveraging recent algorithms that approximate the value function in continuous spaces [24]. In either case our output is the value V at each augmented state (s, b).

Now that we have $V(s, b)$ from Equation (4) we will use this value to find optimal human and robot policies $\pi_{\mathcal{H}}$ and $\pi_{\mathcal{R}}$. Let a^* be the Bayesian Nash Equilibrium at (s, b):

$$a^* = \arg\max_{a} \sum_{i=1}^{N} b(\theta_i) \cdot V\Big(F(s, b, a_{\mathcal{H}}, a_{\mathcal{R},i})\Big) \tag{5}$$

where $a^* = (a^*_{\mathcal{H}}, a^*_{\mathcal{R},1}, \ldots, a^*_{\mathcal{R},N})$ assigns an action to each type of robot, so that robot type θ_i takes action $a^*_{\mathcal{R},i}$. The optimal human and robot take their respective actions within this Bayesian Nash Equilibrium:

$$\pi_{\mathcal{H}}(s, b) = a^*_{\mathcal{H}}, \quad \pi_{\mathcal{R}}(s, b, \theta_i) = a^*_{\mathcal{R},i} \tag{6}$$

Overall, Equations (4)–(6) solve our stochastic two-player Bayesian game to find the optimal pair of policies for the human and robot. In practice, we recognize that actual human users may deviate from their optimal policy $\pi_{\mathcal{H}}$. Humans that exactly follow $\pi_{\mathcal{H}}$ are *rational humans*: in our definition of *rationally opaque* we assume that the ideal human sticks to $\pi_{\mathcal{H}}$. We also emphasize that, when solving Equations (4)–(6) for the robot's optimal policy $\pi_{\mathcal{R}}$, we have assumed the human partner acts rationally.

C. *Proving Opaque Behavior can be Optimal*

Given the HBA algorithm for finding optimal robot policies, we are ready to revisit our fundamental question: is it ever optimal for robots to *hide* their type? Here we apply Equations (4)–(6) to a simulated 1-DoF human-robot team[1]. We first demonstrate that it *can* be optimal for collaborative robots to be fully opaque. Next, we explore the distinction between fully and rationally opaque, and prove that robots which are rationally opaque *may not be* fully opaque.

Example Problem. Consider a 1-DoF version of our motivating example where the human and robot are trying to reach for a block (see Figure 2). The state s is the position of the human-robot team. This state is bounded between $[0, 2]$, where $s = 0$ is the position of the block closer to the robot and $s = 2$ is the position of the farther block. At each timestep the human can take actions $\mathcal{A}_{\mathcal{H}} = \{-0.2, 0, +0.2\}$ to reach left or right. There are two types of robots: a *capable* robot (θ_1) which can reach for either block, and a *confused* robot (θ_2) that can only move for the block at $s = 0$. The capable robot's action set is $\mathcal{A}_{\mathcal{R},1} = \{-0.1, +0.1\}$ and the confused robot's action set is $\mathcal{A}_{\mathcal{R},2} = \{-0.1\}$. The human increases their belief b that the robot is capable if they observe $a_{\mathcal{R}} = +0.1$. The game ends after a total of $T = 5$ timesteps. The human and robot receive reward $r = +1$ if they end the game at state $s = 0$ (the closer block) $r = +2$ if they end the game at $s = 2$ (the farther block). At all other states the reward is $r = 0$.

