

Should Collaborative Robots be Transparent?

Shahabedin Sagheb^{1*}, Soham Gandhi¹ and Dylan P. Losey¹

^{1*}Department of Mechanical Engineering, Virginia Tech, 635 Prices Fork Road,
Blacksburg, 24061, Virginia, USA.

*Corresponding author(s). E-mail(s): shahab@vt.edu;
Contributing authors: sgandhi25@vt.edu; losey@vt.edu;

Abstract

We often assume that robots which collaborate with humans should behave in ways that are transparent (e.g., legible, explainable). These transparent robots intentionally choose actions that convey their internal state to nearby humans: for instance, a transparent robot might exaggerate its trajectory to indicate its goal. But while transparent behavior seems beneficial for human-robot interaction, 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*. In contrast to transparent robots, opaque robots select actions that *withhold* information from the human. 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. We extend this theoretical analysis to user studies across **43** total participants in both online and in-person settings. We find that — during short interactions — 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>

Keywords: Human-Robot Cooperation, Expectation and Intention Understanding, Legibility, Game Theory

1 Introduction

Optimal robots select actions to maximize their objective function. When robots collaborate alongside humans, maximizing this objective requires teamwork: the robot must reason about how the human interprets and reacts to the robot’s behavior in order to seamlessly complete the overall task [37, 13, 6, 15]. In this paper we focus on settings where the robot and the human share the same objective function (i.e., the robot and human are working together to perform a task), but the human does not know exactly how the robot will

behave. This applies to factory floors, manufacturing, and assembly contexts where everyday human workers must collaborate with robot partners [33, 27, 26, 23]. For example, imagine a person teaming up with a robot arm to build a block tower (see Figure 1). Both the human and robot share the same objective: they are trying to maximize the tower’s height without it falling over. But the human is not sure about the robot’s capabilities. If the robot is *capable* it can reach for any block and add it to the tower; on the other hand, if the robot is *confused* it will only be able to add the closer, smaller blocks. Whether the robot

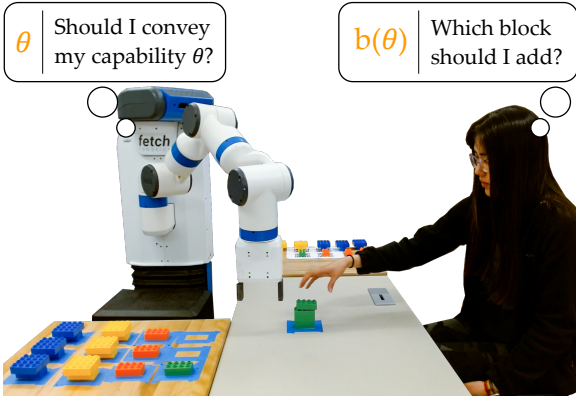


Fig. 1 Collaborative block-stacking task where the human is uncertain about the robot’s internal state θ . Transparent robot actions help the human learn θ and decide what blocks to add to the tower. However, we find that the costs of this transparent behavior may outweigh its benefits

is capable or confused affects the human’s optimal decisions: when working with a capable robot the human should add larger blocks, but for a confused robot the human needs to add smaller blocks to keep the tower from becoming unstable and falling over. Given this uncertainty, it seems intuitive that the robot’s optimal behavior is to pick up blocks that reveal whether it is capable or confused.

In line with this intuition, today’s approaches to human-robot interaction often assume that robot behavior should be *transparent* (e.g., legible, explainable, understandable) [16, 34, 17, 36]. 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 [11, 10, 7]. 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 [14, 39, 29, 18]. 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 get to the block and complete the task. Going one step further, the human teammate may require multiple interactions before they correctly interpret what the robot is trying to convey and update their own behavior in response.

In this paper we explore the situations where transparent behavior is optimal for human-robot teams. We build on related works to introduce *opacity* as the opposite of transparency: opaque robots select actions that withhold information from the human. To determine whether it is optimal for robots to withhold information and select opaque behaviors, or to convey information and select transparent 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, we will prove that under some conditions 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 robot is opaque, and does not take actions to convey to its capabilities to the human.

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 this context we build on prior works to define fully and rationally opaque behavior.

Proving when Opacity is Optimal. We extend a recursive combination of Bayesian Nash Equilibrium and the Bellman equation to find optimal robot policies. We then show that these optimal policies can be opaque. Our analysis and simulations 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 during shorter interactions. Users prefer opaque robots about equally to transparent partners.

2 Related Work

Transparent Robots. Prior work on human-robot interaction often assumes that robots should be *transparent* (e.g., legible or explainable) [16, 34, 17, 11]. Transparent robots actively and intentionally reveal their internal state to nearby humans so that these humans have a more accurate estimate of 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 [10], indicate their intended trajectory [7], and express whether or not they are capable of performing a task [19]. 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 [14, 39, 29].

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 [35, 9]. 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 [25, 5, 28, 1]. By solving this game we seek to identify whether optimal robot behavior reveals or withholds information about the robot’s latent state. When solving similar games, recent research has found that it can be optimal for robots to take actions that influence humans [32, 31, 38] 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 [32] 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.

3 Problem Formulation

We consider industrial applications where a robot is working alongside a human partner. 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 we assume that the human and robot know this reward function (i.e., both agents know the task), 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 building a block tower with a robot arm: the robot can either reach any block (type capable) or the robot can only reach for nearby blocks (type confused). The human does not initially know the robot’s type θ and the robot must decide whether to take actions that reveal its type during interaction.

