LIMIT: Learning Interfaces to Maximize Information Transfer

BENJAMIN A. CHRISTIE, Virginia Tech, USA
DYLAN P. LOSEY, Virginia Tech, USA

Robots can use auditory, visual, or haptic interfaces to convey information to human users. The way these interfaces select signals is typically pre-defined by the designer: for instance, a haptic wristband might vibrate when the robot is moving and squeeze when the robot stops. But different people interpret the same signals in different ways, so that what makes sense to one person might be confusing or unintuitive to another. In this paper we introduce a unified algorithmic formalism for learning *co-adaptive* interfaces from *scratch*. Our method does not need to know the human's task (i.e., what the human is using these signals for). Instead, our insight is that interpretable interfaces should select signals that maximize *correlation* between the human's actions and the information the interface is trying to convey. Applying this insight we develop LIMIT: Learning Interfaces to Maximize Information Transfer. LIMIT optimizes a tractable, real-time proxy of information gain in continuous spaces. The first time a person works with our system the signals may appear random; but over repeated interactions the interface learns a one-to-one mapping between displayed signals and human responses. Our resulting approach is both personalized to the current user and not tied to any specific interface modality. We compare LIMIT to state-of-the-art baselines across controlled simulations, an online survey, and an in-person user study with auditory, visual, and haptic interfaces. Overall, our results suggest that LIMIT learns interfaces that enable users to complete the task more quickly and efficiently, and users subjectively prefer LIMIT to the alternatives. See videos here: https://youtu.be/IvQ3TM1_2fA.

CCS Concepts: • **Computing methodologies → Online learning settings**; • **Human-centered computing → User interface design**.

Additional Key Words and Phrases: Interfaces, Information Theory, Co-Adaption, Human-Robot Interaction

1 Introduction

Imagine a person collaborating with a robotic interface to complete some task. The interface displays signals to convey information to the person, and the person interprets those signals to determine what actions to take. For instance, in Figure 1 the human is searching for their missing phone. The interface knows the phone's location and can signal the human with an array of LEDs. But how does the interface determine which signals to use? One person might think that the left LED strip corresponds to the phone's position in the x-axis and the right strip indicates position in the y-axis. But another user might have the opposite mapping — or interpret the interface's feedback in an entirely different way. For each user, the interface must identify a method for selecting signals (e.g., turning on LEDs) that clearly conveys the desired information (e.g., the phone's location).

In this paper we explore settings where a robotic interface is communicating information to a human operator. Here *interfaces* refer to autonomous systems that provide nonverbal feedback in the form of lights, sounds, augmented reality displays, haptic signals, or robot motion. We assume that the interface has access to some task-related, *hidden* information that the human cannot

This work is supported in part by NSF Grant #2129201.
Authors' Contact Information: Benjamin A. Christie, Virginia Tech, Department of Mechanical Engineering, Blacksburg, VA, USA, benc00@vt.edu; Dylan P. Losey, Virginia Tech, Department of Mechanical Engineering, Blacksburg, VA, USA, losey@vt.edu.

ACM 2573-9522/2024/1-ART1
https://doi.org/10.1145/3675758

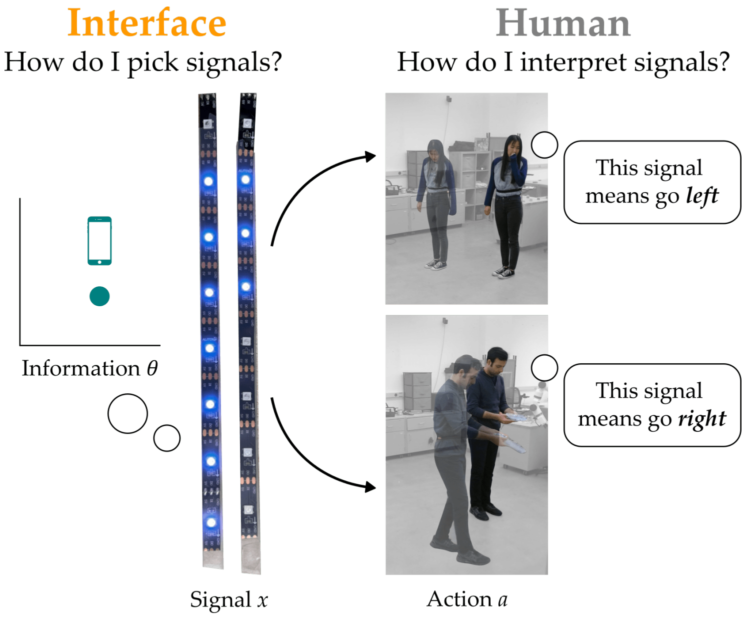
Interface
How do I pick signals?
Human
How do I interpret signals?
Information θ
Signal x
Action a
This signal means go *left*
This signal means go *right*

Fig. 1. Interface selecting signals to convey information to the human operator. Choosing the right feedback is challenging because the way people respond to signals varies across tasks, users, and interface types; e.g., when a person sees this LED pattern should they go left or right? We introduce a unified algorithmic framework that co-adapts to the current user by learning to pick signals that maximize information transfer.

directly observe. Existing research pre-programs these interfaces with a human-engineered and fixed mapping from information to signals [4, 15]. Returning to our motivating example, state-of-the-art methods might tell the robot to always use the left LEDs for x-position and the right LEDs for y-position. But there are two fundamental limitations of this approach. First, using a fixed convention for choosing signals *forces* all humans to learn and follow this specific convention; by contrast, we know that humans have personalized signal preferences and interpretations [3, 13, 37]. Second, these human-engineered mappings must be designed on a *case-by-case* basis, where the designers rely on their intuition and experimental data to decide how the interface will provide feedback for the current task [32, 34, 35].

To overcome these limitations we here introduce a *unified* algorithmic framework for *learning* interfaces from scratch. We do not assume that the interface (a) knows the task the human wants to complete or (b) has a model of how the human will interpret its signals. Instead, our insight is that — in tasks where a robotic interface is sharing *hidden* information with a user:

When interfaces are interpretable *the human's actions are* correlated
with the hidden information that the interface is trying to convey.

Effective feedback signals should guide the user's decisions and inform the human's behaviors. Return to our motivating example where the robotic interface is trying to communicate the position of the human's missing phone. Applied to this example setting, our insight asserts that — if the phone's location changes — the interface should display LED signals that cause the human's actions to also change. For instance, if the phone is on the left side of the room the interface should display different signals (and cause the human to take different actions) as compared to when the phone is on the right side of the room.

We use this insight to develop LIMIT: **L**earning **I**nterfaces to **M**aximize **I**nformation **T**ransfer. LIMIT is an information-theoretic algorithm that learns a real-time interface policy (i.e., a mapping from information to signals) to maximize the mutual dependence between the hidden information and the human actions. LIMIT is not tied to any specific type of interface; as we will show, our unified approach can be applied to visual displays, auditory cues, and haptic arrays. The first time a new user interacts with LIMIT the interface signals may appear random or irregular. But over repeated interactions LIMIT gathers data from the current user, learns online, and personalizes the signals so that humans take different actions for different values of hidden information. Co-adaptive humans exploit these interpretable signals to complete the task and maximize their reward over repeated interactions. Overall, LIMIT is a step towards robots that learn how to convey their own latent, internal information to nearby humans.

In this paper we make the following contributions:

Formalizing Interfaces with Information Transfer. We formulate robotic interfaces as the intersection of human and interface policies. In settings where an interface has access to hidden information that a human does not, we hypothesize that interpretable, task-agnostic interfaces should maximize conditional information gain between hidden information and human actions. We then derive information gain in terms of the agent policies.

Learning Interfaces to Maximize Information Gain. Directly optimizing for information gain is intractable in continuous spaces. We accordingly introduce LIMIT, an online learning approach that closely mirrors the structure of our derived formulation[1]. LIMIT learns an interface policy to correlate the human's actions and hidden information.

Comparing Interfaces in Controlled Simulations. We compare LIMIT to ablations and a state-of-the-art baseline across simulated environments. This includes settings where the interface signal is over-actuated (i.e., the signal has more dimensions than the information) and under-actuated (i.e., the information has more dimensions than the signal).

Testing Interface Interpretability with Online Users. We perform an online user study where 37 participants attempt to find their missing phone using learned interface feedback. Participants more accurately completed the task with LIMIT feedback, and also perceived the LIMIT interface as more helpful, understandable, and intuitive.

Conducting User Studies on Visual, Auditory, and Haptic Interfaces. We put LIMIT to the test with 11 in-person users across three different types of interfaces. Participants guide a robot arm or walk around a room while getting sound, light, and haptic feedback. In each task the interface must co-adapt alongside the human and learn to select meaningful signals. As compared to a state-of-the-art baseline, LIMIT results in better objective performance and subjective ratings.

2 Related Work

We focus on communicating information from robotic interfaces to human operators. The interface knows some information and needs to determine what signals it should display to convey that information to the human. Our goal is to develop a *unified* algorithmic framework that can be applied to different tasks and types of interfaces, and learns to output interpretable signals that are personalized to the current user. Here we discuss related research that leverages fixed, pre-defined interface mappings, as well as interfaces that learn to interpret the human's inputs.