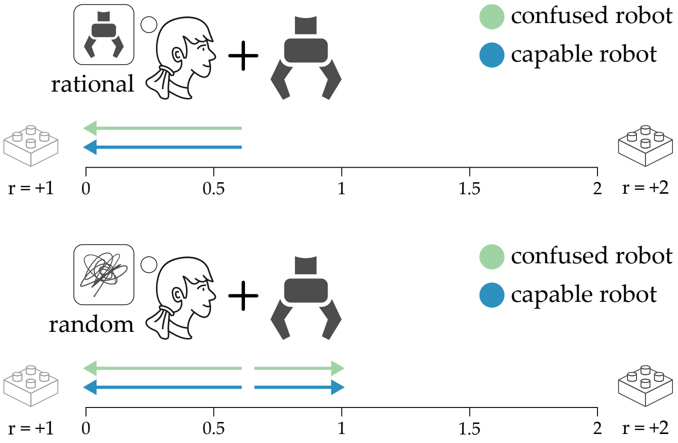

Fig. 2. Example of an optimal, fully opaque robot. The system starts at position $s = 0.6$ and prior $b^0 = 0.2$ (i.e., the human is initially 80% confident the robot is confused). The confused robot can only move towards the left block. (Top) Optimal human and robot solve this stochastic Bayesian game. (Bottom) Optimal robot is paired with a human that takes random actions. Regardless of the human's actions, both capable and confused robots always move towards the left block. Hence, the robot is *fully opaque*.

Known Type. Imagine that the robot's type θ is public knowledge. If we are interacting with a capable robot, both the human and robot should move right at every timestep ($a_{\mathcal{H}} = +0.2$ and $a_{\mathcal{R}} = +0.1$) to reach the farther block (reward $r = +2$). Alternatively, if the human knows they are interacting with the confused robot and they start at a state $s < 1.5$, then they cannot reach the farther block and should move left for the closer block ($a_{\mathcal{H}} = -0.2$ and $a_{\mathcal{R}} = -0.1$). It may therefore seem optimal for the robot to reveal its type so that the human can determine which block to aim for. As we will show, however, this is not always the case.

Optimal Robots can be Fully Opaque. Let initial system state be $s^0 = 0.6$ and let the prior be $b^0(\theta_1) = 0.2$ (i.e., the human is 20% sure the robot is capable). We solve Equations (4)–(6) to find the optimal robot policy $\pi_{\mathcal{R}}$. We then pair this optimal robot with two different humans: a *rational* human that follows the optimal human policy $\pi_{\mathcal{H}}$, and a *random* human that can take any action (see Figure 2). We find that — no matter what action the human takes — the optimal action for both types of robots is to move left and reach for the closer goal. Put another way, when starting at this (s^0, b^0) both robot types always take action -0.1. Rational or random humans cannot distinguish the robot's type: at the end of the interaction

[1] See the complete implementation of this example here: https://github.com/VT-Collab/opaque-example

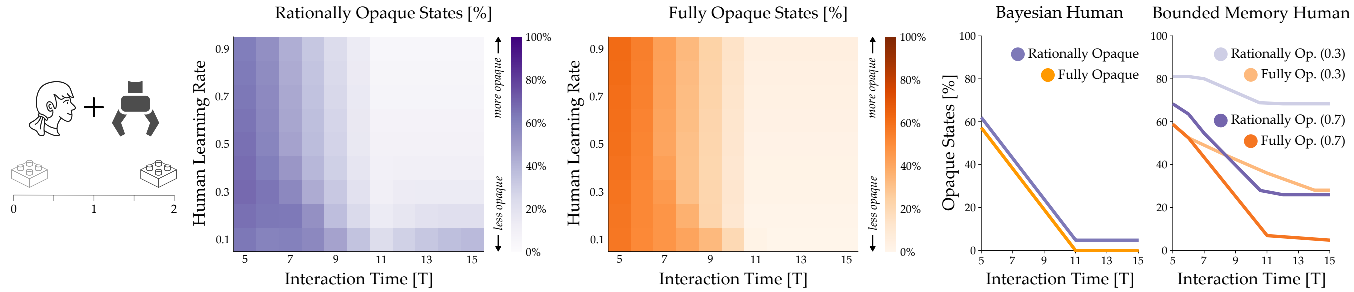

Fig. 3. Simulation results from our 1-DoF Environment. (Left) The human and robot collaborated to reach a goal; the confused robot could only go left while the capable robot could help reach right or left. For each plot we sampled all start states and priors and then calculated the percentage of those augmented states which were opaque; e.g., 50% opaque means that for half of the initial augmented states it was *optimal* for the robot to withhold its type from the human. (Middle) We varied the human's learning rate and the total number of timesteps in each interaction. A higher learning rate indicated that the human uncovered θ more quickly when the actions for each robot type diverged. (Right) We also tested a human that used Bayesian inference to update their belief and two bounded memory humans (with learning rates of 0.3 and 0.7) that forgot what they had learned after each timestep.