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:

$$s^{t+1} = f(s^t, a_{\mathcal{H}}^t, a_{\mathcal{R}}^t) \quad (1)$$

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 $a_{\mathcal{R}}^t$ and $a_{\mathcal{H}}^t$ 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. In our user studies we focus on cases where $N = 2$ and there two possible types of robots (to better measure opacity and transparency), but the formulation we

present extends to settings with an arbitrary number of robot types. At the start of each interaction θ is sampled from the prior $P(\theta)$. The human knows this prior distribution $P(\theta)$ but not the robot’s current type θ .

Belief. The human updates their estimate of the robot’s type during the interaction based on the robot’s behavior. Let $b^{t+1}(\theta) = P(\theta \mid s^{0:t}, a_{\mathcal{R}}^{0:t})$ be the probability that the robot is type θ given that we have visited states $s^{0:t}$ and the robot has taken actions $a_{\mathcal{R}}^{0:t}$ up to the current timestep t . Similar to prior work [40], we assume that the human updates this belief using Bayesian inference:

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

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 . Our approach is not tied to a specific choice of this likelihood function. In our simulations (Section 5) we will test different models for $P(a_{\mathcal{R}} \mid s, \theta)$.

Reward. At each timestep the collaborative human and robot receive the same reward $r(s)$. This reward function captures the task that the agents are trying to complete (e.g., the height of the tower the human and robot are building). 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 complete the task and maximize their reward across the interaction. If we model the human as 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 [1]. We highlight two non-standard aspects of our game-theoretic formulation. First, the 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. Returning to our running example of building a tower: at a given (s, b) a capable robot could reach for the larger block while the confused robot might move for the closer, smaller block.

In Section 4 we will solve for the optimal robot policy. This policy $\pi_{\mathcal{R}}(s, b, \theta)$ will maximize the human’s and robot’s expected reward across an interaction.

4 Should Collaborative Robots be Opaque?

In this section we develop a solution to our stochastic Bayesian game to identify the robot’s optimal policy. Because the human and robot share the same reward function (and are therefore collaborating on a common task), it may seem reasonable to expect that the optimal robot behavior is *transparent*. Recall that transparent motions intentionally improve the human’s estimate of the robot’s latent state [11, 16]. Returning to our motivating example in Figure 1, we might expect the capable robot to take actions that reveal it can reach for larger blocks, and the confused robot to take actions that demonstrate it is limited to the nearby, smaller blocks. 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 4.1). Next, we derive a game-theoretic approach for finding *optimal* robot behavior in collaborative games (Section 4.2). 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 4.3).

4.1 Formalizing Robot Opacity

We start by introducing two definitions for opacity within stochastic Bayesian games where the human is uncertain about the robot’s type θ . These definitions are consistent with prior work on robot legibility and transparency [8]. 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 . Recall that $b(\theta)$ from Equation (2) is the human’s belief over the robot’s type θ , and the interaction has a total of T timesteps. Intuitively, a transparent system causes the human to reach different beliefs when interacting with different types of robots, so that $b^T(\theta_1) \neq b^T(\theta_2)$ [22]. By contrast, an opaque system causes humans to converge to the same final belief regardless of the robot’s actual type, so that $b^T(\theta_1) = b^T(\theta_2)$. More formally, we define Fully Opaque and Rationally Opaque robot’s below:

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

As an example, consider a robot where $a_{\mathcal{R}} = \pi_{\mathcal{R}}(s, b, \theta)$ is always zero; i.e., a robot that always takes action $a_{\mathcal{R}} = 0$. 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 $b^T(\theta_i) = b^T(\theta_j)$ for any choice of i and j . We define this robot as *fully opaque* because, no matter what the human does, they always converge to the same final 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). Define (s^0, b^0) as rationally opaque if — no matter which type $\theta \in \Theta$ the robot actually is — the rational human’s final belief $b^T(\theta)$ 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 can take random actions to perturb the system. 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 underlying type).

4.2 Identifying Optimal Behavior

Now that we have defined opacity, we will next determine if it is ever optimal for robots to be opaque. In Section 3 we formulated the human and robot as agents in a two-player stochastic Bayesian game. In this subsection we present an algorithm for finding optimal human and robot policies $\pi_{\mathcal{H}}$ and $\pi_{\mathcal{R}}$ under this game-theoretic setting. Next, in Section 4.3 we will apply this algorithm to example scenarios and demonstrate that the robot’s optimal behavior in these settings can be 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). Without loss of generality, we combine these two equations into a single dynamics:

$$(s^{t+1}, b^{t+1}) = F(s^t, b^t, a_{\mathcal{H}}^t, a_{\mathcal{R}}^t) \quad (3)$$

where (s, b) is the *augmented state* and F is the augmented dynamics under which the state and belief evolve. Note that F is not a new equation: we are evaluating Equation (1) and Equation (2) to find the next state-belief pair (s^{t+1}, b^{t+1}) given that 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) [2]. 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 reduces to:

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

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

Overall, Equation (4) is an instance of the Bellman equation [30]. 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: recall 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 $a^* = (a_{\mathcal{H}}^*, a_{\mathcal{R},1}^*, a_{\mathcal{R},2}^*, \dots, a_{\mathcal{R},N}^*)$ 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 [30], or by (b) leveraging recent algorithms that approximate the value function in continuous spaces [21]. 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) :

$$a^* = \arg \max_a \sum_{i=1}^N b(\theta_i) \cdot V\left(F(s, b, a_{\mathcal{H}}, a_{\mathcal{R},i})\right) \quad (5)$$

where $a^* = (a_{\mathcal{H}}^*, a_{\mathcal{R},1}^*, \dots, a_{\mathcal{R},N}^*)$ 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, such that:

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

Overall, Equations (4)–(6) solve the 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 *rationaly opaque* we assume that the human executes $\pi_{\mathcal{H}}$. 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.