Pre-Defined Interfaces. Prior works explore how interfaces can convey information to humans through nonverbal cues such as lights, projections, augmented reality, haptic signals, and robot motion [4]. Often the interfaces are designed with a specific task in mind, and programmed with a

[1] Our code for implementing LIMIT is available here: https://github.com/VT-Collab/LIMIT-learning-interfaces

pre-defined mapping from information to signals that is held *constant* throughout human-interface interaction [1, 5, 7, 12, 23, 30, 36, 40, 41]. The resulting signals can be intuitive for users to interpret without much experience or explanation [1, 5, 23, 30, 40, 41]. For example, in [5] a mobile robot drove around a crowded room; when the robot projected a line onto the floor indicating its planned trajectory, humans quickly recognized the robot's intent. In other settings the fixed, hand-designed mappings are high-dimensional or intricate, and users may need practice to correctly interpret the robot's meaning [7, 12, 36]. For instance, in [36] humans were trained to convert tactile signals into 500 different words (similar to Morse code). Rather than pre-defining the mapping from information to signals, alternate research assumes the interface has an accurate *model of the human* [11, 16, 20]. More specifically, these works assume that the robot knows how the human will interpret its motions; the robot then inverts this model to select legible behaviors and convey the desired information. This approach works well when the actual user follows the robot's convention — but we know that different humans will interpret and respond to the same feedback in different ways [3, 13, 37]. Unlike these prior works we do not assume that the interface is given either a pre-defined mapping from information to signals or a human model. Instead, we seek to *learn* an interpretable and personalized mapping from scratch.

Learned Interfaces. Most relevant here is recent research that learns mappings from the *human inputs* (i.e., signals) to the *human's intent* (i.e., information) [9, 22, 28, 29]. For instance, in [28] the human is controlling a drone with a keyboard, and the robot learns how to map the human's key presses to drone motions. Similarly, in [22] a human is controlling an assistive robot arm with a joystick, and the robot learns which joystick directions should be associated with each robot motion. These learned mappings go in the *opposite* of what we are interested in: instead of learning how to extract information from human commands, we want to learn how to convey the interface's information to human operators. Put another way, in our work the interface is sending signals to the human. As the interface learns from and adapts to the human operator, the human will inevitably co-adapt to the interface [17, 24, 26, 39]. Building on this prior work, we recognize that humans are not static operators: our approach must be able update and refine signals as the human learns how to interpret the interface.

Maximizing Information Gain. Under our proposed approach the interface learns to display signals that maximize the correlation between the human's actions and the interface's information. More specifically, we will develop an algorithm where the interface learns to maximize a proxy of *information gain*. Recent works have similarly leveraged information gain (i.e., mutual information) to select robot behaviors during human-robot interaction [14, 18, 19, 21, 28, 31]. For example, in [31] an autonomous car nudges closer to the human's car to see how the human will respond (and actively gather information about the human driver). Likewise, in [19] a social robot communicates with the human when the expected benefits of the human's feedback outweigh the cost of probing the human. Although our proposed approach similarly optimizes for information gain, we do so in the opposite direction: the related works [14, 18, 19, 21, 31] select robot actions to *gain* information from the human, while we will choose interface signals to *convey* information to the human.

3 Problem Formulation

We consider settings where a feedback interface is sending signals to a human. Our approach is not tied to any specific type of interface: e.g., the interface could be a haptic wristband, a light projection, or an augmented reality display. The human is attempting to perform some task. We assume that the interface knows *hidden information* θ that the human cannot directly observe, and the human's task depends on this hidden information. More specifically, we assume that the human should take different actions if the hidden information changes. In our running example (see Figure 1) the interface is a wearable array of LED lights. Here the human's task is to find their phone: only the interface knows the phone's location θ, and the human must interpret the feedback signals to reach θ. Our fundamental challenge is finding a signal mapping that is *interpretable* for the current user. We do not assume that the human and interface have a pre-defined convention for signals (e.g., the human and robot do not assume that the left lights indicate horizontal motion and the right lights indicate vertical motion). Instead, we want to enable interfaces to learn to communicate with the current user from scratch.

Human. Let $s \in \mathcal{S}$ be the system state and let $a \in \mathcal{A}$ be the human's action. In our running example the state is the position of the human and the human's action is their change in position. The state transitions based on the human's action:

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

where f is the deterministic dynamics and t is the current timestep. An interaction lasts a total of T timesteps. We use trajectory $\xi = (s^1, s^2, \ldots, s^T)$ to capture the sequence of states the human visits during the current interaction.

Interface. Let $\theta \in \Theta$ be the hidden information and let $P(\theta)$ be a prior over this information. In our running example θ is the phone's location and $P(\theta)$ is a uniform distribution across the room. At the start of the interaction the interface observes $\theta \sim P(\cdot)$, and this parameter remains constant throughout the rest of the interaction. At each timestep t the interface sends signal $x \in \mathcal{X}$ to the human, where $\mathcal{X}$ is the set of all possible signals the interface can output. For our running example x is the intensity of the LED light array.

Policies. The interface observes the system state s and hidden information θ and then outputs signal x. Accordingly, the interface's policy maps (s, θ) to x:

$$\pi_{\mathcal{R}}(x \mid s, \theta) \tag{2}$$

The human sees states and signals; importantly, the human *cannot* directly observe the hidden information θ. We assume the human gets the current signal x^t before taking action a^t, so that the human's policy is a mapping from states and signals to actions:

$$\pi_{\mathcal{H}}(a \mid s, x) \tag{3}$$

Objective. During each interaction the human has in mind some task that they want to complete. This task depends on the hidden information θ; for instance, perhaps the human wants to locate their phone and θ is the phone's position. Let the human have reward function $R(\xi, \theta) \rightarrow \mathbb{R}$, where higher rewards indicate that the human has better accomplished their task. We *do not assume* that the interface has any knowledge of this task or reward function. Instead, the interface only has access to the data it has directly observed: the states, actions, signals, and hidden information from previous interactions. Based only on this data, we seek to learn an interface policy $\pi_{\mathcal{R}}$ that — when paired with the human policy $\pi_{\mathcal{H}}$ — will maximize the human's reward $R(\xi, \theta)$.

4 Learning Interfaces to Maximize Information Transfer (LIMIT)

Given an interface and hidden information θ, our goal is to find an interface policy $\pi_{\mathcal{R}}$ that helps the human complete their task and maximize their reward. This is challenging because (a) the interface does not know the human's task and (b) different humans respond to the same signals in different ways. Within this section we accordingly develop a *task-agnostic*, *personalized* approach for learning interface mappings. This approach is based on our fundamental insight that the human's actions should be correlated with the information that the interface is trying to communicate. In Section 4.1 we capture the mutual dependence between human actions and hidden information using conditional *information gain*. We then rewrite this information gain in terms of the human and interface policies (Section 4.2). Using these equations we introduce a real-time learning approach that trains the interface policy to personalize to the human's current behavior (Section 4.3). Finally, in Section 4.4 we account for the human's co-adaptation to the changing interface.

4.1 Optimizing for Information Gain

Information gain (i.e., mutual information) is a general metric for correlation: it quantifies the amount of information obtained about one variable by observing another variable [8]. Our central hypothesis is that an effective interface should maximize the correlation between human actions and hidden information. Accordingly, we assert that the interface should maximize the *conditional information gain* between action a and hidden information θ given state s:

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

Here $H(a \mid s)$ is the conditional Shannon entropy of a given s, and $H(a \mid s, \theta)$ is the conditional Shannon entropy of a given s and θ. Intuitively, $H(a \mid s)$ captures how uncertain we are about the human's action at state s, while $H(a \mid s, \theta)$ captures our uncertainty given both s and θ.

Equation (4) is maximized when (a) each action is equally likely at state s but (b) we know exactly which action the human will take once hidden information θ is observed. Consider our running example where a human is standing in the middle of the room. The human could walk in any direction; but once the human knows their phone's location θ, the human goes directly towards that goal. More generally, an interface that maximizes Equation (4) will cause the human operator to take actions a that are correlated with the hidden information θ the interface is trying to convey.

4.2 Writing Information Gain in Terms of Policies

We want to learn an interface policy $\pi_{\mathcal{R}}$ that optimizes the conditional information gain $I(a; \theta \mid s)$. Towards this end, we here rewrite Equation (4) in terms of the human policy $\pi_{\mathcal{H}}$ and the interface policy $\pi_{\mathcal{R}}$. For ease of explanation we treat $\mathcal{S}$, $\mathcal{A}$, $\mathcal{X}$, and Θ as discrete sets: the same result extends to continuous spaces by replacing the following summations with integrals over continuous distributions. We will work in continuous spaces during our simulations and user study.