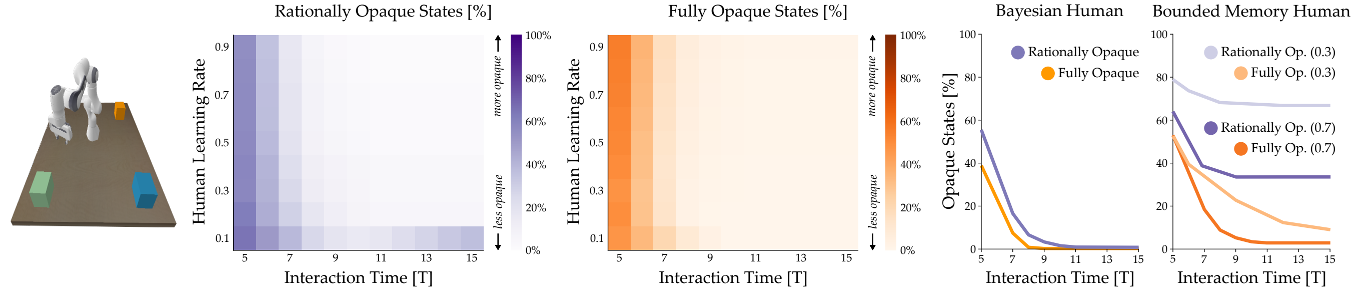

Fig. 4. Simulation results from our robot arm environment. (Left) Humans shared control with a robot arm to reach for goals on the table; the capable robot could go towards any goal while the confused robot could only move down and left. The format of our results follows Figure 3. (Middle) The number of opaque states decreases as the interaction time increases. (Right) The number of opaque states also decreases as the human learns more quickly. Note that the Bayesian human is an *ideal* user that can infer the robot's type from a single timestep; i.e., this human model learns θ as efficiently as possible. When compared to this ideal human, it is more likely for opaque behavior to be optimal when the robot is collaborating with a forgetful user that follows the bounded memory model. Overall, our results show that opaque behavior is more likely to be optimal during short interactions with suboptimal humans.

the human's belief is $b^T(\theta_1) = 0$ after working with both capable and confused robots. Accordingly, given the initial augmented state $s^0 = 0.6$ and $b^0(\theta_1) = 0.2$ the robot's *optimal* behavior is *fully opaque*.

Optimal Robots can be Rationally Opaque. We next prove that an optimal robot may be rationally opaque *but not* fully opaque. Let the initial state be $s^0 = 1.0$ with prior $b^0(\theta_1) = 0.2$ (i.e., the 1-DoF system starts farther to the right than before). When the optimal robot interacts with a rational human the system again reaches for the closer block; both capable and confused robots always take action $a_{\mathcal{R}} = -0.1$, and the rational human's final belief is identical across both types of robots. But this changes when the human is free to take any action. If the random human takes action $a_{\mathcal{H}} = 0.2$ the capable robot switches direction to go towards the farther goal and obtain reward $r = +2$. The random human's final belief is $b(\theta_1) = 0.4$ when interacting with the capable robot and $b(\theta_1) = 0$ with the confused robot. Hence, for this initial state the robot's behavior is only *rationally opaque*.

V. What Conditions Lead to Opaque Robots?

When collaborative robots solve for optimal policies we have shown these optimal policies can be opaque. But what aspects of the problem setting lead to opaque robot behaviors? Here we conduct controlled experiments with simulated humans. We vary the time horizon of the interaction, the human's learning rate, and the way the human learns from the robot's behavior. Across two simulated environments, we observe that shorter interaction times and lower learning rates result in a higher number of optimally opaque states.