4.3 Proving Opaque Behavior can be Optimal

By leveraging our modified HBA algorithm from Section 4.2 we can find the robot’s optimal policy and determine whether it is ever optimal for robots to be opaque. Here we apply Equations (4)–(6) to a simulated 1-DoF human-robot team¹. We first empirically 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 $r = 0$.

Known Type. Imagine that the robot’s type θ is public knowledge. If the human interacts 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

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

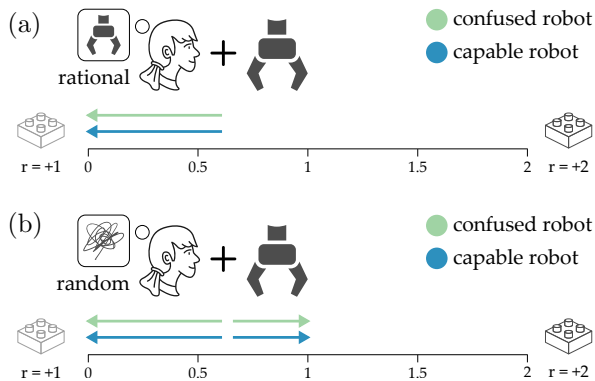


Fig. 2 Example of an optimal, fully opaque robot. The system starts at position $s = 0.6$ and prior $b^0 = 0.2$. The robot has two types θ : confused and capable. The confused robot can only move towards the left block. (a) Optimal human and robot solve this stochastic Bayesian game. (b) Optimal robot 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*, and the human cannot infer θ from the optimal robot’s actions

$r = +2$). 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 $a_{\mathcal{R}} = -0.1$). The human’s optimal actions depend on the robot’s type θ . 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 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. Moving for the closer block results in higher team reward, even though it does not convey information to the human.

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 optimal behavior is only *rationally opaque*.

5 What Conditions Lead to Opaque Robots?

In Section 4 we developed a method for solving our stochastic Bayesian game and identifying optimal robot behavior. We also showed that opaque robot behavior can be optimal in certain problem settings. In this section we analyze *what types of problem settings* lead to optimal, opaque robots. We conduct controlled experiments with simulated humans. We hypothesize that the duration of the interaction and the speed of the human’s learning will determine whether it is optimal for robots to be transparent or opaque. Accordingly, 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. All code for replicating these environments and reproducing our simulations is available online². The 1-DoF human-robot team matches the example from

²See <https://github.com/VT-Collab/opaque-example>

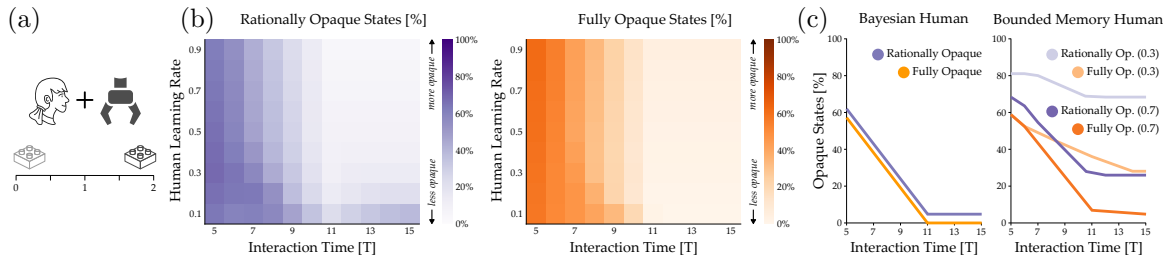


Fig. 3 Simulation results from our 1-DoF Environment. (a) The human and robot collaborated to reach a goal; the confused robot could only go left while the capable robot could help reach right or left. For each plot we sampled all start states and priors and then calculated the percentage of those augmented 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 θ from the human. (b) 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. (c) 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

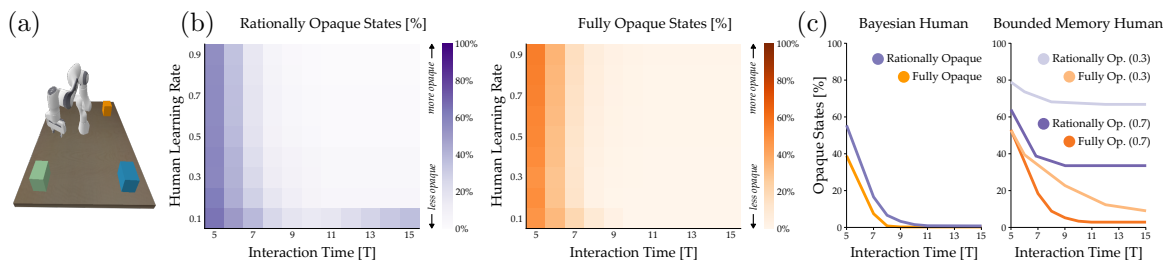


Fig. 4 Simulation results from our robot arm environment. (a) Humans shared control with a robot arm to reach for goals on the table; the capable robot could go towards any goal while the confused robot could only move down and left. The format of our results follows Figure 3. (b) The number of opaque states decreases as the interaction time increases. (c) 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

Section 4.3 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. One type of robot can reach either goal, and the other type of robot can only move left.