By definition Equation (4) is equal to [8]:

$$I(a\,;\ \theta \mid s) = \sum_{\mathcal{S},\mathcal{A},\Theta} P(s, a, \theta) \log \frac{P(a \mid \theta, s)}{P(a \mid s)} \tag{5}$$

Marginalizing over the interface signal x at each term we find that:

$$I(a\,;\ \theta \mid s) = \sum_{\mathcal{S},\mathcal{A},\Theta} \left(\sum_{\mathcal{X}} P(s, a, x, \theta) \right) \log \frac{\sum_{\mathcal{X}} P(a, x \mid \theta, s)}{\sum_{\mathcal{X}} P(a, x \mid s)} \tag{6}$$

Remember that the human and interface policies are probability distributions: $\pi_{\mathcal{R}}$ is the probability of signal x given s and θ, and $\pi_{\mathcal{H}}$ is the probability of action a given s and x. Because Equation (3) only depends on s and x, we further have that $P(a \mid s, x, \theta) = \pi_{\mathcal{H}}(a \mid s, x)$. Using the chain rule and plugging in Equation (2) and Equation (3), we get that $I(a\,;\ \theta \mid s)$ from Equation (4) is equal to:

$$\sum_{\mathcal{S},\mathcal{A},\Theta} P(s, \theta) \left(\sum_{\mathcal{X}} \pi_{\mathcal{H}}(a \mid s, x) \cdot \pi_{\mathcal{R}}(x \mid s, \theta) \right) \cdot \log \frac{\sum_{\mathcal{X}} \pi_{\mathcal{H}}(a \mid s, x) \cdot \pi_{\mathcal{R}}(x \mid s, \theta)}{\sum_{\mathcal{X}} \pi_{\mathcal{H}}(a \mid s, x) \sum_{\Theta} \pi_{\mathcal{R}}(x \mid s, \theta') P(\theta')} \tag{7}$$

Equation (7) re-expresses conditional mutual information in terms of the human policy $\pi_{\mathcal{H}}$ and interface policy $\pi_{\mathcal{R}}$.

We will gain additional insight by separating Equation (7) into two terms: *convey* and *distinguish*. We refer to the first term as *convey*:

$$\mathcal{T}_{conv} = \sum_{\mathcal{X}} \pi_{\mathcal{H}}(a \mid s, x) \cdot \pi_{\mathcal{R}}(x \mid s, \theta) \tag{8}$$

Note that $\mathcal{T}_{conv}$ appears twice in Equation (7): once outside of the log and again in the numerator of the log. Next, we refer to our second term as *distinguish*:

$$\mathcal{T}_{dist} = \sum_{\mathcal{X}} \pi_{\mathcal{H}}(a \mid s, x) \sum_{\Theta} \pi_{\mathcal{R}}(x \mid s, \theta') P(\theta') \tag{9}$$

For clarity, we show how these *convey* and *distinguish* terms are derived from Equation (7) below:

$$I(a\,;\,\theta \mid s) = \sum_{\mathcal{S},\mathcal{A},\Theta} P(s, \theta) \cdot \mathcal{T}_{conv} \cdot \log \frac{\mathcal{T}_{conv}}{\mathcal{T}_{dist}} \tag{10}$$

$\mathcal{T}_{conv}$ captures how likely it is that the human takes action a given state s and hidden information θ. By contrast, $\mathcal{T}_{dist}$ expresses the likelihood of action a at the current state across any choice of θ. From Equation (10), we see that an interface $\pi_{\mathcal{R}}$ that optimizes for information gain must *maximize* $\mathcal{T}_{conv}$ and *minimize* $\mathcal{T}_{dist}$. Intuitively, we want the human's action a to be likely for a specific choice of θ (increasing $\mathcal{T}_{conv}$), but not likely for every possible θ (decreasing $\mathcal{T}_{dist}$). Return to our motivating example and imagine that the hidden phone is on the left side of the room. For this θ, the interface should select signals x that always cause the human to walk left. However, if θ changes (i.e., the phone is now on the right side) the robot's signals should not cause the human to keep walking left (and take the same action a). Instead, different human actions a should be likely for different choices of interface information θ.

4.3 Learning to Maximize Information Gain

With Equations (8)-(10) we now have a formula for information gain in terms of the human and interface policies. Ideally we would optimize over these equations to find the interface policy $\pi_{\mathcal{R}}$ that maximizes conditional information gain. Unfortunately, this is not possible for two reasons: (a) we do not know the human's current policy $\pi_{\mathcal{H}}$ and (b) it is intractable to evaluate information gain in continuous $\mathcal{S}$, $\mathcal{A}$, $\mathcal{X}$, and Θ spaces [2, 27, 33]. Instead of directly computing the information gain, we here introduce **L**earning **I**nterfaces to **M**aximize **I**nformation **T**ransfer (**LIMIT**). LIMIT is a real-time, personalized *learning* approach that closely mirrors the structure of Equations (8)-(10). As LIMIT gathers data alongside the human operator, it continually learns and updates the interface policy $\pi_{\mathcal{R}}$ to correlate the human's actions and the hidden information. We emphasize that LIMIT does not have access to the human's task or reward function: this task-agnostic approach learns policies to optimize a *proxy* of conditional information gain.

Models. LIMIT consists of three neural networks. We introduce the first two networks here: let $\mathcal{H}_\phi$ be a model of the human's policy with weights ϕ, and let $\mathcal{R}_\psi$ be the interface's learned policy with weights ψ. The structure of these models corresponds to Equation (2) and Equation (3):

$$\mathcal{H}_\phi : \mathcal{S} \times \mathcal{X} \rightarrow \mathcal{A}, \qquad \mathcal{R}_\psi : \mathcal{S} \times \Theta \rightarrow \mathcal{X} \tag{11}$$

so that $\mathcal{H}_\phi$ maps states and signals to actions and $\mathcal{R}_\psi$ maps states and hidden information to signals[2]. It should be noted that $\mathcal{H}_\phi$ is not the human's actual policy (which the robotic interface never knows). However, while the interface does not observe the human's policy $\pi_{\mathcal{H}}$ or even their current task, the interface does have access to data from previous interactions. Let $\mathcal{D}$ =

[2] When using LIMIT, the interface policy $\pi_{\mathcal{R}}$ is the learned model $\mathcal{R}_\psi$.

$\{(s, x, a, \theta^0), \dots (s^N, x^N, a^N, \theta^N)\}$ be the dataset of observed states, signals, actions, and hidden information across all previous interactions.

Below we introduce the two loss functions used to train $\mathcal{H}_\phi$ and $\mathcal{R}_\psi$ on dataset $\mathcal{D}$. These loss functions are analogous to the terms $\mathcal{T}_{conv}$ and $\mathcal{T}_{dist}$ from Equation (10).

Convey. Remember from Section 4.2 that interfaces which optimize information gain will maximize the *convey* term in Equation (8). Intuitively, $\mathcal{T}_{conv}$ expresses the probability of the human's observed action a given s and θ. We here introduce a loss term analogous to Equation (8) that specifically applies to our deterministic human and robot models:

$$\mathcal{L}_{conv}(\phi, \psi) = \sum_{(s,a,\theta)\in\mathcal{D}} \left\| a - \mathcal{H}_\phi\big(s, \mathcal{R}_\psi(s, \theta)\big) \right\|^2 \tag{12}$$

For any (s, a, θ) tuple in our dataset, Equation (12) asserts that the human model and interface policy should map s and θ to an action that closely matches the human's actual behavior a. Put another way, $\mathcal{R}_\psi(s, \theta)$ should output a signal x that causes our human model to take the observed action a. When comparing Equation (12) and Equation (8) we recognize that maximizing $\mathcal{T}_{conv}$ and minimizing $\mathcal{L}_{conv}$ accomplish similar things: both assert that hidden information θ and action a should be correlated at state s, so that if the system observes (s, θ) the human always outputs a.

Distinguish. We next develop a proxy loss function for the *distinguish* term in Equation (9). $\mathcal{T}_{dist}$ implies that an optimal interface will result in a one-to-one mapping between hidden information and human actions. Another way to put this is — given the states and actions output by our models — the system should be able to accurately infer θ. For instance, during one interaction our LED interface may display signals that guide the human to the left of the room. Based on this sequence of states and human actions we should be able to infer the unique θ that the interface had in mind (i.e., the phone is on the left of the room). Let s and θ be sampled from $\mathcal{D}$, and let $\tau(s, \theta) = \big((s, a), (s', a'), \dots\big)$ be a sequence of k counterfactual states and actions starting at s:

$$a = \mathcal{H}_\phi\big(s, \mathcal{R}_\psi(s, \theta)\big), \qquad s' = f(s, a) \tag{13}$$

Here we use our learned human and interface models and the system dynamics from Equation (1) to rollout a hypothetical interaction: given that the human starts at s with hidden information θ, sequence $\tau(s, \theta)$ predicts how the system will behave. We then decode this sequence to try and infer θ from the states and actions:

$$\mathcal{L}_{dist}(\phi, \psi, \sigma) = \sum_{(s,\theta)\in\mathcal{D}} \left\| \theta - \Delta_\sigma\big(\tau(s, \theta)\big) \right\|^2 \tag{14}$$

where $\Delta_\sigma : \mathcal{S}^k \times \mathcal{A}^k \rightarrow \Theta$ is a decoder model with weights σ. This decoder is the third and final network within LIMIT. Equation (14) is minimized when the interface policy $\mathcal{R}_\psi(s, \theta)$ outputs signals x that cause the human model $\mathcal{H}_\phi$ to take different actions for different θ, enabling the decoder Δ_σ to successfully identify the hidden information θ behind these actions. Our loss function $\mathcal{L}_{dist}$ is therefore analogous to $\mathcal{T}_{dist}$: minimizing both terms encourages a one-to-one mapping between hidden information and human actions.