Environments. Our simulated environments are shown in Figure 3 and Figure 4. The 1-DoF human-robot team matches the example from Section IV-C with one slight difference: now the human actions $\mathcal{A}_{\mathcal{H}} = \{-0.1, 0, +0.1\}$ have the same magnitude as the robot actions. As a reminder, in this 1-DoF setting the human and robot are collaborating to reach one of the goals, and the confused robot can only move left.

We extend this environment to create a **robot arm** simulation (see Figure 4). As before, the human and robot must collaborate to reach a goal. There are three different goals on the table; the confused robot can only move down or to the left, while the capable robot can autonomously move in any direction. Goals that are farther away from the robot's base have a higher reward. However, the confused robot may not be able to coordinate with the human to reach these goals, and thus the human needs to determine the robot's type to figure out which goal to aim for.

Procedure. For each environment, time horizon, and simulated human we first solve Equations (4)–(6) to find the optimal policies. We then sample all the discrete states and priors, and test whether the augmented start states (s^0, b^0) are fully opaque, rationally opaque, or neither. In what follows we report (a) the percentage of *rationally opaque* start states and (b) the percentage of *fully opaque* start states.

A. Varying Interaction Time

Remember that the human-robot interaction ends after T total timesteps. When interactions are short (i.e., as $T \to 0$) it may not make sense for the robot to take the time to reveal its type θ to the human: here the robot must leverage its actions to complete the task. Conversely, when interactions are long (i.e., as $T \to \infty$) the robot should take actions to communicate its type: these transparent actions may reduce rewards in the short-term but can facilitate collaboration and improve overall, long-term reward. Building on this intuition, we conducted experiments where we held the simulated human constant and varied the interaction time T. Our results across the x-axes of Figure 3 and Figure 4 generally follow the expected trend: the percentage of rationally opaque and fully opaque states *decreases* as T *increases*.

B. Varying Human Learning Rate

The human infers the robot's type based on the robot's behaviors. For example, if the *capable* robot takes an action that is not possible for the *confused* robot, then the human should become more confident that $\theta = \theta_{capable}$. Here we adjust how rapidly the simulated human's belief changes during a single timestep (i.e., the human's *learning rate*). For a learning rate of 0.1 the next belief $b^{t+1}(\theta) = b^t(\theta) \pm 0.1$, and for a learning rate of 0.9 the belief similarly updates in increments of 0.9. Increasing the learning rate corresponds to a human that is more sensitive to differences in robot behavior. From Figures 3 and 4, we find that the percentage of rationally opaque and fully opaque states *decreases* as the learning rate *increases*. We explain this result in connection with the interaction length T: as the learning rate decreases it takes the robot more timesteps to reveal its type θ, making transparent behavior less efficient. At the extreme the human does not learn at all from the robot's actions: in this case, there is no advantage to transparency.

C. Varying How the Human Learns

We finally test two alternative human models. First, we simulate an *ideal* human that leverages Bayesian inference to update their belief. Here the human knows the robot's policy and treats $\pi_{\mathcal{R}}$ as the likelihood function in Equation (2). This ideal human learns as quickly as possible: at augmented states where $\pi_{\mathcal{R}}(s, b, \theta_i)$ and $\pi_{\mathcal{R}}(s, b, \theta_j)$ output different actions, the Bayesian human immediately distinguishes types θ_i and θ_j. Second, we apply the bounded memory model [25] to simulate a human that only remembers the robot's most recent behavior (i.e., the robot's last action). This human updates their belief b during each timestep at a fixed learning rate, and then resets b to the prior between timesteps (i.e., this simulated human *forgets* what they have learned). Our results for ideal and forgetful humans are shown on the right side of Figures 3 and 4. When the interaction lasts only a few timesteps T, even for an ideal human learner it is still optimal for the robot to select opaque behaviors. As the duration of the interaction *increases* we again find that the percentage of opaque states *decreases* for both the ideal and forgetful humans. We also find that bounded memory humans with a lower learning rate lead to an increased number of opaque states.

Overall, our simulations suggest that it is more likely for opaque robot behavior to be optimal when (a) the interaction is brief and (b) the human learns slowly.