We extend this environment to create a **robot arm** simulation (see Figure 4). In accordance with our motivating example of assembling a tower, the human and robot must collaborate to reach blocks (e.g., goals) around the table. There are three different goals on the table; the confused robot (θ_1) can only move down or to the left, while the capable robot (θ_2) 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 robot 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 transparent. In what follows we report (a) the percentage of *rationally opaque* start states and (b) the percentage of *fully opaque* start states. Our overall results are summarized in Figure 3 and Figure 4.

5.1 Varying Interaction Time

The human-robot interaction ends after T total timesteps. To understand how the duration of the interaction affected the optimality of opaque behaviors, we held the simulated human constant

and varied the interaction time T . When interactions are short (i.e., as $T \rightarrow 0$) we find that the robot’s optimal behavior is *opaque* for an increasing number of initial conditions. Consider the x -axes of Figure 3 and Figure 4: the percentage of rationally opaque and fully opaque states *increases* as T *decreases*. Conversely, when interactions are long (i.e., as $T \rightarrow \infty$) the robot’s optimal behavior becomes *transparent*. Again looking at the x -axes of Figure 3 and Figure 4, the percentage of rationally opaque and fully opaque states *decreases* as T *increases*.

To explain this result we highlight a trade-off between completing the task and communicating hidden information. Using the robot’s actions to convey the robot’s type θ requires additional timesteps: the robot must exaggerate its motion (and potentially move away from goals) to indicate θ to the human teammate. But when the interactions are short, this time is not available — the robot must leverage all of its actions to directly complete the task and maximize the team’s reward. Hence, we suggest that transparent behavior is more suitable for settings where the human and robot will work together for long periods of time, while opaque behavior is more likely to be optimal when the human and robot are collaborating across a single-shot task.

5.2 Varying Human Learning Rate

The human updates their estimate $b(\theta)$ of the robot’s type θ as they observe the robot’s behaviors. For example, if the robot reaches for a block (i.e., a goal) that is farther away, the human should become more confident that the robot is capable of reaching these distant goals. During our controlled simulations we adjusted how slowly or rapidly the simulated human’s belief changed during a single timestep — i.e., we adjusted 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; by contrast, lowering the learning rate towards zero corresponds to a human that does not update their estimate regardless of the robot’s actions. Looking at Figures 3 and 4, we find that the percentage of rationally opaque and fully opaque states

decreases as the learning rate *increases*. When the simulated human learns more rapidly, *transparent* behavior becomes optimal. On the other hand, when the simulated human learns more gradually, it is optimal for robots to remain *opaque*.

We explain this result in connection with the interaction length T from Section 5.1. When the learning rate increases the robot can more rapidly convey information to the human, and thus the robot does not need to devote as many actions and timesteps to communicate its type θ . At the other end of the spectrum, when the human learns gradually the robot must commit multiple actions to convey θ , and during these actions the human’s response is not necessarily aligned with the robot’s capabilities. At the extreme the human does not learn at all from the robot’s actions: in this case, there is no advantage to transparency. Our simulation results suggest that transparent robots are more likely to be optimal when the human is sensitive to the robot’s behavior, and opaque behavior is more likely to be optimal when the human learns slowly or ignores the robot’s actions.

5.3 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 [4]. 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 [24, 3] 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.

These results demonstrate that opaque robot behavior is not tied to a specific model of human learning; we find that opaque behavior can be optimal when the human learns as quickly as possible (*Bayesian Human*) or when the human forgets previous actions (*Bounded Memory Human*). In line with the results from Section 5.1 and Section 5.2, we again observe the connection between interaction time T and the speed of the human’s learning. *Transparent* behavior is more likely to be optimal as T increases and the human learns more efficiently, while *opaque* behavior is suitable during short interactions where the human is less sensitive to the robot’s behavior. In Section 6 we will move beyond these simulated, controlled environments and test whether these optimization procedures and design principles apply with real-world users.

6 User Studies

In this section we test how humans collaborate with opaque or transparent robot partners. Our analysis from Section 4 and simulations from Section 5 suggest that there are situations where opaque robot behaviors lead to better team performance. In the following experiments we use Equations (4)–(6) to solve the stochastic two-player Bayesian game and find optimal robot policies. We consider single-shot driving environments and the tower-building task from Figure 1. Here the short interaction time T results in optimal opaque behaviors — i.e., when solving the system of equations the robot finds that opaque behaviors will lead to higher rewards. But it is not clear whether *actual humans* will follow this game-theoretic model, and if the behaviors that the robot predicts will result in higher rewards actually improve the team’s performance. In addition to the shared reward, we also monitor the human’s subjective response. Even if the human does receive higher rewards with an opaque robot, they may prefer working with transparent partners that take actions to convey their latent state. To test both the team’s performance and the human’s perception we performed two separate user studies: an online user study with autonomous driving (Section 6.1), and an in-person experiment with collaborative block stacking (Section 6.1). Below

we first introduce our independent and dependent variables shared across both user studies, and then describe the protocol, results, and discussion for each specific experiment.

Independent Variables. During each trial one human interacted with one robot. There were two types of robots ($N = 2$) that the human could interact with: these robots were not visually different, and the human had to infer the robot’s type based on its actions. For ease of reference we will refer to these types as the *confused* and *capable* robots. During each interaction the robot’s actual type was randomized: in half of the trials participants worked with the *confused* robot, and the other interactions were with the *capable* robot. As such, participants were unsure about the robot’s current type θ at the start of each interaction.