Loss Function. We finally combine Equation (12) and Equation (14) to generate the loss function for training LIMIT:

$$\mathcal{L}(\phi, \psi, \sigma) = \mathcal{L}_{conv}(\phi, \psi) + \mathcal{L}_{dist}(\phi, \psi, \sigma) \tag{15}$$

This loss function is used to update the human model $\mathcal{H}_\phi$, the interface policy $\mathcal{R}_\psi$, and the decoder Δ_σ. Note that this loss function is over *continuous space*, not over discrete space like the relation presented in Equation (10). The human model and decoder are purely for training purposes. During interaction the LIMIT interface selects signals x according to the learned model $\mathcal{R}_\psi$, and then

Algorithm 1 LIMIT: Learning Interfaces to Maximize Information Transfer

```
Initialize model weights $\phi$, $\psi$, and $\sigma$
Initialize dataset $\mathcal{D} \leftarrow \{\}$
for each interaction do
    $\theta \sim P(\theta)$                                    ▷ Interface observes hidden information
    for timestep $t = 0 \dots T$ do
        if $length(\mathcal{D}) \geq batch\_size$ then
            Sample batch of recent $(s, a, \theta) \in \mathcal{D}$
            Train $\mathcal{H}_\phi$, $\mathcal{R}_\psi$, and $\Delta_\sigma$ to minimize $\mathcal{L}$
        end if
        $s^t \leftarrow$ measured system state
        $x^t \leftarrow \mathcal{R}_\psi(s^t, \theta)$                          ▷ Display signal $x^t$ to human
        $a^t \leftarrow$ human action
        $\mathcal{D} \leftarrow (s^t, x^t, a^t, \theta^t)$                     ▷ Append data to dataset
        $s^{t+1} \leftarrow f(s^t, a^t)$                          ▷ Transition to next state
    end for
end for
```

displays these signals to the actual human operator. Because the model structure and loss function of LIMIT closely mirror Equation (10), our proposed LIMIT approach learns to maximize a real-time, tractable proxy of conditional information gain. See Algorithm 1 for an outline of LIMIT. To download an implementation of LIMIT, see our repository: https://github.com/VT-Collab/LIMIT-learning-interfaces. The code in this respository corresponds to the 2D simulations from Section 5.

As an aside, we recognize that recent works also attempt to estimate information gain [2, 27, 33]. However, these learning approaches are not applicable to our problem setting because (a) they require offline training data and (b) they maintain a static estimate of information gain. LIMIT learns online, from the current user, and co-adapts alongside the human to maximize a *proxy* of conditional information gain.

4.4 Accounting for Human Co-Adaptation

So far we have focused on the interface's perspective, and learned an interface policy that maximizes information gain. Importantly, this interface does not know the human's task or reward function: the interface is learning to correlate humans actions with hidden information, and not necessarily to perform the task correctly. Consider our running example where the human is looking for their missing phone. The interface could learn to turn on the *blue light* when the phone is on the *right* side of the room and the *red light* when the phone is on the *left* side. But the human initially interprets this feedback with the *opposite* mapping: perhaps the human goes left for the blue light and right for the red light. From the interface's perspective this is interpretable behavior that maximizes information gain (i.e., human actions are correlated with θ). But from the human's perspective this is the exact opposite of what we wanted: instead of maximizing reward, the human is guided away from their phone!

We recognize that humans are not static agents; over time, the human will inevitably adapt to the interface. At the end of each interaction the human observes their reward $R(\xi, \theta)$. By reasoning over this reward and the previous signals and actions, the human may shift their policy $\pi_{\mathcal{H}}$ to improve performance [17, 24, 26, 39]. Returning to our example, once the human realizes that they are going away from the phone using the initial $\pi_{\mathcal{H}}$, they may switch their interpretation so that they correctly go right for the blue light and left for the red light. Of course, this adaptation is not a one-way street; as the human adapts to the interface, LIMIT should also adapt to the human's changing policy. We explicitly encourage personalization by biasing the system's learning towards *recent* human data. When sampling (s, a, θ) tuples in Algorithm 1, we set the probability of sampling the most recent data (s^N, a^N, θ^N) as exponentially more likely than sampling the first datapoint (s^0, a^0, θ^0). Overall, LIMIT learns an interface policy that transfers hidden information to the human; the human is then responsible for taking advantage of this information and maximize task reward. We will test how LIMIT adapts to the human — and how the human co-adapts to the learning interface — through our simulations and user studies.

5 Simulations

We first compare our LIMIT algorithm to naive alternatives and a state-of-the-art baseline across controlled simulations. Within these simulations the interface knows the hidden information θ (e.g., the location of the human's phone), and the human is trying to maximize their task reward $R(\xi, \theta)$ (e.g., reach their phone by the end of each interaction). The interface's signal is a vector x, and the simulated human interprets this signal to select their own action a. Importantly, our simulated humans are adaptive agents; they change how they interpret the signals between interactions based on their past experiences and observed rewards. The interface must learn to personalize alongside these shifting humans and accurately convey the hidden information.

Interface Algorithms. We compare five different methods for selecting the feedback signals:

- **Naive.** The interface multiplies the vector (s, θ) by a randomized matrix to get signal x. The matrix elements are uniformly randomly sampled between $[-1, +1]$.
- **Bayes [28].** The interface uses a matrix to map (s, θ) to x. At the end of each interaction the interface observes the task reward $R(\xi, \theta)$. The elements of the matrix are updated to maximize this reward using Bayesian optimization [25].
- **Convey.** An ablation of our approach where the interface policy is only trained with loss $\mathcal{L}_{conv}$ in Equation (12).
- **Distinguish.** An ablation of our approach where the interface policy is only trained with loss $\mathcal{L}_{dist}$ in Equation (14).
- **LIMIT.** Our proposed approach from Algorithm 1. For the implementation of LIMIT used in these simulations see https://github.com/VT-Collab/LIMIT-learning-interfaces

We note that the **Bayes** method adopted from [28] has access to the human's reward function, and uses this reward function when designing the feedback signals. By contrast, within our problem setting we *do not assume* any knowledge of $R(\xi, \theta)$. So while we believe that **Bayes** is the closest existing alternative to **LIMIT**, it is important to remember that **Bayes** knows the human's reward function while **LIMIT** is task-agnostic.

Simulated Humans. In Sections 5.1–5.3 we pair the interface with two types of simulated humans: rotate and align. Both types of simulated humans take actions based on signal x, where x is a vector with elements bounded between $[-1, +1]$.

- **Rotate.** This simulated human *rotates* x to get action a. In our 1D environment the human multiplies x by $+1$ or -1 (to change the sign). In the 2D environment the human multiplies x by an $SO(2)$ rotation matrix.
- **Align.** This simulated human *rotates and scales* x to get action a. The rotation is the same as **Rotate**. For scaling, the human multiplies x by a value between $[-1, +1]$.

Both types of simulated humans update their rotation (and scaling) at the end of each interaction to *co-adapt* to the interface. In Sections 5.1–5.3 the human is attempting to reach their missing phone: the reward function is the negative distance between the human's final position and the

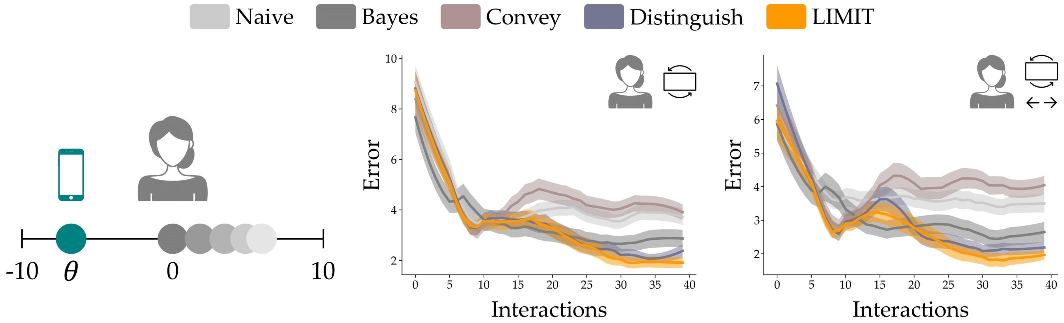

Fig. 2. Simulation results in 1D environment. (Left) Error is the distance between the human's final position and the phone location θ. (Middle) Interfaces paired with the **Rotate** human. A repeated measures ANOVA reveals that the interface type had a significant effect on error ($F(4, 396) = 9.2, p < .001$), with **LIMIT** resulting less error than the alternatives ($p < .05$). (Right) Interfaces paired with the **Align** human. The interface algorithm affects error ($F(4, 396) = 16.8, p < .001$); pairwise comparisons show that **LIMIT** leads to less error than all alternatives besides **Distinguish** ($p < .05$).

phone position: $R(\xi, \theta) = -\|s^t - \theta\|^2$. In order to co-adapt to the interface the human randomly samples N recent interactions and finds the rotation matrix (and scalar) that would have maximized reward over those N interactions. Put another way, the human co-adapts so that their mapping from signals to actions would have increased their task reward over recent interactions.

Environments. Our simulated environments are shown in Figure 2 and Figure 3. The 1D environment is a number line and the 2D environment is an x-y plane. At the start of each interaction the human begins at the origin and the interface samples a random phone position θ. The human and interface interact for 10 timesteps: during each timestep the interface displays signal x, the human takes action a, and the state transitions according to $s^{t+1} = s^t + a^t$. At the end of the interaction we measure the distance $\|s^t - \theta\|$ between the human and their missing phone. We emphasize that $\mathcal{S}$, $\mathcal{A}$, $\mathcal{X}$, and Θ are continuous spaces, and the hidden information θ is known only by the interface.