VI. User Studies

In this section we put opaque robots to the test by pairing them with *actual* humans. Our analysis and simulations suggest that opaque behaviors can be optimal: i.e., humans can gain higher rewards when their robot partners withhold latent information. Here we explore whether this is really the case, and measure the human's performance with opaque and transparent robots. But even if the human does get higher scores with an opaque robot, they may not *like* working with this optimal robot; in other words, are users willing to accept less reward if it helps them identify the robot's type? To test both hypotheses we performed two separate user studies: an online user study with autonomous driving, and an in-person experiment with collaborative block stacking.

Independent Variables. In both studies there were two types of robots: a *confused* robot and a *capable* robot. During each interaction the robot's actual type was randomized: half the time people worked with *confused*, and the other interactions were with *capable*. As such, participants were unsure about the robot's current type θ. We compared our proposed game-theoretic approach (**Opaque**) to a **Transparent** baseline [4], [5]. The **Transparent** algorithm modified the robot's reward to incentivize revealing actions; more formally, in [4], [5] the robot assigned a bonus reward to actions that conveyed θ to the user. By contrast, the reward for **Opaque** was the exact same as the human's reward. **Opaque** applied Equations (4)–(6) to find optimal, opaque behaviors.

Dependent Measures. We measured the team's final *Reward*. Here reward corresponded to the score in each environment: e.g., the distance the shared car traveled, or the height of the block tower. We displayed the human's reward in real-time so that participants could track their own performance. We also measured the human's subjective *Preference* for each robot algorithm. Preference was measured on a 1-7 scale, where higher scores indicated the human had a stronger preference for working with that robot algorithm.

Hypotheses. We had two main hypotheses:

> **H1.** *Humans will obtain more reward with* ***Opaque*** *robots than with* ***Transparent*** *robots.*
>
> **H2.** *Humans will prefer* ***Opaque*** *robots, even though these robots withhold information.*

A. Online: Sharing Control of an Autonomous Car

We first conducted an online survey where participants collaborated with a virtual robot to share control over a car. At each timestep the human the human clicked their input $a_{\mathcal{H}}$, the robot selected its action $a_{\mathcal{R}}$, and then the autonomous car moved using the combined action $a = a_{\mathcal{H}} + a_{\mathcal{R}}$.

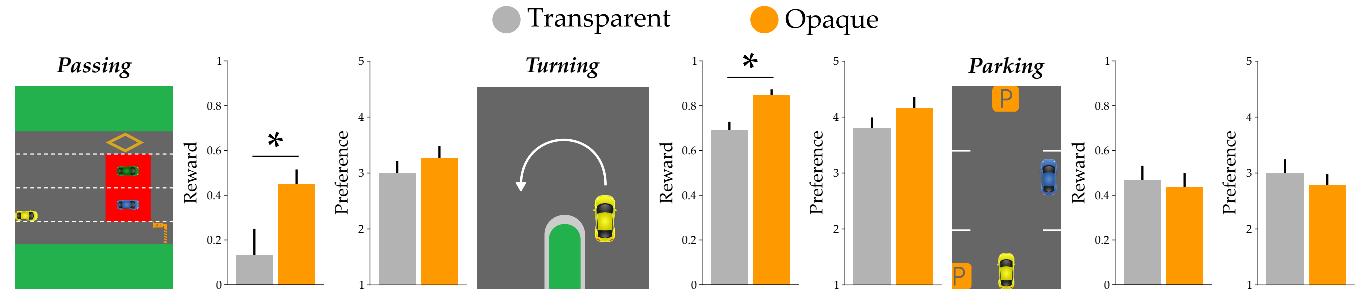

Fig. 5. Task results from our online user study. Participants collaborated with a virtual agent to drive a car in *Passing*, *Turning*, and *Parking* environments. Error bars show standard error and an $*$ denotes statistical significance ($p < .05$).

Experimental Setup. Participants teamed up with the robot to drive in three settings: passing, turning, and parking (see Figure 5). In *Passing* the team got rewards for making lane progress, staying on the road, and avoiding a collision. In *Turning* the reward was the car's velocity plus the total angle the car turned. Finally, in *Parking* the team obtained a reward for either (a) parallel parking directly above the start position or (b) driving straight ahead to an open parking place.