We varied the robot’s algorithm to compare **Opaque** and **Transparent** robot policies. To find the **Opaque** policy we applied Equations (4)–(6) from Section 4.2 and solved for the robot’s optimal behavior $\pi_{\mathcal{R}}$. We then confirmed that this optimal policy $\pi_{\mathcal{R}}$ was fully opaque at the environments’ start states according to the definition from Section 4.1. To find **Transparent** robot behavior we leveraged a baseline from related works [11, 10]. The **Transparent** algorithm modified the robot’s reward to incentivize revealing actions; more formally, in [11, 10] the robot assigns a bonus reward to actions that convey θ to the user. We incorporated this bonus reward into our stochastic Bayesian game, and then solved for the optimal (and transparent) robot behaviors. We confirmed that the **Transparent** robot took actions to communicate its intent and was not either fully or rationally opaque.

Dependent Measures. During each interaction we measured the human-robot team’s final *Reward*. Here reward corresponded to the score in each environment: e.g., the distance the shared car traveled in the online study, or the height of the block tower in the in-person study. We normalized these rewards between 0 and 1 for consistent reporting. We also displayed the human’s reward in real-time so that participants could track their own performance.

Next, we measured the human’s subjective *Preference* for each robot algorithm. Preference was measured on a 1-7 Likert scale, where higher

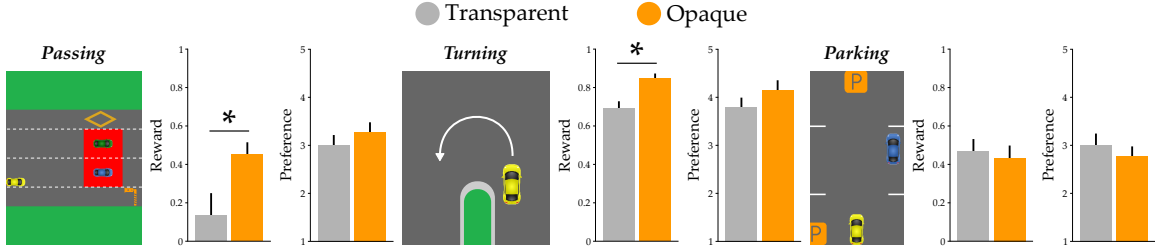


Fig. 5 Task results from our online user study. Participants collaborated with a virtual agent to drive a car in *Passing*, *Turning*, and *Parking* environments. Error bars show standard error and an * denotes statistical significance ($p < .05$)

scores indicated the human had a stronger preference for working with that robot algorithm.

Hypotheses. Throughout our user studies we had two main hypotheses:

H1. *Humans will obtain more reward with **Opaque** partners than with **Transparent** partners.*

H2. *Humans will prefer **Opaque** robots, even though these robots withhold information.*

6.1 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 clicked their input $a_{\mathcal{H}}$, the robot selected its action $a_{\mathcal{R}}$, and then the autonomous car moved using the combined action $a = a_{\mathcal{H}} + a_{\mathcal{R}}$.

Experimental Setup. Participants teamed up with the robot to drive in three settings: passing, turning, and parking (see Figure 5). In *Passing* the team received 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. Each reward was normalized into a 0-1 range.

Participants completed every scenario with **Opaque** and **Transparent** robots. The robot’s type θ was randomized and counterbalanced — in half of the interactions the robot was type confused, and in the other half the robot was type capable. Accordingly, there were four equally balanced experimental conditions: **Opaque** policies with confused and capable robots, and **Transparent** policies with confused and capable robots.

Interactions lasted three timesteps ($T = 3$). During each timestep we first showed the participant an image of the current state s^t and prompted the user to select their action from a multiple choice menu (e.g., turning left, accelerating forward). After the user selected their action $a_{\mathcal{H}}^t$, the robot acted simultaneously with $a_{\mathcal{R}}^t$ and we showed an image of the next state s^{t+1} . We displayed the user’s reward throughout the game, and then at the end asked if the user “preferred sharing control with this car.”

Participants. 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 qualifying 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 randomized and counterbalanced so that half of the users started each scenario with the **Opaque** robot and the other half started with **Transparent**. Participants received a \$5 USD gift card after completing the survey.

Results. Our results are summarized in Figures 5 and 6. Taken across all tasks and both robot types, paired t-tests reveal that the team’s overall reward was significantly higher when humans collaborated with **Opaque** partners than when humans collaborated with **Transparent** partners ($t(179) = 2.62, p < .05$). Looking more specifically at the individual driving 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.

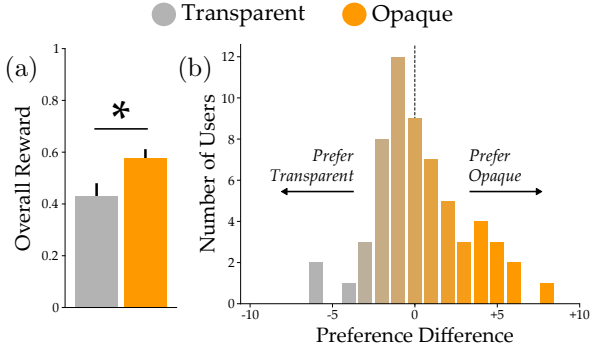


Fig. 6 Overall results from our online user study. (a) The average reward across all three tasks. (b) 30 participants completed each task twice with **Opaque** and twice with **Transparent**, resulting in 60 pairs of datapoints. For each pair we subtracted the total preference scores with **Transparent** 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 (the Preference Difference was near zero). Error bars show standard error and an * denotes statistical significance ($p < .05$)

On the other hand, the users’ perceptions of the robots were divided. Across all three tasks, paired t-tests showed that the differences in preference scores were not statistically significant ($t(179) = 1.06, p = .29$). We initially hypothesized 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, and the overall distribution is unimodal. Hence, in response to hypothesis **H2** we find that online users did not clearly prefer either **Opaque** or **Transparent** robots.

