Accelerating Interface Adaptation with User-Friendly Priors

Benjamin A. Christie, Heramb Nemlekar, and Dylan P. Losey

***Abstract*— Robots often need to convey information to human users. For example, robots can leverage visual, auditory, and haptic interfaces to display their intent or express their internal state. In some scenarios there are socially agreed upon conventions for what these signals mean: e.g., a red light indicates an autonomous car is slowing down. But as robots develop new capabilities and seek to convey more complex data, the meaning behind their signals is not always mutually understood: one user might think a flashing light indicates the autonomous car is an aggressive driver, while another user might think the same signal means the autonomous car is defensive. In this paper we enable robots to *adapt* their interfaces to the current user so that the human's personalized interpretation is aligned with the robot's meaning. We start with an information theoretic end-to-end approach, which automatically tunes the interface policy to optimize the correlation between human and robot. But to ensure that this learning policy is intuitive — and to accelerate how quickly the interface adapts to the human — we recognize that humans have *priors* over how interfaces should function. For instance, humans expect interface signals to be proportional and convex. Our approach biases the robot's interface towards these priors, resulting in signals that are adapted to the current user while still following social expectations. Our simulations and user study results across 15 participants suggest that these priors improve robot-to-human communication. See videos here: https://youtu.be/Re3OLg57hp8**

I. Introduction

Consider a robot communicating with a human. For example, in Figure 1 a human driver is attempting to pass an autonomous car. The human does not know the autonomous car's driving style — is the autonomous car an aggressive driver (that will merge in front of the human) or a defensive driver (that will stay in its lane to give the human space)? To communicate the robot uses an *interface*. In Figure 1 the autonomous car's interface is an LED light mounted on its roof that signals the robot's driving style. This light can vary from fully on (an orange light) to fully off (a black light). Importantly, there are multiple ways humans might interpret these signals: perhaps for one human an orange light means the robot is aggressive, while for another human the same signal indicates the robot is defensive. To successfully communicate, here the robot must *adapt* its interface signals to align with the current human's interpretation (e.g., switch between using orange or black for aggressive driving).

More generally, our work focuses on learning a mapping from the information the robot wants to convey to the signals the robot uses to convey that information. In some scenarios this mapping is pre-established and mutually understood —

This work is supported in part by NSF Grant #2129201.

The authors are members of the Collaborative Robotics Lab (Collab), Dept. of Mechanical Engineering, Virginia Tech, Blacksburg, VA 24061. e-mail: {benc00, hnemlekar, losey}@vt.edu

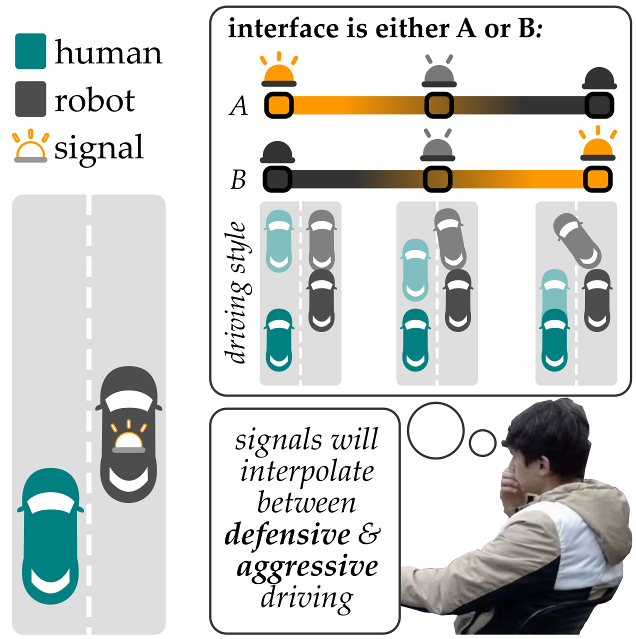

Fig. 1. A human is attempting to pass an autonomous car. The autonomous car *communicates* its policy (i.e., its driving style) using an LED mounted on the top of the car. Different users will interpret these LED signals in different ways. Although the interface does not know *a priori* how any specific human will interpret its signals, we recognize that there are underlying patterns all users expect interfaces to follow. Here the interface signals should recognize that there are two intuitive options that interpolate between (at one extreme) defensive driving and (at the other extreme) aggressive driving.

e.g., a red tail light communicates that a car is slowing down. But we explore settings where humans interact with robotic interfaces (e.g., visual, auditory, haptic), and the meaning behind the interface's signals is not clear. Existing works address this challenge in two ways. On the one hand, some methods [1]–[6] assume the robot has a model of either the human or the task, and fine-tune the interface's signals based on that model. But these methods require domain-specific knowledge and can result in mappings that are only able to convey a single task. On the other hand, approaches like [7] and [8] learn interface signals from scratch by gathering data from the operator across multiple interactions. This removes the need to know the human's task — but learning an interface end-to-end is time-consuming, and the learned mappings can turn out to be complex and unintuitive.

In this paper we bridge the gap between these approaches. We seek to learn a task-invariant interface which can (a) communicate information without needing to know the human's intent while also (b) adapting rapidly and intuitively to the current user. To achieve this goal, our hypothesis is that:

Although different humans may have different *interpretations of the same signal, there are underlying patterns all users expect interfaces to follow.*

Consider our running example from Figure 1. We naturally expect one extreme of the LED light to indicate the robot is aggressive, the other extreme to indicate it is defense, and the intermediate signals should interpolate between aggressive and defensive driving. In what follows, we encode this idea as *priors* over the space of interface mappings. We then derive an information theoretic approach that adapts signals to the current user while ensuring that the interface satisfies the given priors. This leads to interfaces that are *self-adjusting*: these interfaces tune their own signals to help the human correctly interpret the robot's information. Returning to Figure 1, under our approach the robot adapts to use the orange light to convey aggressive driving with one human, and the black light to convey the same data to another user.