Participants completed every scenario with confused and capable **Opaque** robots, as well as with confused and capable **Transparent** robots. Interactions lasted three timesteps. During each timestep we first showed the participant an image of the current state s^t and prompted the user to select their action (e.g., turning left, accelerating forward). After the user selected their action $a^t_{\mathcal{H}}$, the robot acted simultaneously with $a^t_{\mathcal{R}}$ and showed an image of the next state s^{t+1}. We displayed the user's score throughout the game, and then at the end asked if the user "preferred sharing control with this car."

Participants. For the online study we recruited 44 anonymous participants. All participants read the instructions prior to starting the study; we then asked qualifying questions to test their understanding. Below we report the results for the 30 participants who passed these questions and completed the survey. Each participant performed 12 driving interactions (3 scenarios, twice with **Opaque** and twice with **Transparent**). The order of presentation was counterbalanced so that half of the users started each scenario with the **Opaque** robot and the other half started with **Transparent**.

Results. See Figures 5 and 6. Taken across all tasks and both robot types, the team's overall reward was significantly higher for **Opaque** than for **Transparent** ($t(179) = 2.62$, $p < .05$). Looking more specifically at the individual tasks, in both *Passing* ($t(59) = 2.37$, $p < .05$) and *Turning* ($t(59) = 3.33$, $p < .05$) users reached higher rewards when paired with **Opaque** partners. These results support hypothesis **H1** and suggest that real participants can reach higher rewards when collaborating with optimal but opaque robots.

On the other hand, the users' perceptions of the robots were divided. Across all three tasks the differences in preference scores were not significant ($t(179) = 1.06$, $p = .29$). We initially thought that there might be a bi-modal split in participants: if half strongly preferred opaque robots and the other half strongly preferred transparent, then these two halves would cancel out when computing the average scores. But calculating each individual's preferences in Figure 6 we find that most users are roughly on the fence. Hence, in response to hypothesis **H2** we have that online users did not clearly prefer either **Opaque** or **Transparent** robots.

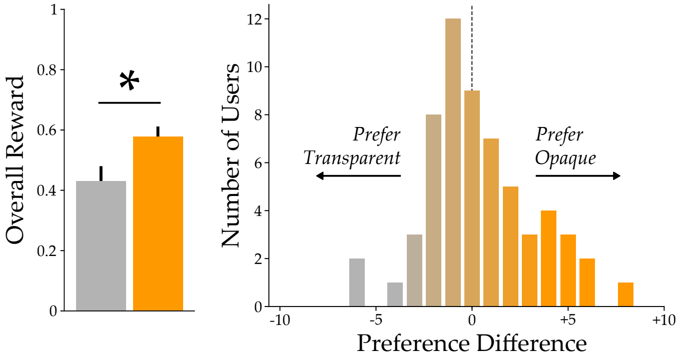

Fig. 6. Overall results from our online user study. (Left) The average reward across all three tasks. (Right) 30 participants completed each task twice with **Opaque** and twice with **Transparent**, resulting in 60 pairs of datapoints. For each pair we subtracted the total preference scores with **Transparent** from the total preference scores with **Opaque**. Positive numbers indicate that individual user ranked **Opaque** as better than **Transparent**, and negative values indicate the opposite. We found that most users perceived the two algorithms as roughly equal (Preference Difference near zero).

B. In-Person: Stacking Blocks with a Robot Arm

In our second user study in-person participants collaborated with a robot arm to build a tower (see Figure 1). At each timestep the human picked up and added one block to the tower, $a_{\mathcal{H}}$, and then the robot stacked its block on top, $a_{\mathcal{R}}$. The state s was the sequence of blocks in the tower.