5.1 Single-DoF Environment

We start with the 1D environment. Here the state, signal, action, and hidden information are all scalars. The simulated human and interface collaborate across 40 interactions: at the start of each interaction θ is sampled uniformly at random from $[-10, +10]$, and we measure the *error* between the final state and θ. The averaged results across 100 simulations are reported in Figure 2. For humans that co-adapt using **Rotate** or **Align** we find that **LIMIT** leads to the lowest average error (i.e., **LIMIT** users most accurately reach their phone).

5.2 Two-DoF Environment

We next perform the same experiment in a 2D environment where the state, signal, action, and θ are all two-dimensional and continuous vectors. The simulated human and interface work together across 100 interactions, and the human's position is reset to the origin $(0, 0)$ at the start of each interaction. We display the averaged results for 50 simulations in Figure 3. Decreasing *error* indicates that — for each interface algorithm — the distance between the human's final state and the hidden θ decreases over interactions. However, we again find that **LIMIT** has the lowest mean error with our **Rotate** and **Align** humans.

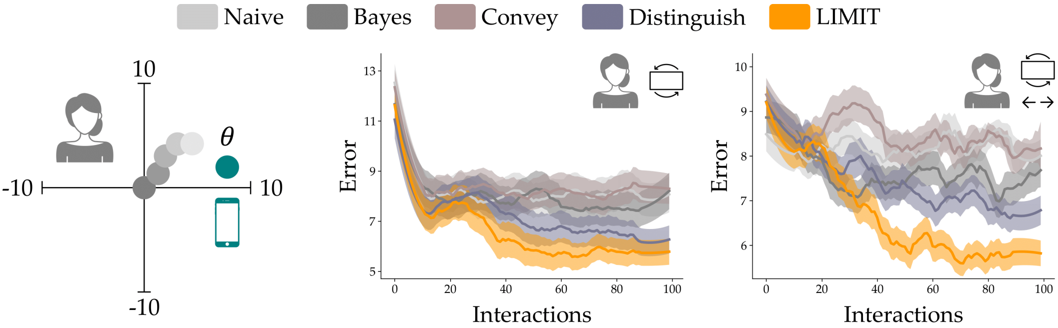

Fig. 3. Simulation results in 2D environment. (Left) The human observes vector x and tries to reach hidden location θ. (Middle) Results with **Rotate** human: differences here are not statistically significant. (Right) Interface paired with an **Align** human. Here interface type has a significant effect on error ($F(4, 196) = 9.0$, $p < .001$), and humans using **LIMIT** have less error by the final interaction than humans using **Naive**, **Bayes**, or **Convey** ($p < .001$).

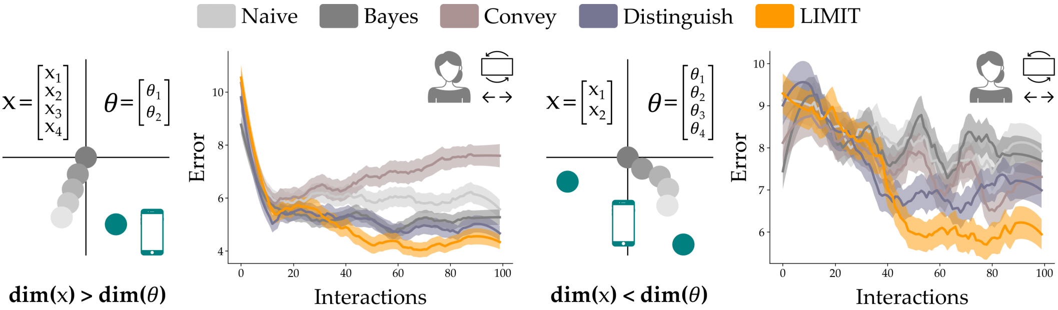

Fig. 4. Simulation results in the 2D environment when the signal x and hidden information θ have different dimensions. (Left) The interface signal is 4-dimensional, but only two dimensions are necessary to convey position θ. Interface type has a significant effect on error ($F(4, 196) = 10.2$, $p < .001$) and **LIMIT** results in lower final error than either **Naive** or **Convey** ($p < .05$). (Right) Now the hidden information is 4 dimensional, and the interface must embed this θ to a lower-dimensional signal x (e.g., the interface is trying to convey two phone locations). Humans reach different errors with different methods ($F(4, 196) = 4.6$, $p < .001$), but **LIMIT** yields less error than all baselines ($p < .05$).

5.3 Mismatch between Signals and Information

In our next simulation we focus on the 2D environment and the **Align** human. We vary the dimensions of x and θ to test scenarios where the interface has additional feedback channels, $dim(x) > dim(\theta)$, or where the hidden information is more complex than the interface, $dim(x) < dim(\theta)$. On the left of Figure 4 the signal x is a 4-dimensional vector: because θ here is only an (x, y) position, the robot must learn how to harness two additional feedback dimensions. The second scenario is shown on the right side of Figure 4. Here θ is a 4D vector specifying the position of two hidden phones: the interface must learn to embed this higher-dimension hidden information into a 2-dimensional signal x. Within these controlled environments with simulated users, we again observe that **LIMIT** outperforms the baselines.

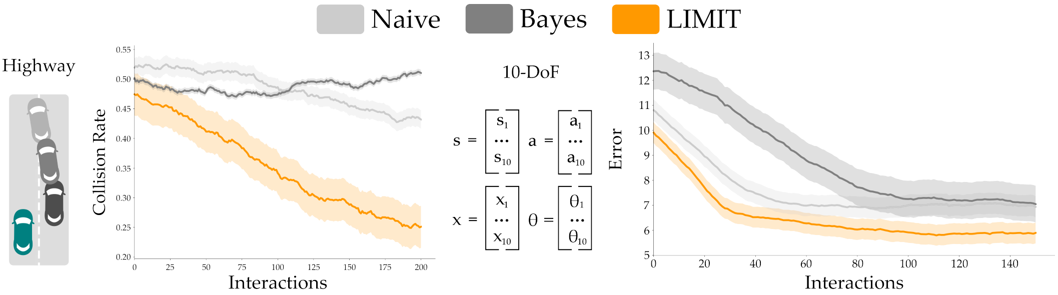

Fig. 5. (Left) Simulation results from the autonomous driving task. The autonomous car has four different driving policies, and must signal its current policy to the human in order to help the human driver avoid a collision. Interfaces generated by **LIMIT** result in a lower collision rate than either **Naive** or **Bayes** ($p < 0.001$). (Right) Results from a 10-dimensional environment. This simulation extends Section 5.2 to a high-dimensional setting where the states, actions, signals, and hidden information are all 10-dimensional vectors in a continuous space. Despite this increase in dimension, the interfaces generated by **LIMIT** still result in a lower error at the end of an interaction than those generated by **Naive** or **Bayes** ($p < 0.001$).

5.4 LIMIT in more Complex Tasks

To test the effectiveness of LIMIT in more complex tasks, we present two additional simulations. These simulations vary in complexity along two axes: the type of information the interface needs to convey to the human, and the dimensionality of the problem setting. We start with an autonomous driving scenario, where the interface attempts to convey the policy of an autonomous car to a nearby human driver. We then end this section by returning to the phone example, but now in a 10-dimensional state-action space. Note that — in these more complex settings — we model the human as a multi-layer perceptrons with one hidden layer. At the end of each interaction, this simulated human co-adapts to the interface by updating the weights of its model to maximize the measured reward. We also conduct supplementary simulations to explore the adaptation rate and human variability in the Appendix (Section C).

Autonomous Driving Task. In this simulation the human agent is driving along a two-lane one-way highway (see Figure 5). Ahead of the human is an autonomous vehicle that the human would like to avoid; the autonomous vehicle may be in either lane of the highway. The autonomous vehicle attempting to signal its *policy* (i.e., how it will change lanes). There are four discrete policies that the autonomous vehicle could be using to drive along the highway:

(1) The robot will always stay in the right lane (θ_1)
(2) The robot will always stay in the left lane (θ_2)
(3) The robot will merge into the human's lane (θ_3)
(4) The robot will merge into the opposite lane of the human (θ_4)

Each interaction lasts five timesteps, after which the human car and autonomous car are reset. The autonomous car's policy is randomly sampled at the start of the interaction: this policy is the autonomous car's hidden information θ. The human agent is penalized when it collides with the autonomous car, so it should learn to infer θ from the signals produced by the interface. To test the effectiveness of LIMIT, we compared the performance of LIMIT to our **Naive** baseline and the state-of-the-art **Bayes** baseline. Figure 5 (Left) shows the average collision rate per interaction over time; LIMIT outperforms the baselines, approaching a collision rate of 0. This result suggests that LIMIT can be successfully applied in scenarios where the interface needs to convey more complex information to the human (i.e., the policy of another agent).

10-DoF Environment. In this simulation, we replicate the experiments from Sections 5.1 and 5.2, but now increase the dimension of the state, action, and hidden information so that each component is a 10-dimensional continuous space. The interface knows the goal position θ, and attempts to convey this hidden information through 10-dimensional signals. The simulated human attempts to reach the hidden goal position over 10 timesteps based on the signals from the interface. Like before, we compare the performance of LIMIT against the **Naive** and **Bayes** algorithms. Figure 5 (Right) shows that LIMIT outperforms the baselines — despite the high-dimensionality of the problem setting — while the performance of **Bayes** and **Naive** follow random error. This result indicates that LIMIT can be extended to higher-dimensional problem settings.