Overall, we make the following contributions:

Communication as Optimization. We formulate an optimization problem for identifying the interface signals in task-agnostic settings. Under this formulation the robot learns an interface mapping that (a) maximizes the correlation between the human's behavior and the information the robot is trying to convey, while also (b) conforming to user-friendly priors.

Incorporating User-Friendly Priors. We propose intuitive priors that humans expect interfaces to follow (such as proportionality and convexity). The interface biases its signals towards these priors through a real-time learning approach that personalizes to the current human's behavior.

Comparisons in Simulation and User Studies. We compare our resulting framework against state-of-the-art baselines and ablations of our own approach. Across a 15 person user study we find that participants prefer working with interfaces that leverage our approach, and that these interfaces learn to communicate the robot's information in ways that objectively improve the human's task performance.

II. Related Work

Prior research explores how robots can use interfaces to communicate information to human partners. In settings such as robotic surgery [9], [10] or collaborative assembly [11], [12], these interfaces are typically pre-defined and held constant across all users. Here *the human must adapt to the interface* (i.e., the human must learn what meaning the robot ascribes to each signal). Often these pre-defined interfaces are intuitive — and require little human adaptation — because they build upon socially accepted conventions. For instance, in [13]–[16] the robot communicates data to a human collaborator using signals that are unambiguous (e.g., an arrow pointing to the robot's goal).

However, when humans interact with learning systems or interfaces that integrate novel types of auditory, visual, and haptic feedback, it is not always obvious what the robot is attempting to convey [17]–[20]. Within these contexts the *interface should also adapt to the human*, and change the meaning of its signals to match the human's interpretations. In what follows we survey two fundamental approaches for adapting and personalizing the robot's interface.

Model-Based Interface Adaptation. One group of related works automatically tunes the interface's signals based on either (a) models of the human user or (b) the human's task performance. For example, in machine teaching a robot attempts to convey its latent state to a human: the robot has a model of how the human interprets its actions, and the robot reasons over this model to select communicative actions [3]–[5]. When the human model changes — e.g., the robot recognizes that the human is confused by complex signals — the robot autonomously adapts its signals to align with the new model [6]. In practice, however, the effectiveness of these machine teaching approaches is limited by the accuracy of their human models. If a user deviates from the robot's expectations, the robot may unintentionally send signals that convey the incorrect information. To address this challenge recent methods treat the human as a black box, and optimize the interface signals based on task performance [1], [2]. Here the robot trades-off between exploration and exploitation: the robot applies a set of signals, measures the human's achieved reward, and then decides whether to use those signals again or try different signal mappings. But not only is this process potentially inefficient and time consuming, it can also lead to interfaces that over-fit to a given task. Consider our driving example: if the interface tunes its mapping to help the human pass an autonomous car, that same interface may fail when the human is trying to merge behind the autonomous car.

End-to-End Interface Adaptation. A second body of research — including our prior work [8] — takes the opposite approach. Instead of relying on a human model or given task, these methods use end-to-end structures to learn the interface mapping from scratch. The advantage of these approaches is that the learned signals are task-invariant and highly customized to the current user. But the downside is that end-to-end learning is data intensive: the robot requires many interactions with the current user to adapt its signals, and the resulting mappings may not be intuitive or user-friendly. In this paper we seek to address this challenge by combining end-to-end learning with priors over the space of interface mappings. Existing research in cognitive science suggests that humans have common biases in their communication [21], [22], and recent robotics research has applied similar priors to different problem settings [7]. Our hypothesis is that incorporating these priors will accelerate the robot's adaptation, leading to more effective communication interfaces than purely model-based or end-to-end approaches.

III. Problem Statement

We consider scenarios in which a robotic interface is trying to communicate hidden information to a human operator. For instance, in our running example an autonomous vehicle seeks to convey its driving style to a nearby human-driven car. The interface has access to the hidden information (e.g., the robot knows whether it is an aggressive or defensive driver), and the human needs this hidden information to perform their task correctly (e.g., the human must determine the robot's driving style to safely pass). Within this context, we specifically focus on settings where there are not obvious or mutually agreed upon signals for the robot to convey its hidden information. Our fundamental challenge is determining how — in the absence of established conventions — a robotic interface should adapt its signals over time to convey information to humans. Below we formalize this problem in terms of the interface and human policies.

Repeated Interaction. We assume that the human works with the robotic interface multiple times. At the start of each interaction the robot observes the hidden information $\theta \in \Theta$ (e.g., a vector of parameters capturing the autonomous car's current driving style). This hidden information can change between interactions, but we assume that θ remains constant within an interaction. Every interaction lasts for a total of T timesteps. Over the course of this interaction the interface displays signals to the human, and the human uses these signals to infer θ and complete their task (e.g., safely driving around the autonomous car).

Interface. Let s^t be the system state at timestep t. Within our running example the system state is the position of both the autonomous car and the human-driven car. The robotic interface observes state s and hidden information θ, and then chooses signals $x \in \mathcal{X}$ according to its policy:

$$x \sim \pi_{\mathcal{R}} \left(\circ \mid s, \theta \right) \tag{1}$$

where x^t is the signal that the interface outputs at timestep t. This formulation is not tied to a specific type of signal: here x could be a visual display, an auditory pattern, or a haptic rendering. We emphasize that the robotic interface is collaborating with the human, and seeks to render signals that are intuitive and meaningful for the human.

Human. The human operator cannot directly observe the hidden information θ. Instead, the human sees the interface signals x and the system state s, and then interprets these signals to determine what actions they should take (e.g., the human driver accelerating or changing lanes). Let a^t be the human's action at timestep t, where this action is sampled from the human's policy:

$$a \sim \pi_{\mathcal{H}} \left(\circ \mid s, x \right) \tag{2}$$

The human's actions and the robot's actions cause the system state to transition. The state dynamics are:

$$s^{t+1} = f(s^t, a^t, u^t) \tag{3}$$

where u is the robot's action (e.g., the autonomous car accelerating or changing lanes). In some of our experiments the robot does not take actions and $u = 0$.