Discussion. When sharing control of the autonomous car, users needed to quickly coordinate with their robot partner to successfully complete the task (i.e., to pass, turn, or park). If the human and robot attempted to complete the task in different ways, their conflicting inputs resulted in lower rewards. Hence, our findings here are aligned with our simulation results from Section 5.1, and suggest that transparent actions are not desirable when the human and robot must quickly collaborate.

One limitation of this online study is that the participants may not have been invested in

the outcomes. There were no consequences if the human-robot failed the task outside of a lower user score. For example, if humans were driving actual vehicles, we anticipate that their behavior may have been more conservative. To help address this limitation we next performed an in-person user study where participants worked directly with a physical robot arm.

6.2 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). This experiment was motivated by our larger application of industrial settings where human workers must collaborate with robot partners. At each timestep the human picked up and added one block to the tower, a_H , and then the robot stacked its block on top, a_R . The state s was the sequence of blocks in the tower. See videos of this user study here: <https://youtu.be/u8q1Z7WHUuI>

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 +0.5 if they picked the same block, and -0.5 if they picked different blocks. Each taller block that was stacked on the tower added a reward of +0.1. In Figure 7 we normalize the interaction reward between 0 and 1 for ease of comparison with online user study results. 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. To familiarize participants with the robot and alleviate any hesitation in interacting with it, we asked the participants to engage in an unrecorded practice run which did not include any algorithms from the study.

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.

Participants and Procedure. We recruited 13 participants from the Virginia Tech community (5

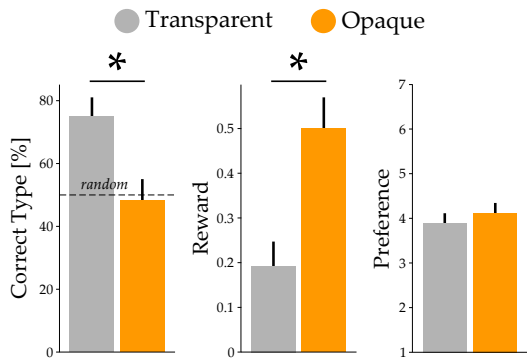


Fig. 7 Results from our in-person user study. Participants collaborated with a robot arm to stack blocks (see Figure 1). *Correct Type* is the percentage of trials where users correctly identified that the robot was capable or confused: by guessing randomly users would be right 50% of the time. As expected, the **Transparent** algorithm more effectively conveys the robot’s type θ to the participants. However, when performing the collaborative assembly task with a physical robot arm, users achieved higher rewards with the **Opaque** algorithm. Users did not indicate a clear preference for either **Opaque** or **Transparent** partners. Error bars show standard error and an * denotes statistical significance ($p < .05$)

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). We informed the participation about the performance based compensation before the study began. This compensation was designed to motivate participants to try and maximize their reward while working with the robot (i.e., to assemble the best tower possible).

After completing each tower we asked users if they could determine which type of robot they had worked with (i.e., if they could determine whether θ was 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. We emphasize that users were never told which algorithm they were working with, or what the current θ was. The algorithms and robot types were presented in a randomized and counterbalanced order. Participants were never told which robots were confused and which robots were capable.

Results. See our [video](#) and Figure 7. To confirm that 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 inferred whether the robot was confused or capable. For reference, users that guessed the robot’s type θ completely at random would be right 50% of the time. Our results in Figure 7 show that with **Opaque** users were not able to determine robot’s type: their responses were on par with this random baseline (48% correct). As expected, with the **Transparent** partner users identified θ correctly 75% of the time. This supports our experimental design and suggests that the **Opaque** robot did indeed withhold the robot’s type θ , while the **Transparent** robot communicated this type to the participants.

We next proceeded to evaluate hypotheses **H1** and **H2**. Our results from Figure 7 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 ($t(51) = .48, p = .63$). In practice, when working with the **Opaque** the robot would select the same block regardless of the robot’s type θ . This consistency may have helped participants predict how the robot would behave and enabled users to coordinate with the robot. By contrast, with **Transparent** the robot selected different blocks to add to the tower for each type θ . By changing the block the robot communicated its type to the human. But this change may have also made it more difficult to anticipate what the robot would do, resulting in an increased number of interactions where the human was unsure how best to coordinate with the robot partner.

Discussion. Our results across 13 in-person participants suggest that robot arms in assembly settings can achieve higher rewards when withholding latent information. To support this finding, we checked whether *learning effects* may have skewed our results. Because participants worked with the robot a total of 8 times, it is possible that users picked up on patterns over the course of the experiment and then anticipated how the robot would behave when assembling their final towers. Using paired t-tests we compared the final reward from the first four interactions and the last four interactions. We found that the differences

were not statistically significant with the **Opaque** robot ($t = 0.622$, $p = 0.54$) or the **Transparent** robot ($t = -1.33$, $p = 0.20$). Hence, we conclude that there was not a significant learning effect during the experiment, and the towers that users assembled were not affected by repeated interactions.