6 Can Users Understand Learned Interfaces?

Our simulations from Section 5 suggest LIMIT learns interface policies that better convey hidden information than the alternatives. However, these tests were run with simulated humans in controlled environments. How does LIMIT fare with actual users? To evaluate the real-world performance, we first conducted an online user study via Google Forms where participants observed colored signals and attempted to guess the 2D position of their missing phone. Here the interfaces were trained offline — using synthetic human data — and we measured whether participants (a) found the learned mappings intuitive and (b) adapted to the interface policy. Our results across 37 online participants indicate that LIMIT interfaces are more interpretable than a randomized baseline.

Independent Variables. We compared **LIMIT** to a **Naive** baseline. During each interaction the human started at state $s = (0, 0)$, observed a signal x, and then clicked *once* to indicate where they thought the phone was hidden. The interface had access to the phone's hidden location θ. **Naive** multiplied the vector θ by a randomized 2×2 matrix to get signal x. **LIMIT** was pre-trained offline using the procedure from Section 4.2. After training, we recorded the signals produced by **Naive** and **LIMIT** for nine different θ positions.

Experimental Setup. At the start of each interaction the online participants were shown a picture of their current position s and two colored bars for signal x (see Figure 6). After observing the state and signal, the participants selected the x-y coordinates that they thought the interface was trying to convey. We then displayed the phone's actual location θ. Because we revealed the hidden information θ after each interaction, users could adapt to the interface over time.

Each participant completed nine interactions with the **Naive** approach and nine interactions with **LIMIT**. The order of the methods was counterbalanced: half of the participants started with **Naive** and the other half started with **LIMIT**.

Dependent Variables. To determine how participants objectively performed with each interface, we measured the distance between their guess s^1 and the phone's actual position θ. Specifically, we defined *Error* as $||s^1 - \theta||$. To understand how participants subjectively perceived each interface, we administered a 7-point Likert scale survey each time the users completed a method. Questions were organized along three scales: how confident users were that their performance *improved* over time, if they *understood* what the interface was trying to convey, and whether the interface was *intuitive*. The exact items on the survey are listed in Section A.

Participants. A total of 38 adults took part in this online survey. At the start of the survey we asked participants to read the instructions, follow an example interaction, and then answer three qualifying questions to test their understanding. Below we report the results for 37 participants who correctly answered these questions and finished the survey.

Results. Our results are summarized in Figure 7. We first measured the *error* between the human's predictions with **Naive** and **LIMIT** across nine rounds of interaction. Paired t-tests revealed that users were significantly more accurate with **LIMIT** than with **Naive** ($t(295) = 2.17$, $p < .05$).

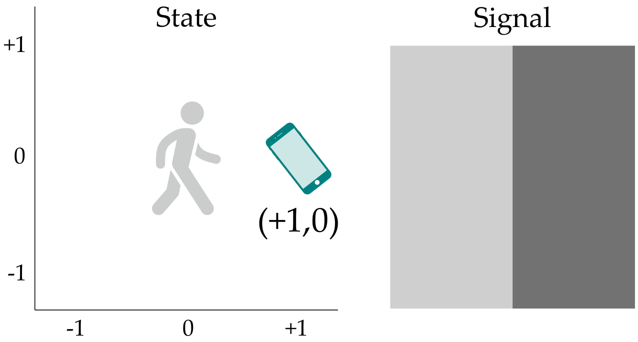

Fig. 6. Example of a state and signal displayed to online study participants in Section 6. At first the phone's location θ (in blue) was *not* shown: participants just saw the human and the signal, and tried to guess the phone's location based on the color of the two bars. For instance, perhaps the color of the first bar corresponded to the x-axis and the color of the second bar corresponded to the y-axis. We then revealed θ and moved on to a new state and signal.

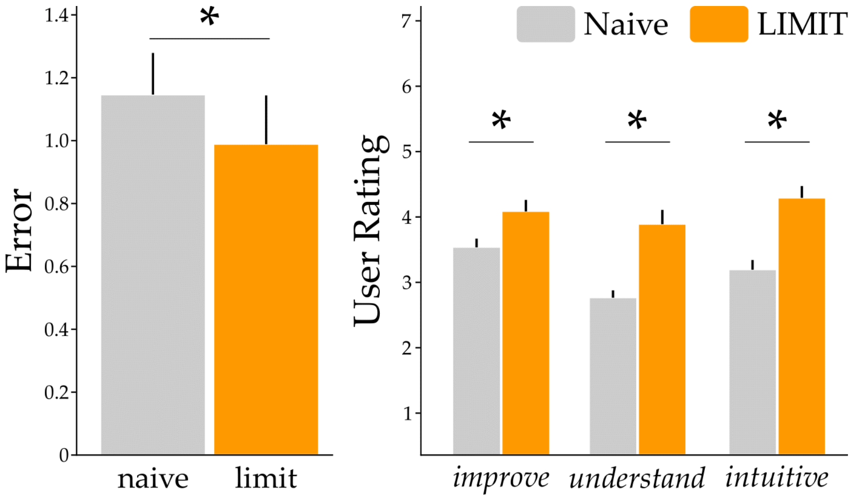

Fig. 7. Results from the online user study in Section 6. (Left) Our 37 participants interpreted the signals in Figure 6 to predict the phone's hidden position. The error between their guess and the actual position was lower with **LIMIT** than with **Naive**. (Right) With **LIMIT** participants thought they better improved over time, better understood what the interface was trying to convey, and overall perceived the interface as more intuitive. Error bars show standard error and $*$ denote statistical significance ($p < .05$).

Put another way, when working with **LIMIT** the online participants were better able to infer the hidden information and select high-reward states. The subjective responses suggested that participants perceived a difference between the methods. To analyze the Likert results we first confirmed that our three multi-item scales (*improve*, *understand*, and *intuitive*) were reliable with a Cronbach's $\alpha > 0.7$. We then grouped each scale into combined scores and performed paired t-tests. Overall, our 37 users thought that they *improved* more with **LIMIT** ($t(36) = -2.41, p < .05$), better *understood* what the **LIMIT** interface was trying to communicate ($t(36) = -4.31, p < .001$), and found the **LIMIT** interface to be more *intuitive* ($t(36) = -4.16, p < .001$).

We emphasize that — within this online user study — each interaction only lasted a single timestep, and the pre-trained interface policy did not learn or adapt alongside actual user data. However, our results across 37 participants are a first step towards confirming that the interfaces learned using **LIMIT** are interpretable for everyday human users.

7 User Study

Our online study was a first step towards evaluating LIMIT. To understand how LIMIT performs across repeated human-robot interaction with different types of interfaces, we next performed an *in-person* user study. In this study participants completed three separate tasks that each had different interface modalities: *sounds*, *lights*, and *haptics*. The interface knew some hidden information θ (e.g., the correct joint position for a robot arm), and selected feedback signals to convey θ to the human. There were no immediately obvious conventions for interpreting this feedback. For instance, the interface played musical notes of varying pitches to indicate where to guide a robot arm: one person might assume higher notes indicate moving the arm to the right, while others might think lower notes denote the same motion. Over multiple timesteps and interactions the interface and participant co-adapted. The interface had to learn how to map θ to signals, and users had to learn to interpret these signals and complete the task.

Independent Variables. We compared two algorithms for selecting feedback: **Bayes** and **LIMIT**. **Bayes** is a state-of-the-art approach adapted from [28] that treats the mapping from signals to rewards as a black box. The interface explores this black box by using Bayesian optimization [25] to search for signals that maximize task reward. We note that **Bayes** knows what task that the human is trying to complete (i.e., in this baseline the interface has access to the human's reward function). As such, **Bayes** actually has more information than **LIMIT**, where the interface never knows the human's objective. For **LIMIT** we used our proposed approach from Algorithm 1. Both methods were pre-trained offline with simulated partners using the procedure from Section 4.2.

Interfaces. To demonstrate that our work is not tied to any specific type of interface, we performed tests with interfaces that employed sounds, lights, or haptics (see Figure 8).

- **Sounds.** At each timestep the system played a musical note (G) through speakers and headphones. The signal x was the 1-DoF pitch of this musical note: the interface could continuously vary the pitch along two octaves.
- **Lights.** This interface was similar to the online user study in Section 6. Users carried a interface with two strips of LED lights. The 2-DoF signal x was the brightness of the lights on each strip: at each timestep the interface could change the number and intensity of illuminated LEDs.
- **Haptics.** Here we wrapped three pneumatic bags around a robot arm [38]. Users kinesthetically interacted with these bags as they moved the robot. The 3-DoF signal x was the pressure of each bag: at every timestep the system could increase or decrease the pressures between 0 and 3 PSI.

We emphasize that the interfaces were used separately. Users completed tasks with only sounds, lights, or haptics, and did not interact with more than one interface at a time.

Experimental Setup. We divided the study into three tasks, one for each of the interfaces (see Figure 8). Note that the interface used in the task corresponds to the task name.