Reward. The human has in mind a task that they want to complete. This task directly or indirectly depends on the hidden information θ; returning to our driving example, the human driver wants to safely pass the autonomous car, and θ captures how the autonomous car will drive and change lanes. We therefore formulate the human's objective as a reward function $R(s^t, \theta)$. From the human's perspective, inferring θ is important so that the human can maximize their reward and successfully complete their task. But — from the robot's perspective — we do not assume access to this reward function. Put another way, the robot knows that it should communicate θ to the human, but the robot has no knowledge about how the human will use this information to complete the task (or even what task the human is attempting to complete). Accordingly, over repeated interactions we seek to learn a *task-agnostic* interface policy $\pi_{\mathcal{R}}$ that efficiently communicates θ to the human. The robot does not initially know how the human will interpret its signals (i.e., the robot does not know the human's policy $\pi_{\mathcal{H}}$), and so the robot must *adapt its signals* so that those signals are interpretable and meaningful for the current user.

IV. Using User-Friendly Priors to Accelerate Signal Adaptation

The interface is not sure how the human will interpret its signals. Here we apply our hypothesis that — even though different humans interpret the same signals in different ways — there exist underlying patterns that all users expect signals to follow. Returning to our example where the autonomous car displays its driving style with an LED light: we do not know if the human will interpret an orange light as an aggressive or defensive driver. But we can expect the human to interpret these signals in a proportional and convex manner: i.e., one extreme of the LED light indicates the robot is aggressive, and the other extreme indicates it is defensive, and the intermediate signals linearly interpolate between aggressive and defensive driving. We encode this expectation as a *prior* over the space of interface policies $\pi_{\mathcal{R}}$. By itself, this prior is still not sufficient for the interface to identify the correct $\pi_{\mathcal{R}}$ for the current user. However, we will leverage this prior to narrow down the range of mappings the interface explores and ultimately accelerate the co-adaptation between the learning interface and human operator. In Section IV-A we review a task-agnostic method to learn interfaces over repeated interaction. Then in Section IV-B we theoretically derive how priors can be incorporated to accelerate human-robot co-adaptation. Finally, in Section IV-C we formulate intuitive priors that humans might expect their interfaces to follow (e.g., proportional and convex signals).

A. Correlating Hidden Information and Human Actions

We seek a task-agnostic interface policy $\pi_{\mathcal{R}}$ that maps hidden information θ to signals x in a way that the human can interpret. More specifically, the human's actions a should be *correlated* with the hidden information θ that the interface is trying to convey. Building on our preliminary work [8], we formalize this objective as maximizing the conditional mutual information between θ and a:

$$I(a\,;\,\theta \mid s) = H(a \mid s) - H(a \mid \theta, s) \tag{4}$$

Within this definition $H(Y \mid Z)$ denotes the conditional entropy of variable Y given Z. In practice, if the conditional mutual information $I(a\,;\,\theta \mid s)$ increases it means that — at state s — observing θ tells us what action a the human will take (i.e., variables a and θ are correlated). Consider our driving example. When $I(a\,;\,\theta \mid s)$ is maximized the human driver chooses different actions a for different types of autonomous cars θ: e.g., if the autonomous car is aggressive the human drives more slowly, and if the autonomous car is defensive the human can safely pass. It is still up to the human to decide what actions to take to complete their task successfully. But optimizing for Equation (4) ensures that the interface provides an interpretable connection between the hidden information and the human operator.

In practice it is intractable to evaluate Equation (4) in continuous $\mathcal{S}$, $\mathcal{A}$, $\mathcal{X}$, and Θ spaces [23]. Our preliminary work [8] therefore introduces three neural network models to approximately optimize $I(a\,;\,\theta \mid s)$. The first two models match the interface and human policies from Equation (1) and Equation (2). Let $\mathcal{R}_\psi : \mathcal{S} \times \Theta \rightarrow \mathcal{X}$ be the interface policy (i.e., an instantiation of $\pi_\mathcal{R}$) with weights ψ, and let $\mathcal{H}_\phi : \mathcal{S} \times \mathcal{X} \rightarrow \mathcal{A}$ be the interface's model of the human policy $\pi_\mathcal{H}$ with weights ϕ. We emphasize that the interface does not have access to the human's actual policy. Instead — as we will show in Section IV-B — the interface learns the weights ϕ so that $\mathcal{H}_\phi$ matches the observed human behaviors. Our final model is a decoder that inputs a sequence of k system states and human actions, and then attempts to recover the hidden information θ. We define this decoder as $\Delta_\sigma : \mathcal{S}^k \times \mathcal{A}^k \rightarrow \Theta$ with weights σ. In what follows we will identify the rules for training these three models and learning the weights ψ, ϕ, and σ. For clarity, we note that Section IV-A has summarized our prior work [8], but the theoretical and experimental findings from the remaining sections are novel contributions of this paper.

B. Learning Interface Policies that Incorporate Priors

Maximizing correlation offers a first-pass solution for learning an interpretable interface policy. However, this existing approach does not take advantage of our hypothesis that there are user-friendly priors operators expect their interfaces to follow. For example, if we optimize for Equation (4) alone the autonomous car could learn to use an orange LED light to convey aggressive driving, a partially illumined LED light for defensive driving, and then turn the light off for balanced aggressive-defensive driving. Although users might adapt to this interface given enough time, it is not intuitive: it would better align with our expectations for the extremes of the signal to capture the extremes of the robot's behavior.