Experimental Setup. We placed blocks of two different sizes and four different colors near the human and robot (see Figure 1). The capable robot could pick up any of the blocks, while the confused robot was only able to stack the smaller blocks. At each timestep the human-robot team obtained a reward of $+5$ if they picked the same block, and -5 if they picked different blocks. Each taller block that was stacked on the tower added a reward of $+1$. In Figure 7 we normalize the reward between 0 and 1. To prevent the human from cheating (i.e., changing their block in response to the robot) the robot waited to move until after the human.

Each participant built towers with confused and capable **Opaque** robots, and with confused and capable **Transparent** robots. In total participants built 8 towers where each tower contained 6 blocks. We displayed the user's current score in real-time on a monitor next to the towers.

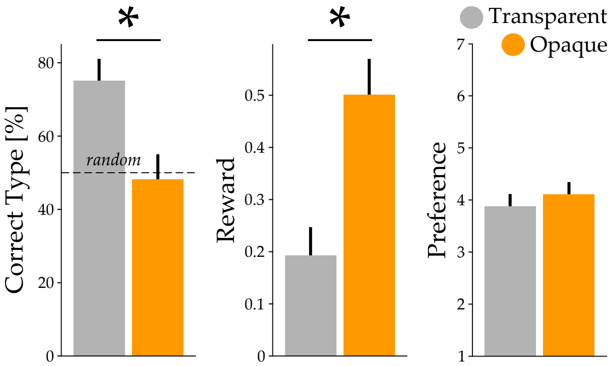

Fig. 7. Results from our in-person user study. Participants collaborated with a robot arm to stack blocks (see Figure 1). *Correct Type* is the percentage of trials where users correctly identified that the robot was capable or confused: by guessing randomly users would be right 50% of the time.

Participants and Procedure. We recruited 13 participants from the Virginia Tech community (5 female, ages 24.31±3.45 years). These participants provided informed consent under IRB#20-755 and did not take part in the online study. We used a performance-based compensation model. Every user received a \$10 gift card for taking part in the study; for each tower they built with a reward higher than 10 points, they received a performance bonus of ¢50 (USD).

After completing each tower we asked users if they could determine which type of robot they had worked with (i.e., capable or confused). We grouped the towers into pairs of two, where each pair contained one **Opaque** robot and one **Transparent** robot. After every pair we asked if the user preferred the first robot or the second robot. The algorithms and robot types were presented in a counterbalanced order. Participants were never told which robots were confused and which robots were capable.

Results. See our video and Figure 7. To confirm the **Opaque** algorithm withheld the robot's type — and the **Transparent** robot revealed its type — we first recorded the percentage of trials where participants correctly guessed whether the robot was confused or capable. Users that guessed completely at random would be right 50% of the time: with **Opaque** user guesses were on par with random (48% correct), while with **Transparent** users identified θ correctly 75% of the time.

We then moved on to evaluate hypotheses **H1** and **H2**. Our results here are in line with the online user study: users reached higher rewards when working with **Opaque** partners ($p < .001$), and participants rated **Opaque** robots about the same as **Transparent** partners ($t(51) = .48$, $p = .63$).

VII. Conclusion

We have introduced theoretical and experimental analysis to demonstrate that collaborative robots *should not always* be transparent. We derived a modified version of Harsanyi-Bellman *ad hoc* coordination to identify optimal robot policies in stochastic Bayesian games. When applied to human-robot interactions with short time horizons or gradually learning humans, we proved that *opaque* robot behavior becomes optimal. In user studies participants reached higher scores with opaque partners, and did not perceive opaque robots as subjectively worse than their transparent counterparts.

References

[1] T. Hellström and S. Bensch, "Understandable robots–what, why, and how," *Journal of Behavioral Robotics*, vol. 9, pp. 110–123, 2018.

[2] S. Sebo, B. Stoll, B. Scassellati, and M. F. Jung, "Robots in groups and teams: A literature review," *ACM Transactions on Human-Computer Interaction*, vol. 4, pp. 1–36, 2020.

[3] G. Hoffman, "Evaluating fluency in human–robot collaboration," *IEEE Transactions on Human-Machine Systems*, vol. 49, pp. 209–218, 2019.