In *Sounds* participants physically interacted with a 7-DoF Franka Emika robot arm. The robot started each interaction in its home position, and users kinesthetically guided the robot to move its end-effector. The system state s was the position of the end-effector. Users were free to select their final x and z coordinates; here θ corresponded to the correct position along the y-axis. The robot used auditory feedback (the pitch of a musical note) to indicate θ.

Lights matches our motivating example of a person walking around a room to find their missing phone. Participants moved in an empty 10ft by 10ft space while wearing an HTC Vive position

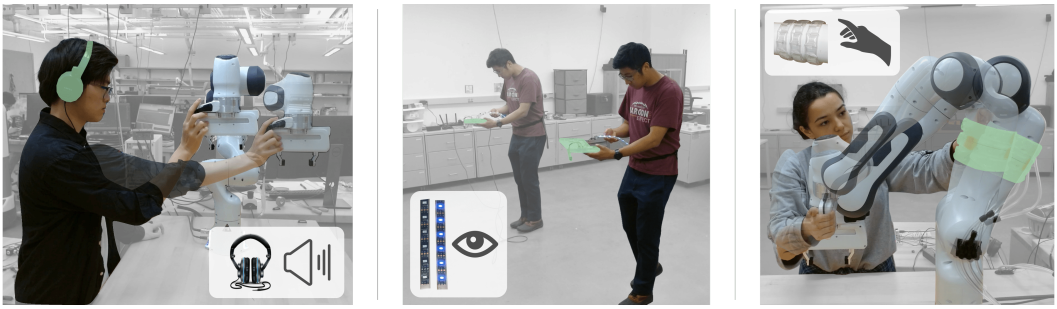

Fig. 8. Interfaces and tasks for our in-person study. (Left) In *Sounds* the human moved a robot arm back and forth along a single axis. The interface played a musical note of varying pitches to convey the correct arm position. (Middle) In *Lights* participants walked around an empty space. The interface illuminated two LED strips to indicate the location of their missing phone. (Right) in *Haptics* we wrapped pneumatic bags around a robot arm [38]. Participants needed to guide the robot to the correct height, orientation, and distance from their body: the interface inflated and deflated the haptic bags to convey these features.

tracker for real-time measurements: the state s was the user's position. Hidden information θ corresponded to the correct x-y position of their missing phone. As users walked they carried our lights interface with two LED strips. Once the user was confident that they had reached the correct position (i.e., they thought they had found their phone), they informed the proctor and the interaction ended.

For *Haptics* participants again interacted with the 7-DoF Franka Emika robot arm. In this task the robot followed a parameterized trajectory: the robot had a fixed start, but it was up to the user to correct the robot's goal. Here state s was the position and orientation of the end-effector. The robot knew where the endpoint the trajectory should be: hidden information θ contained the correct height, yaw, and distance from the person. As users physically corrected the robot's trajectory they got haptic feedback from the wrapped display. Note that in *Sounds* θ and x are 1-dimensional, in *Lights* they are 2-dimensional, and in *Haptics* they are 3-dimensional.

Each task contained multiple interactions. At the start of the interaction we reset the environment (i.e., the robot returned to its home position or the human walked to the center of the room). The interface then sampled a random $\theta \sim P(\cdot)$ from a uniform prior: this θ was held constant during the interaction but changed between interactions. The interaction lasted multiple timesteps as the human physically guided the robot or walked around the room. At the end of the interaction the system revealed the actual θ back to human: in *Sounds* and *Haptics* the robot arm moved to the correct pose, and in *Lights* the proctor showed the person the correct place to stand. The *Sounds* and *Lights* tasks included 10 interactions, and the *Haptics* task had 5 interactions. Note that the input and output dimensions of each algorithm were changed to accommodate the change in DoF between tasks; all other architectural changes are addressed in Appendix B.

Participants and Procedure. We recruited 11 participants (3 female, ages 25.5 ± 5.15) from the Virginia Tech community. Prior to the experiment all participants provided informed written consent consistent with university guidelines (IRB #20-755). Two of the eleven participants had never interacted with robots before. None of these in-person users took part in the online study.

Each participant completed all three tasks twice: once with **Bayes** and once with **LIMIT**. The order of the tasks was counterbalanced: e.g., some users started with *Sounds* while others started with *Haptics*. The order of the algorithms was also counterbalanced: for each task, half of the users

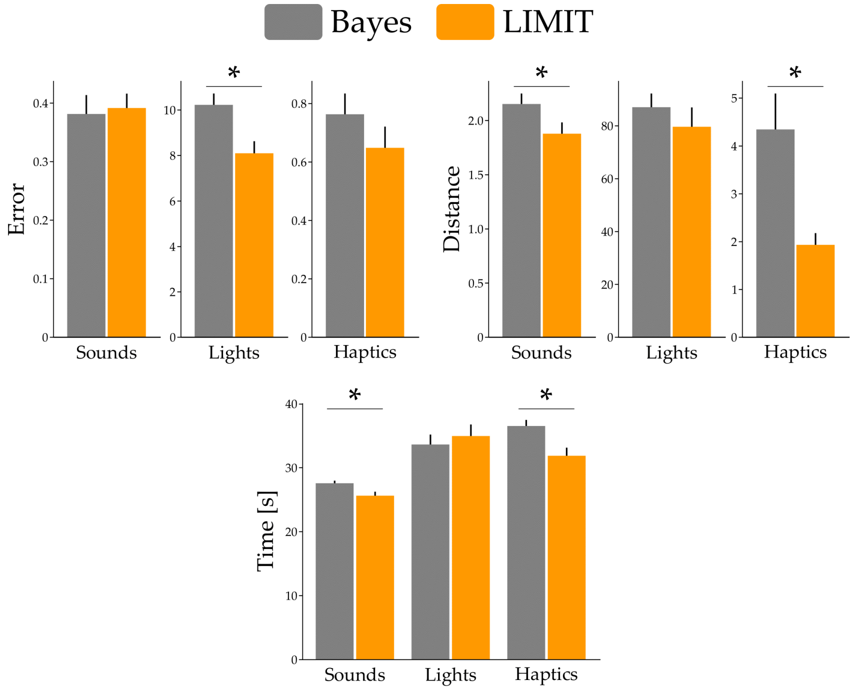

Fig. 9. Objective results from our in-person user study. These results are broken down by task; *Sounds* and *Haptics* are performed with a robot arm. (Left) Error between θ and human's final state s^T. (Middle) Distance the human travels during an interaction. For error and distance the units vary between tasks: in *Sounds* the units are meters, in *Lights* the units are feet, and in *Haptics* the units are meters (end-effector position) plus radians (end-effector orientation). (Right) Time taken to complete an interaction. Error bars show standard error and an $*$ denotes statistical significance ($p < .05$).

started with **LIMIT**. Participants were never told which algorithm they were working with during the experiment, and did not know which algorithm was our approach.

Dependent Measures – Objective. We measured the states, signals, and actions during each timestep. Recall that in every task the human was trying to reach θ (e.g., the correct robot position). We therefore recorded the *Error* between the system's final state s^T and the desired position θ, so that $Error = \|s^T - \theta\|$. To assess how the human behaved within the task, we also measured the total *Distance* they traveled during an interaction: $Distance = \sum_{t=1}^{T} \|s^t - s^{t-1}\|$. Here lower distances indicate that the human understood the interface and went directly to the goal, while higher distances suggest the human often backtracked or changed directions. Finally, we measured the *Time* it took to complete the task.

Dependent Measures – Subjective. After each task and algorithm participants completed a 7-point Likert scale survey. This survey was designed to measure the user's subjective perception of the interface along four multi-item scales. We asked users if they felt like their performance *improved* over time, if they could *understand* what the interface was trying to communicate, if the signals seemed *consistent*, and whether they thought they would continue to improve if they kept working with this interface (*intuitive*).

Hypotheses. We had two hypothesis for the user study:

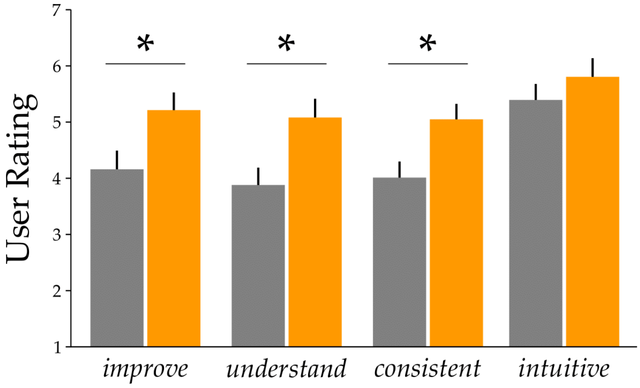

Fig. 10. Subjective results from our 7-point Likert scale survey. Participants thought that they improved more with **LIMIT**, and that **LIMIT** signals were more understandable and consistent. Users also scored **LIMIT** as more intuitive than **Bayes**, but this was not statistically significant. An $*$ denotes significance ($p < .05$).