We here encode these types of expectations as a prior $P_0(x \mid s, \theta)$. This prior captures how the human might expect the robotic interface to map hidden information to signals x. From an information theoretic perspective, the correlation between a and θ with the prior P_0 is *greater than or equal to* the correlation between a and θ without this prior:

$$I(a\,;\,\theta \mid s, P_0) \geq I(a\,;\,\theta \mid s) \tag{5}$$

Equation (5) follows directly from the properties of mutual information given that both the robotic interface and human operator observe the prior P_0 [24]. Comparing Equation (5) to Equation (4), we see that incorporating priors theoretically increases the correlation between human actions and hidden information. To now reformulate $I(a\,;\,\theta \mid s, P_0)$ as an optimization problem over the space of interface policies, we introduce the objective:

$$\arg\max_{\pi_\mathcal{R}} \Big[I(a\,;\,\theta \mid s) - D_{KL}(\pi_\mathcal{R} \| P_0) \Big] \tag{6}$$

where $D_{KL}(\pi_\mathcal{R} \| P_0)$ is the Kullback-Leibler divergence between the interface distribution $\pi_\mathcal{R}$ and the prior distribution P_0. Maximizing the right side of Equation (6) corresponds to increasing correlation between the human's actions and the hidden information while minimizing the divergence of the interface policy $\pi_\mathcal{R}$ from the human's expectations P_0. Rewriting this optimization problem in terms of the interface and human policies, we reach:

$$\arg\max_{\pi_\mathcal{R}} \int_{x,s,\theta} \bigg(\int_a \Big(p(a,s,x,\theta) \log T \Big) + \pi_\mathcal{R} \log P_0 \bigg) \tag{7}$$

In the above we abbreviate $\pi_\mathcal{R}(x \mid s, \theta)$ as $\pi_\mathcal{R}$ and $\pi_\mathcal{H}(a \mid s, x)$ as $\pi_\mathcal{H}$. Term T is defined as:

$$T = \int_{x'} \pi_\mathcal{H} \pi_\mathcal{R} \bigg(\int_{x'} \pi_\mathcal{H} \pi_\mathcal{R} \int_{\theta'} \pi_\mathcal{R} p(\theta') \bigg)^{-1} \tag{8}$$

Intuitively, Equations (7) and (8) capture our objectives of (a) maximizing correlation between human and interface while (b) biasing the interface towards the prior, and then express these objectives in terms of the interface and human policies.

To tractably solve Equations (7) and (8) we return to our neural networks $\mathcal{R}_\psi$, $\mathcal{H}_\phi$, and Δ_σ from Section IV-A. Let $\mathcal{D} = \{(s^1, a^1, x^1, \theta^1), \ldots, (s^n, a^n, x^n, \theta^n)\}$ be a dataset of n past tuples. The interface continually collects these tuples over repeated interactions as it collaborates with the current user. We will leverage dataset $\mathcal{D}$ to adjust our models in a way that approximately maximizes $I(a\,;\,\theta \mid s, P_0)$. More specifically, we train $\mathcal{R}_\psi$, $\mathcal{H}_\phi$, and Δ_σ to minimize three loss functions that approximate Equations (7) and (8).

Prior Loss. To approximate the $\pi_\mathcal{R} \log P_0$ term that appears Equation (7), we introduce the prior loss $\mathcal{L}_P$:

$$\mathcal{L}_P(\psi) = \sum_\mathcal{D} \|\hat{x} - \mathcal{R}_\psi(s,\theta)\|^2, \quad \hat{x} \sim P_0(\circ \mid s, \theta) \tag{9}$$

Minimizing Equation (9) modifies the deterministic interface policy to match the prior (i.e., the interface outputs signals x that are likely under the human's prior expectations). We note that minimizing the loss $\mathcal{L}_P$ is analogous to maximizing the original term $\pi_\mathcal{R} \log P_0$.

Policy Loss. To approximate the numerator of T in Equation (8) we introduce the policy loss. Intuitively, the numerator $\int \pi_\mathcal{H} \cdot \pi_\mathcal{R}$ is the probability of a human's observed action given s and θ. Hence, the policy loss $\mathcal{L}_\pi$ is:

$$\mathcal{L}_\pi(\phi) = \sum_{(s,a,\theta) \in \mathcal{D}} \Big\| a - \mathcal{H}_\phi\left(s, \mathcal{R}_\psi\left(s, \theta\right)\right) \Big\|^2 \tag{10}$$

For any (s, a, θ) tuple in our dataset, Equation (10) asserts that the human model should map the interface's signal to the actual human behavior a. This drives the human model $\mathcal{H}_\phi$ towards the actual human policy $\pi_\mathcal{H}$. Minimizing $\mathcal{L}_\pi$ is analogous to maximizing $\int \pi_\mathcal{H} \cdot \pi_\mathcal{R}$ and increasing the value of T. We note that we do not use this loss function to update the weights of the interface policy R_ψ — only the human model weights are adjusted.

Decoder Loss. To approximate the denominator of T in Equation (8) we introduce the decoder loss $\mathcal{L}_\Delta$. This method approximates $\int \pi_\mathcal{H} \pi_\mathcal{R} \int \pi_\mathcal{R} p(\theta')$ by generating a collection of counterfactual state-action pairs starting at a (s, θ) pair that is sampled from our dataset $\mathcal{D}$:

$$\begin{gathered} \tau(s,\theta) = \{(s,a)^0, \ldots, (s,a)^k\} \\ a^t = \mathcal{H}_\phi\left(s^t, \mathcal{R}_\psi\left(s^t,\theta\right)\right), \quad s^{t+1} = f\left(s^t, a^t\right) \end{gathered} \tag{11}$$

Within Equation (11) we apply the learned interface and human policies to predict the counterfactual human action, and then transition to the next state using the known system dynamics f. The resulting trajectory τ rolls-out a hypothetical interaction between the interface and human. We then decode this trajectory of states and human actions to attempt to recover the θ the interface was trying to convey:

$$\mathcal{L}_\Delta(\psi,\sigma) = \sum_{(s,\theta)\in\mathcal{D}} \left\|\theta - \Delta_\sigma\left(\tau(s,\theta)\right)\right\|^2 \tag{12}$$

Equation (12) is minimized when the interface signals cause the human to take unique actions for specific hidden information; this is equivalent to minimizing the denominator of T, where a given human action a should be likely for one choice of θ and unlikely for other choices of θ. We note that back-propagating the derivative of Equation (12) through the network is only possible because we have generated a simulated interaction with Equation (11).