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

[5] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa, "Effects of robot motion on human-robot collaboration," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2015, pp. 51–58.

[6] C. Bodden, D. Rakita, B. Mutlu, and M. Gleicher, "A flexible optimization-based method for synthesizing intent-expressive robot arm motion," *The International Journal of Robotics Research*, vol. 37, no. 11, pp. 1376–1394, 2018.

[7] S. Habibian and D. P. Losey, "Encouraging human interaction with robot teams: Legible and fair subtask allocations," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6685–6692, 2022.

[8] S. A. Wu, R. E. Wang, J. A. Evans, J. B. Tenenbaum, D. C. Parkes, and M. Kleiman-Weiner, "Too many cooks: Bayesian inference for coordinating multi-agent collaboration," *Topics in Cognitive Science*, vol. 13, no. 2, pp. 414–432, 2021.

[9] A. Roncone, O. Mangin, and B. Scassellati, "Transparent role assignment and task allocation in human robot collaboration," in *IEEE Int. Conference on Robotics and Automation*, 2017, pp. 1014–1021.

[10] M. Kwon, S. H. Huang, and A. D. Dragan, "Expressing robot incapability," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2018, pp. 87–95.

[11] S. Sreedharan, A. Kulkarni, and S. Kambhampati, "Obfuscatory behavior and deceptive communication," in *Explainable Human-AI Interaction: A Planning Perspective*, 2022, pp. 121–136.

[12] A. Dragan, R. Holladay, and S. Srinivasa, "Deceptive robot motion: Synthesis, analysis and experiments," *Autonomous Robots*, 2015.

[13] S. Nikolaidis, S. Nath, A. D. Procaccia, and S. Srinivasa, "Game-theoretic modeling of human adaptation in human-robot collaboration," in *ACM/IEEE international Conference on Human-Robot Iteraction*, 2017, pp. 323–331.

[14] S. Bansal, J. Xu, A. Howard, and C. Isbell, "Bayes–Nash: Bayesian inference for nash equilibrium selection in human-robot parallel play," *Autonomous Robots*, vol. 46, no. 1, pp. 217–230, 2022.

[15] L. Peters, D. Fridovich-Keil, C. J. Tomlin, and Z. N. Sunberg, "Inference-based strategy alignment for general-sum differential games," in *International Conference on Autonomous Agents and MultiAgent Systems*, 2020, pp. 1037–1045.

[16] S. V. Albrecht, J. W. Crandall, and S. Ramamoorthy, "Belief and truth in hypothesised behaviours," *Artificial Intelligence*, pp. 63–94, 2016.

[17] S. Sagheb, Y.-J. Mun, N. Ahmadian, B. A. Christie, A. Bajcsy, K. Driggs-Campbell, and D. P. Losey, "Towards robots that influence humans over long-term interaction," in *IEEE International Conference on Robotics and Automation*, 2022.

[18] 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*, 2016.

[19] T. Willi, A. H. Letcher, J. Treutlein, and J. Foerster, "Cola: Consistent learning with opponent-learning awareness," in *International Conference on Machine Learning*, 2022, pp. 23 804–23 831.

[20] D. P. Losey and D. Sadigh, "Robots that take advantage of human trust," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2019, pp. 7001–7008.

[21] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning," in *AAAI*, 2008.

[22] S. V. Albrecht and S. Ramamoorthy, "A game-theoretic model and best-response learning method for ad hoc coordination in multiagent systems," in *International Conference on Autonomous Agents and Multi-Agent Systems*, 2013, pp. 1155–1156.

[23] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Pearson Education, Inc., 2022.

[24] M. Lutter, S. Mannor, J. Peters, D. Fox, and A. Garg, "Value iteration in continuous actions, states and time," in *International Conference on Machine Learning*, 2021, pp. 7224–7234.

[25] S. Nikolaidis, A. Kuznetsov, D. Hsu, and S. Srinivasa, "Formalizing human-robot mutual adaptation: A bounded memory model," in *ACM/IEEE Int. Conf. on Human-Robot Interaction*, 2016, pp. 75–82.