For hypothesis **H2**, we again found that the users did not have a clear preference for either **Transparent** or **Opaque** robots. We suggest that two factors may be at play here: (a) participants may prefer the **Opaque** robot because it leads to higher task reward, while simultaneously (b) participants may prefer the **Transparent** robot because it conveys its capabilities to the human. Perhaps this trade-off between performance and communication cancelled out, and users were left with similar preferences for both algorithms.

7 Conclusion

We considered settings where a robot is collaborating with a human, and both the human and robot share the same objective. Although prior work suggests that it is often optimal for collaborative robots to be transparent — and take actions that convey their latent state — we have introduced theoretical and experimental analysis to demonstrate that there are situations where transparency is not optimal. We built on related works to formalize opaque robot behavior, and then developed a modified version of Harsanyi-Bellman *ad hoc* coordination to identify optimal robot policies in stochastic Bayesian games. We proved that this optimal policy can be either fully or rationally opaque. Our simulations suggest that opaque robot behavior is more likely to be optimal in settings where the human and robot collaborate across a short time horizon, or where the human learns gradually from the robot’s actions. We find experimental support for our analysis across two user studies with 43 total users. Within these experiments participants reached higher rewards when collaborating with opaque robots, and users did not perceive opaque robots as subjectively worse than their transparent counterparts.

7.1 Limitations and Future Work

One limitation of our experiments is that we focused on settings where there were two possible

latent states (i.e., $N = 2$). We designed our experiments with this binary choice in mind so that we could clearly explain the types to our participants: psychological research suggests that too much variability here can lead to random user choices [12]. Similarly, prior work on using transparent robot motion often communicates the robot’s goal in settings where there are only two different options [11, 10, 7]. By restricting our user studies to $N = 2$, we were able to mitigate user confusion and ensure that the experiments accurately tested the transparent and opaque algorithms. Moving forward we are interested in conducting experiments with an increasing number of latent states. We anticipate that — during these experiments — we may need additional measures of the human’s cognitive workload and attention to understand whether opaque robot’s simplify the human’s decision making.

Acknowledgments. Not applicable.

Declarations

- Funding
Not applicable.
- Conflict of interest/Competing interests
The authors have no competing interests to declare.
- Ethics approval
Not applicable.
- Consent to participate
All participants were adults who were informed that they could withdraw from the experiment at any time. The participants provided informed consent under Virginia Tech IRB#20 – 755.
- Consent for publication
Not applicable.
- Availability of data and materials
Supplementary video is available at the following link: <https://youtu.be/u8q1Z7WHUuI>
Simulation data is available in the following repository: <https://github.com/VT-Collab/opaque-example>
- Code availability
Experimental and analysis codes are available in the following repository: <https://github.com/VT-Collab/opaque-example>
- Authors’ contributions
Not applicable.

References

- [1] Stefano V Albrecht, Jacob W Crandall, and Subramanian Ramamoorthy. Belief and truth in hypothesised behaviours. *Artificial Intelligence*, pages 63–94, 2016.
- [2] Stefano V Albrecht and Subramanian 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*, pages 1155–1156, 2013.
- [3] Robert J Aumann and Sylvain Sorin. Cooperation and bounded recall. *Games and Economic Behavior*, pages 5–39, 1989.
- [4] Chris Baker, Rebecca Saxe, and Joshua Tenenbaum. Bayesian theory of mind: Modeling joint belief-desire attribution. In *Cognitive Science Society*, 2011.
- [5] Shray Bansal, Jin Xu, Ayanna Howard, and Charles Isbell. Bayes–Nash: Bayesian inference for nash equilibrium selection in human-robot parallel play. *Autonomous Robots*, 46(1):217–230, 2022.
- [6] Erdem Biyik, Anusha Lalitha, Rajarshi Saha, Andrea Goldsmith, and Dorsa Sadigh. Partner-aware algorithms in decentralized cooperative bandit teams. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(9):9296–9303, Jun. 2022.
- [7] Christopher Bodden, Daniel Rakita, Bilge Mutlu, and Michael Gleicher. A flexible optimization-based method for synthesizing intent-expressive robot arm motion. *The International Journal of Robotics Research*, 37(11):1376–1394, 2018.
- [8] Tathagata Chakraborti, Anagha Kulkarni, Sarath Sreedharan, David E Smith, and Subbarao Kambhampati. Explicability? legibility? predictability? transparency? privacy? security? the emerging landscape of interpretable agent behavior. In *Proceedings of the international conference on automated planning and scheduling*, volume 29, pages 86–96, 2019.
- [9] Anca Dragan, Rachel Holladay, and Siddhartha Srinivasa. Deceptive robot motion: Synthesis, analysis and experiments. *Autonomous Robots*, 2015.
- [10] Anca D Dragan, Shira Bauman, Jodi Forlizzi, and Siddhartha S Srinivasa. Effects of robot motion on human-robot collaboration. In *ACM/IEEE International Conference on Human-Robot Interaction*, pages 51–58, 2015.
- [11] Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. Legibility and predictability of robot motion. In *ACM/IEEE International Conference on Human-Robot Interaction*, pages 301–308, 2013.
- [12] Ido Erev and Greg Barron. On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological review*, 112(4):912, 2005.
- [13] Jakob Foerster, Richard Y Chen, Maruan Al-Shedivat, Shimon Whiteson, Pieter Abbeel, and Igor Mordatch. Learning with opponent-learning awareness. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 122–130, 2018.
- [14] Soheil Habibian and Dylan P Losey. Encouraging human interaction with robot teams: Legible and fair subtask allocations. *IEEE Robotics and Automation Letters*, 7(3):6685–6692, 2022.
- [15] Raul Hakli. Cooperative human-robot planning with team reasoning. *International Journal of Social Robotics*, 9:643–658, 2017.
- [16] Thomas Hellström and Suna Bensch. Understandable robots—what, why, and how. *Journal of Behavioral Robotics*, 9:110–123, 2018.
- [17] Guy Hoffman. Evaluating fluency in human-robot collaboration. *IEEE Transactions on Human-Machine Systems*, 49:209–218, 2019.
- [18] Taemie Kim and Pamela Hinds. Who should i blame? effects of autonomy and transparency on attributions in human-robot interaction. In *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pages 80–85, 2006.
- [19] Minae Kwon, Sandy H Huang, and Anca D Dragan. Expressing robot incapability. In *ACM/IEEE International Conference on Human-Robot Interaction*, pages 87–95, 2018.
- [20] Dylan P Losey and Dorsa Sadigh. Robots that take advantage of human trust. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 7001–7008, 2019.