Overall Loss. Putting these three loss functions together to approximate Equations (7) and (8), we reach the overall loss function $\mathcal{L}$ used to train our models:

$$\mathcal{L}(\psi,\phi,\sigma) = \lambda_1 \mathcal{L}_P(\psi) + \lambda_2 \mathcal{L}_\pi(\phi) + \lambda_3 \mathcal{L}_\Delta(\psi,\sigma) \tag{13}$$

where λ_1, λ_2, and λ_3 are design parameters in $[0, \infty)$. In practice, the robot updates $\mathcal{R}_\psi$, $\mathcal{H}_\phi$, and Δ_σ online in order to minimize $\mathcal{L}$. The interface then applies the learned policy $\mathcal{R}_\psi$ to choose the signals it displays to the human. Increasing the value of λ_1 relative to λ_2 and λ_3 places more weight on the interface prior, causing the learned interface policy to remain similar to the mutually understood prior P_0.

C. User-Friendly Interface Priors

In Section IV-B we presented an information theoretic approach to learn interface policies that are interpretable for the current user and biased towards a given prior. However, it is still not clear what these priors should be. Put another way, what distribution(s) should the interface use for P_0 in Equation (9)? Below we propose two user-friendly properties that can be leveraged to derive interface priors. Our approach is not tied to these specific priors — the method that we have developed can be applied to any prior distribution provided by the designer. However, our experiments will focus on the two priors described below: proportionality and convexity.

Proportionality. Consider our running example in which an autonomous car is trying to signal its driving style to a nearby human. If the autonomous car slightly modifies its driving style θ (e.g., becomes incrementally more aggressive), the change in signal x should also be small. Conversely, when the autonomous significantly alters θ (e.g., switches from aggressive to defensive driving), then the change in signals x should also be significant. We formalize this intuition by asserting that the change in signals x are *proportional* to the change in hidden information θ:

$$\begin{gathered} x_1 \sim \pi_\mathcal{R}(\circ \mid s, \theta_1), \; x_2 \sim \pi_\mathcal{R}(\circ \mid s, \theta_2) \\ \|x_1 - x_2\|^2 \propto \|\theta_1 - \theta_2\|^2 \end{gathered} \tag{14}$$

To enforce this proportionality, we choose $\mathcal{L}_P(\psi)$ as:

$$\sum_{(s,\theta_1,\theta_2)\in\mathcal{D}} \left\|\mathcal{R}_\psi(s,\theta_1) - \mathcal{R}_\psi(s,\theta_2)\right\|^2 e^{\gamma\|\theta_1-\theta_2\|^2} \tag{15}$$

where γ is a hyperparameter that tunes the sensitivity to changes in θ. This choice of $\mathcal{L}_P$ scales the difference in signals proportionally to the difference in hidden information.

Convexity. Returning to our running example, at one extreme of θ the autonomous car is fully aggressive, and at the other extreme of θ the autonomous car is fully defensive. Our intuition is that these opposite values of θ should correspond to opposite signals x; e.g. if the LED light is fully illuminated to signal an aggressive car, then it is turned off to signal the defensive car. We formalize this user-friendly prior by making the space of signals *convex* in magnitude. Let $g(s,\theta) = \|\mathcal{R}_\psi(s,\theta)\|$ be the magnitude of the signals output by the interface policy at a given state s. To make $g(s,\theta)$ a convex function of θ (i.e., larger θ correspond to larger x) we select the prior loss $\mathcal{L}_P(\psi)$ to be:

$$\mathcal{L}_P(\psi) = \sum_{(s,\theta)\in\mathcal{D}} \left\|\mathcal{R}_\psi(s,\theta) + \mathcal{R}_\psi(s,-\theta)\right\|^2 \tag{16}$$

This choice of $\mathcal{L}_P$ together with Equation (12) ensures that $g(s,\theta)$ forms a convex set where the global minimum occurs at $\theta = 0$. Put another way, the smallest signal occurs at $\theta = 0$ and the signal x increases in magnitude as θ increases in magnitude. Looking at Equation (16), we also notice that opposite values of θ correspond to opposite signals x.

Summary. To summarize our proposed approach, we use the combined loss function in Equation (13) to tractably approximate Equation (6). Optimizing this loss function produces interface policies that (a) maximize the correlation between human actions and θ, and (b) minimize the divergence from human-friendly priors (e.g., proportionality or convexity). In practice, the interface is initialized with a policy consistent with the prior and then starts collaborating with the human. As the interface gathers data from the human operator, it updates the dataset $\mathcal{D}$ and retrains the interface policy online to minimize the loss $\mathcal{L}$ across $\mathcal{D}$. This leads to an interface that adapts its signals over time to establish correlation

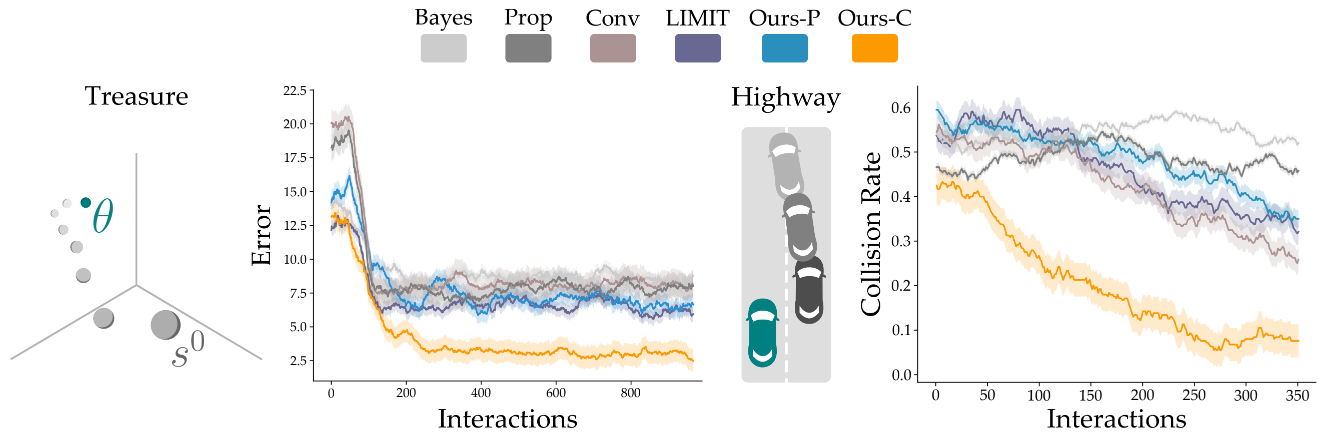

Fig. 2. Results from the **Treasure** (left) and **Highway** (right) simulations. **Treasure**: The human starts in state s^0 and tries to reach a hidden state θ that only the robot knows. The results shown here are averaged across 5 simulations of 1000 interactions in a 3D environment. These results are consistent from 2D to 8D environments. Ours-C outperforms all other baselines in terms of average error ($p < 0.001$). **Highway**: The human (teal) tries to pass the autonomous vehicle (grey) without a collision. Here the results are averaged over 10 simulations of 350 interactions. Ours-C significantly outperforms baselines and Ours-P ($p < 0.001$). We note that the combination of *Proportionality* and *Convexity* are not shown: we found that these two priors conflict with one another during learning. Intuitively, this occurs because *Convexity* attempts to spread signals out, while *Proportionality* groups signals together.

with the current user, while also maintaining the intuitive properties that humans expect interfaces to follow.

V. Simulations

We first compare our proposed approach for adapting interface signals to state-of-the-art baselines [8], [25] and ablations of our method across controlled simulations. We consider two different environments — Treasure and Highway — described later in this section. In each environment, the interface knows the hidden information θ (e.g., location of the treasure) and must generate signals to communicate this hidden information to a simulated human. The simulated humans are adaptive agents that change how they interpret the signals between interactions. Accordingly, the interface must adapt its signals to align with the simulated human in order to accurately convey θ. Code for implementing these simulations can be found here: https://github.com/VT-Collab/interfaces-human-priors.

Interface Algorithms. We compare the following methods for adapting interface signals:

- **Bayes**: The interface generates signals according to $x = \boldsymbol{A}\begin{bmatrix} s & \theta \end{bmatrix}^T$, where $\boldsymbol{A}$ is a matrix found using *Bayesian Optimization* [1], [25]. The interface adjusts $\boldsymbol{A}$ to maximize the total observed reward.
- **Prop**: The interface generates signals *only* according to the *Proportionality* prior from Section IV-C
- **Conv**: The interface generates signals *only* according to the *Convexity* prior from Section IV-C
- **LIMIT**: The interface generates signals to maximize the correlation between human and robot using an existing end-to-end approach [8].
- **Ours-P**: The interface generates signals according to our proposed approach with the *Proportionality* prior.
- **Ours-C**: The interface generates signals according to our proposed approach with the *Convexity* prior.

Note that — unlike the other methods — **Bayes** has access to the human's reward function and uses it to optimize the robot's interface. This gives **Bayes** more domain knowledge than the alternative approaches.

Simulated Human. We pair each algorithm with a simulated agent. These simulated agents are multi-layer perceptrons consistent with Equation (2): $\mathcal{M} : \mathcal{X} \times \mathcal{S} \mapsto \mathcal{A}$. Hence the simulated agents learns an mapping from the interface's signals x and system state s to human actions a. To enforce human-like biases in signal interpretation, we pretrain the human with corresponding interface structures (**Bayes**, *Convexity*, etc.). This process is separate from the testing regime; simulated humans are not tested with the interfaces that they were pretrained with.

Treasure. In this environment the simulated human is navigating a continuous space in $\mathbb{R}^n$, and is attempting to reach the treasure at $s = \theta$ within 10 timesteps. The human state s and treasure location θ are randomly sampled from $U([-10, 10]^n)$ at the start of each interaction. The human cannot observe θ and must rely on the n-dimensional interface signals to infer the hidden location and choose goal-directed actions. We test the interface algorithms in environments ranging from two to eight dimensions. The left side of Figure 2 shows the performance of each algorithm in a 3D environment ($n = 3$). We measure performance as the error between θ and the final human state at the end of an interaction. The results indicate that our proposed approach with *Convexity* prior (**Ours-C**) better communicates the hidden goal location than the alternatives ($p < 0.001$).

Highway. In this environment the simulated human is driving along a two-lane one-way highway. In front of the human is an autonomous vehicle that the human must avoid. The autonomous vehicle's policy is parameterized by θ and chosen from several discrete policies:

(a) The autonomous car will always stay in the right lane
(b) The autonomous car will always stay in the left lane
(c) The autonomous car will merge into the lane of the human at the previous timestep

(d) The autonomous car will merge into the *opposite* lane of the human at the previous timestep

The autonomous car's driving style (e.g., policy parameters θ) is sampled before the start of an interaction. Every interaction lasts for 10 timesteps. At each time step, the autonomous vehicle attempts to signal its policy using a 1D interface with signals $x \in \mathbb{R}$. The human needs to infer the next action of the autonomous vehicle from its signals and then choose an action that avoids collision. The right side of Figure 2 shows the average *collision rate* for each interaction. Once again, we find that **Ours-C** outperforms all other interface algorithms, achieving a lower mean collision rate ($p < 0.001$).

We hypothesize that the performance gap between **Ours-C** and other algorithms is due to the symmetries that *Convexity* imposes on the signal manifold: observing signals for a particular θ reveals what signals for opposite θ may look like. The same is not true for *Proportionality*: although the difference between signals should be proportional to the difference between hidden information, there are not intrinsic symmetries about the signal manifold.