- [21] Michael Lutter, Shie Mannor, Jan Peters, Dieter Fox, and Animesh Garg. Value iteration in continuous actions, states and time. In *International Conference on Machine Learning*, pages 7224–7234, 2021.
- [22] Aleck M MacNally, Nir Lipovetzky, Miquel Ramirez, and Adrian R Pearce. Action selection for transparent planning. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1327–1335, 2018.
- [23] Debasmita Mukherjee, Kashish Gupta, Li Hsin Chang, and Homayoun Najjaran. A survey of robot learning strategies for human-robot collaboration in industrial settings. *Robotics and Computer-Integrated Manufacturing*, 73:102231, 2022.
- [24] Stefanos Nikolaidis, Anton Kuznetsov, David Hsu, and Siddhartha Srinivasa. Formalizing human-robot mutual adaptation: A bounded memory model. In *ACM/IEEE Int. Conf. on Human-Robot Interaction*, pages 75–82, 2016.
- [25] Stefanos Nikolaidis, Swaprava Nath, Ariel D Procaccia, and Siddhartha Srinivasa. Game-theoretic modeling of human adaptation in human-robot collaboration. In *ACM/IEEE international Conference on Human-Robot Interaction*, pages 323–331, 2017.
- [26] Linda Onnasch and Eileen Roesler. A taxonomy to structure and analyze human-robot interaction. *International Journal of Social Robotics*, 13(4):833–849, 2021.
- [27] Sagar Parekh and Dylan P Losey. Learning latent representations to co-adapt to humans. *Autonomous Robots*, pages 1–26, 2023.
- [28] Lasse Peters, David Fridovich-Keil, Claire J Tomlin, and Zachary N Sunberg. Inference-based strategy alignment for general-sum differential games. In *International Conference on Autonomous Agents and MultiAgent Systems*, pages 1037–1045, 2020.
- [29] Alessandro Roncone, Olivier Mangin, and Brian Scassellati. Transparent role assignment and task allocation in human robot collaboration. In *IEEE Int. Conference on Robotics and Automation*, pages 1014–1021, 2017.
- [30] Stuart J Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson Education, Inc., 2022.
- [31] Dorsa Sadigh, Shankar Sastry, Sanjit A Seshia, and Anca D Dragan. Planning for autonomous cars that leverage effects on human actions. In *Robotics: Science and Systems*, 2016.
- [32] Shahabedin Sagheb, Ye-Ji Mun, Neema Ahmadian, Benjamin A Christie, Andrea Bajcsy, Katherine Driggs-Campbell, and Dylan P Losey. Towards robots that influence humans over long-term interaction. In *IEEE International Conference on Robotics and Automation*, 2022.
- [33] Lindsay Sanneman, Christopher Fourie, Julie A Shah, et al. The state of industrial robotics: Emerging technologies, challenges, and key research directions. *Foundations and Trends in Robotics*, 8(3):225–306, 2021.
- [34] Sarah Sebo, Brett Stoll, Brian Scassellati, and Malte F Jung. Robots in groups and teams: A literature review. *ACM Transactions on Human-Computer Interaction*, 4:1–36, 2020.
- [35] Sarath Sreedharan, Anagha Kulkarni, and Subbarao Kambhampati. Obfuscatory behavior and deceptive communication. In *Explainable Human-AI Interaction: A Planning Perspective*, pages 121–136. 2022.
- [36] Jonathan Vitale, Meg Tonkin, Sarita Herse, Suman Ojha, Jesse Clark, Mary-Anne Williams, Xun Wang, and William Judge. Be more transparent and users will like you: A robot privacy and user experience design experiment. New York, NY, USA, 2018. Association for Computing Machinery.
- [37] Woodrow Z. Wang, Mark Beliaev, Erdem Bıyık, Daniel A. Lazar, Ramtin Pedarsani, and Dorsa Sadigh. Emergent prosociality in multi-agent games through gifting. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 434–442. International Joint Conferences on Artificial Intelligence Organization, 8 2021. Main Track.
- [38] Timon Willi, Alistair Hp Letcher, Johannes Treutlein, and Jakob Foerster. Cola: Consistent learning with opponent-learning awareness. In *International Conference on Machine Learning*, pages 23804–23831, 2022.
- [39] Sarah A Wu, Rose E Wang, James A Evans, Joshua B Tenenbaum, David C Parkes, and

Max Kleiman-Weiner. Too many cooks: Bayesian inference for coordinating multi-agent collaboration. *Topics in Cognitive Science*, 13(2):414–432, 2021.

- [40] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, and Anind K Dey. Maximum entropy inverse reinforcement learning. In *AAAI*, 2008.