VI. User Study

Our controlled simulations suggest that incorporating a convexity prior over the space of interface mappings accelerates adaptation. To evaluate how Ours-C performs with real users, we next conducted an in-person study with a robotic interface. This study took place using the same environments from Section V. In each task the interface knew the hidden information θ (e.g., the policy of an autonomous vehicle), and selected signals to convey θ to the actual human.

Independent Variables. We compared three algorithms for generating signals: **Bayes**, **LIMIT**, and **Ours-C**. As described in Section V, **Bayes** observes the human's reward at the end of each interaction and updates its mapping from (s, θ) to signals to maximize this reward. **LIMIT** learns signals end-to-end to maximize the correlation between θ and the human actions. Finally, **Ours-C** uses our proposed approach to optimize for correlation while also minimizing deviation from the convexity prior in Section IV-C.

Experimental Setup. We divided the study into two tasks: **Treasure** and **Highway**. These tasks were identical to their simulated counterparts. A collection of monochromatic bars of varying color were used to communicate the interface's signals to the human, similar to the interface shown in Figure 1. This signal was one-dimensional for the Highway task and two-dimensional for the Treasure task.

Participants and Procedure. We recruited 15 participants (2 female, age 25.9 ± 5.2 years) from the Virginia Tech community. All participants provided informed written consent as per the university guidelines (IRB #20-755).

Each participant completed the tasks three times: once for each method tested. Both the order of methods tested and the order of tasks completed were counterbalanced. Participants were never told which algorithm they were working with.

Dependent Measures — Objective. We recorded the states, signals, actions, and hidden information during each interaction. To assess user performance in the **Treasure** task, we calculate the final state error of each interaction $\|s^T - \theta\|^2$. A lower final state error indicates that the user correctly inferred the hidden information (the position of the treasure) from the interface's signals. To assess the performance in the **Highway** task, we determine the collision rate of an interaction: $\sum \left(\mathbb{1}[s^t_{\mathcal{R}} = s^t_{\mathcal{H}}]\right) / T$. A lower collision rate indicates that the human is able to avoid the autonomous vehicle more often by inferring the robot's policy from the interface signals.

Dependent Measures — Subjective. After each task and algorithm, participants completed a 7-point Likert scale survey. This survey measured the user's subjective preferences along three multi-item scales. We asked users:

1) If they felt like their performance *improved* over time,
2) If they thought that the interface's signals were *intuitive*,
3) If the interface's signals were *consistent*.

Altogether, these questions gauged how well users thought they collaborated with the interface.

Hypotheses. We had two hypotheses for the user study:

> **H1.** *Users will perform the task better when receiving signals from an interface that adapts with* **Ours-C** *than with either* **Bayes** *or* **LIMIT**.
>
> **H2.** *Users will subjectively prefer interfaces that adapt using* **Ours-C**.

Results. The objective results of our user study are summarized in Figure 3. To address **H1**, we measured the final state error for **Treasure** and the collision rate for **Highway**. For both metrics, lower values indicate better performance. An effective interface will help the human in navigating to the position of the treasure or minimizing their collision with the autonomous car. One-way ANOVA tests showed that participants had significantly lower error when working with interfaces generated by **Ours-C** in both the Treasure ($F(3, 627) = 9.361, p < 0.001$) and Highway ($F(3, 837) = 2.971, p \to 0.05$) environments. These results indicate that interfaces generated with **Ours-C** provided more helpful signals than the **Bayes** and **LIMIT** baselines.

Regarding **H2**, we present the subjective results from our Likert-scale survey in Figure 4. A one-way ANOVA analysis of the users' responses showed a significant difference in the perceived *Consistency* of the interface signals in the Treasure task ($F(2, 39) = 3.919, p < 0.05$). Although we did not see a significant difference in user ratings for the other items, overall, users stated that they preferred interfaces generated with **Ours-C** in both tasks. Compared to the baselines, the signals generated by **Ours-C** were more consistent with user expectations — generating opposite signals for opposite values of θ by optimizing Equation (16).

Summary. In this study we evaluated the **Bayes**, **LIMIT**, and **Ours-C** algorithms with real users. Our experimental results indicate that interfaces generated using **Ours-C** are more effective in conveying the hidden information, leading

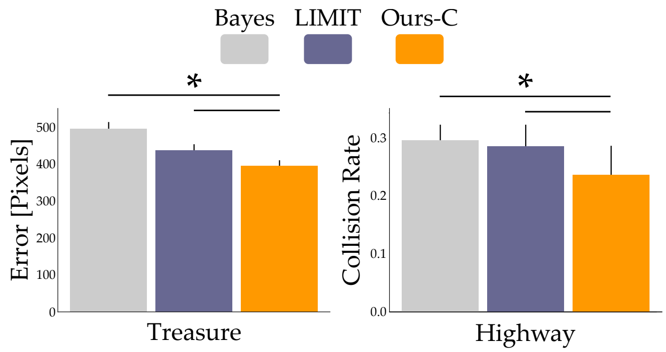

Fig. 3. Ours-C enables users to achieve a better performance in the Treasure (left) and Highway (right) tasks than LIMIT or Bayes. The difference in performance between methods is significant ($p < 0.0001$ (Treasure), $p \rightarrow 0.05$ (Highway)). An asterisk (*) denotes significance.

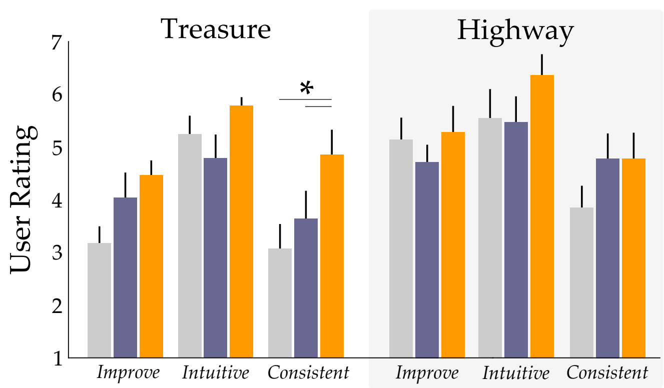

Fig. 4. Participants subjectively preferred Ours-C to the alternatives. Users indicated that interfaces generated using the convexity prior (Ours-C) are more *consistent* and *intuitive*. Furthermore, participants generally felt like they *improved* more over time with Ours-C.

to a greater improvement in user performance than the baselines. Users also subjectively preferred interfaces generated with **Ours-C** stating they are more *intuitive*, *consistent*, and helpful in *improving* their performance over time.

VII. Conclusion

In this paper we studied settings where a robotic interface is attempting to convey information to a human, and there are no pre-existing conventions for how the human should interpret the interface's signals. Although different humans may interpret the same signal in different ways, we hypothesized that all humans have underlying patterns they expect the interface to follow (i.e., priors over the space of signal mappings). We leveraged this insight to propose an information theoretic approach that maximizes the correlation between interface and human while biasing the interface towards user-friendly priors. In practice, this resulted in self-adjusting interfaces that adapted their own signals to better convey information to the human. When compared to model-based and end-to-end baselines across simulations and a user study with 15 participants, we found that incorporating priors accelerated adaptation and improved communication from robot to human.

References

[1] S. Reddy, S. Levine, and A. D. Dragan, "First contact: Unsupervised human-machine co-adaptation via mutual information maximization," in *Advances in Neural Information Processing Systems*, 2022.

[2] T. Kaupp, A. Makarenko, and H. Durrant-Whyte, "Human–robot communication for collaborative decision making — A probabilistic approach," *Robotics and Autonomous Systems*, 2010.

[3] 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.

[4] S. H. Huang, D. Held, P. Abbeel, and A. D. Dragan, "Enabling robots to communicate their objectives," *Autonomous Robots*, 2019.

[5] D. S. Brown and S. Niekum, "Machine teaching for inverse reinforcement learning: Algorithms and applications," in *AAAI*, 2019.

[6] M. S. Lee, H. Admoni, and R. Simmons, "Machine teaching for human inverse reinforcement learning," *Frontiers in Robotics and AI*, 2021.

[7] M. Li, D. P. Losey, J. Bohg, and D. Sadigh, "Learning user-preferred mappings for intuitive robot control," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2020, pp. 10 960–10 967.

[8] B. A. Christie and D. P. Losey, "LIMIT: Learning interfaces to maximize information transfer," *arXiv:2304.08539*, 2023.

[9] A. Simorov, R. S. Otte, C. M. Kopietz, and D. Oleynikov, "Review of surgical robotics user interface: What is the best way to control robotic surgery?" *Surgical Endoscopy*, vol. 26, pp. 2117–2125, 2012.

[10] E. Rozeboom, J. Ruiter, M. Franken, and I. Broeders, "Intuitive user interfaces increase efficiency in endoscope tip control," *Surgical Endoscopy*, vol. 28, pp. 2600–2605, 2014.

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

[12] E. Cha, Y. Kim, T. Fong, and M. J. Mataric, "A survey of nonverbal signaling methods for non-humanoid robots," *Foundations and Trends in Robotics*, vol. 6, no. 4, pp. 211–323, 2018.

[13] T. Weng, L. Perlmutter, S. Nikolaidis, S. Srinivasa, and M. Cakmak, "Robot object referencing through legible situated projections," in *International Conference on Robotics and Automation*, 2019.

[14] R. S. Andersen, O. Madsen, T. B. Moeslund, and H. B. Amor, "Projecting robot intentions into human environments," in *IEEE International Symposium on Robot and Human Interactive Communication*, 2016.

[15] M. Walker, H. Hedayati, J. Lee, and D. Szafir, "Communicating robot motion intent with augmented reality," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2018, pp. 316–324.

[16] J. F. Mullen, J. Mosier, S. Chakrabarti, A. Chen, T. White, and D. P. Losey, "Communicating inferred goals with passive augmented reality and active haptic feedback," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 8522–8529, 2021.

[17] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, and F. Tanaka, "Social robots for education: A review," *Science Robotics*, vol. 3, no. 21, p. eaat5954, 2018.

[18] N. Gasteiger, M. Hellou, and H. S. Ahn, "Factors for personalization and localization to optimize human–robot interaction: A literature review," *International Journal of Social Robotics*, 2021.

[19] N. Dunkelberger, J. L. Sullivan, J. Bradley, I. Manickam, G. Dasarathy, R. Baraniuk, and M. K. O'Malley, "A multisensory approach to present phonemes as language through a wearable haptic device," *IEEE Transactions on Haptics*, vol. 14, no. 1, pp. 188–199, 2020.

[20] A. A. Valdivia, S. Habibian, C. A. Mendenhall, F. Fuentes, R. Shailly, D. P. Losey, and L. H. Blumenschein, "Wrapping haptic displays around robot arms to communicate learning," *IEEE Transactions on Haptics*, vol. 16, no. 1, pp. 57–72, 2023.

[21] V. S. Ramachandran and E. M. Hubbard, "Synaesthesia–A window into perception, thought and language," *Journal of Consciousness Studies*, vol. 8, no. 12, pp. 3–34, 2001.

[22] A. Ćwiek, S. Fuchs, C. Draxler, E. L. Asu, D. Dediu, K. Hiovain, S. Kawahara, S. Koutalidis, M. Krifka, P. Lippus, *et al.*, "The bouba/kiki effect is robust across cultures and writing systems," *Philosophical Transactions of the Royal Society B*, 2022.

[23] J. Song and S. Ermon, "Understanding the limitations of variational mutual information estimators," in *International Conference on Learning Representations*, 2019.

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2012.

[25] F. Nogueira, "Bayesian Optimization: Open source constrained global optimization tool for Python," 2014–. [Online]. Available: https://github.com/fmfn/BayesianOptimization