H1. *Users will have less error and complete the task more efficiently with **LIMIT**.*

H2. *Users will subjectively prefer interfaces that use **LIMIT** to learn feedback signals.*

Results. The objective results are summarized in Figure 9, and the subjective results are displayed in Figure 10. Please also see videos of the user study here: https://youtu.be/IvQ3TM1_2fA

To explore hypothesis **H1** we analyzed the error, distance travelled, and time taken when getting feedback from **Bayes** or **LIMIT**. For each of these metrics lower was better: an effective interface should help the human complete the task correctly and efficiently. Paired t-tests revealed that participants reached significantly lower error when working with **LIMIT** in the *Lights* task ($t(109) = 3.02$, $p < .05$). Interestingly, the difference in error was not significant for either *Sounds* or *Haptics*. Instead, our proposed interface enabled users to complete these tasks more efficiently. **LIMIT** resulted in significantly less distance traveled for *Sounds* ($t(109) = 2.24$, $p < .05$) and *Haptics* ($t(54) = 3.67$, $p < .001$). Along the same lines, **LIMIT** led to shorter interactions for *Sounds* ($t(109) = 3.1$, $p < .05$) and *Haptics* ($t(54) = 3.98$, $p < .001$). We conclude that interfaces which co-adapted to participants using **LIMIT** selected more helpful signals than **Bayes**. But we also recognize that the way in which **LIMIT** helped users differed from one interface to another. For the *Lights* interface **LIMIT** improved the human's task reward without significantly changing the distance traveled or time taken. By contrast, for the *Sounds* and *Haptics* interfaces **LIMIT** and **Bayes** both obtained similar task reward, but **LIMIT** enabled users to reach this reward more quickly and efficiently.

To illustrate how **Bayes** and **LIMIT** affected the human's performance we highlight a *Lights* example in Figure 11. Across 10 repeated interactions the human walked to reach hidden goals θ. When the interface used **Bayes** to select signals x, the human's error was roughly constant from one interaction to another. But under **LIMIT** this interface and human co-adapted so that the human could more accurately interpret the interface's signal after four interactions.

We now turn to hypothesis **H2** and our Likert-scale survey in Figure 10. Our survey items are listed in Appendix A. To process the subjective results we first confirmed that each of the four scales were reliable (Cronbach's $\alpha > 0.7$). We then grouped each multi-item scale into a combined score and performed paired t-tests to determine whether **LIMIT** received significantly higher scores than **Bayes**. Users reported that their performance *improved* more with **LIMIT** ($t(32) = 2.75$, $p < .05$), and they better *understood* what the **LIMIT** interface was trying to convey

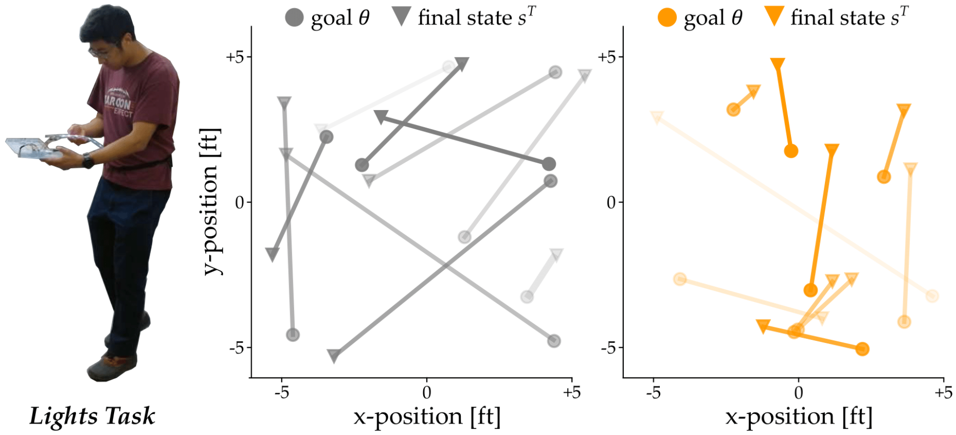

Fig. 11. Example data from *Lights*. (Middle) For one user we plot the error between their final position and the goal position as a function of interaction number for both **Bayes** and **LIMIT**. (Right) Ideally the final state should match the goal θ. We visualize this same data by displaying the user's final x-y position and the actual x-y location of the goal. The user completed a total of 10 interactions: lighter lines signify their first interactions and darker lines their final interactions. Error generally decreases over time with **LIMIT**.

($t(32) = 3.08$, $p < .05$). The participants also rated **LIMIT** as having more *consistent* feedback than **Bayes** ($t(32) = 2.58$, $p < .05$). When asked to explain their preferences, users commented that LIMIT "*was a lot more intuitive to adjust to*," in part because it "*seemed more consistent*" and maintained a distinct, one-to-one mapping from information θ to signals x.

Discussion. The experimental results from our in-person user study suggests that **LIMIT** outperforms **Bayes** in an ensemble of interfaces, both preferentially and numerically. However, we note that the difference between performance (albeit significant) is not as large as one may expect. We emphasize that the **Bayes** algorithm *has access to the human's intent* (i.e., their goal or reward function), giving it a significant advantage to **LIMIT**. However, as is evidenced empirically, users were able to adapt to signals produced by **LIMIT** more easily and achieved better results in higher-dimensional settings. In more complex tasks, the performance gap between **LIMIT** and **Bayes** is more clear, even when **Bayes** has access to the human's intent (see Section 5.4).

Summary. Our in-person user study evaluated **LIMIT** across three types of interfaces: audio feedback, visual feedback, and haptic feedback. The experimental results suggest that **LIMIT** learns to select meaningful and interpretable signals that help users complete their tasks. Across all objective and subjective metrics, we found that **LIMIT** scored as well as or better than a state-of-the-art baseline that has access to the task reward.

8 Conclusion

In this paper we introduced LIMIT, a co-adaptive approach to learn interface mappings from scratch. Learning interfaces is challenging because the way people respond to signals varies across tasks, users, and interface types. To address these challenges we hypothesized that interfaces should learn policies that maximize correlation between the human's actions and the interface's information. We derived a learning algorithm that updates the interface's signals in real-time to optimize for a tractable proxy of information gain. We then put LIMIT to the test across controlled simulations, an online survey, and in-person user studies. When compared to naive baselines and a state-of-the-art alternative with auditory, visual, and haptic interfaces, we found that LIMIT results in better task performance and higher subjective ratings.

Limitations. LIMIT is a step towards robots that autonomously personalize their feedback for the current user. One advantage of LIMIT is that it does not need to know what task the human is trying to complete (i.e., what the human is using the signals for). Our key assumption here is that — no matter what task the human has in mind — the human should respond in different ways to different hidden information θ. While maximizing information gain makes sense for the experiments presented in this paper, there are also settings where the human should maintain the *same* actions even when the hidden information *changes*. For example, imagine a driving scenario where the interface is communicating the location of nearby cars and pedestrians. The human is driving straight ahead, and should not change their actions if another car passes by or if a pedestrian is walking on the opposite sidewalk. When applied to this context, LIMIT may learn to cause the human driver to slow down or speed up, even though these changes are not necessary. Our future work will explore how LIMIT can be combined with task-specific objectives to ensure that the signals are always necessary and meaningful. Our initial hypothesis here is that the task-specific reward function $R(\xi, \theta)$ could be incorporated within Equation (15) so that LIMIT trains interfaces to simultaneously maximize interpretability and performance.

Future Works. A key application of LIMIT could be in sensory substitution, a scenario where neither the engineer nor the user has an informed prior over interface design. For example, brain-computer interfaces (BCI) have been connected with proprioceptive feedback [10]. Here the proprioceptive feedback (e.g., haptic stimulation) could be tuned using an approach like LIMIT to improve the accuracy of the BCI interface. Along similar lines, vibrotactile biofeedback interfaces [6] could be made more effective using LIMIT to adjust the mapping between vibrotactile stimuli and user inputs. Moving forward, we see LIMIT as a step towards robots and interfaces that autonomously change their feedback to make the system more intuitive, understandable, and user-friendly.

References

[1] Rasmus S Andersen, Ole Madsen, Thomas B Moeslund, and Heni Ben Amor. 2016. Projecting robot intentions into human environments. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 294–301.

[2] Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and Devon Hjelm. 2018. Mutual information neural estimation. In *International Conference on Machine Learning*. 531–540.

[3] Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. 2018. Social robots for education: A review. *Science Robotics* 3, 21 (2018), eaat5954.

[4] Elizabeth Cha, Yunkyung Kim, Terrence Fong, and Maja J Mataric. 2018. A survey of nonverbal signaling methods for non-humanoid robots. *Foundations and Trends in Robotics* 6, 4 (2018), 211–323.

[5] Ravi Teja Chadalavada, Henrik Andreasson, Robert Krug, and Achim J Lilienthal. 2015. That's on my mind! Robot to human intention communication through on-board projection on shared floor space. In *European Conference on Mobile Robots*. 1–6.

[6] Aniruddha Chatterjee, Vikram Aggarwal, Ander Ramos, Soumyadipta Acharya, and Nitish V Thakor. 2007. A brain-computer interface with vibrotactile biofeedback for haptic information. *Journal of neuroengineering and rehabilitation* 4, 1 (2007), 1–12.

[7] Yuhang Che, Allison M Okamura, and Dorsa Sadigh. 2020. Efficient and trustworthy social navigation via explicit and implicit robot–human communication. *IEEE Transactions on Robotics* 36, 3 (2020), 692–707.

[8] Thomas M Cover and Joy A Thomas. 2012. *Elements of Information Theory*. John Wiley & Sons.

[9] Dalia De Santis. 2021. A framework for optimizing co-adaptation in body-machine interfaces. *Frontiers in Neurorobotics* 15 (2021), 662181.

[10] Darrel R Deo, Paymon Rezaii, Leigh R Hochberg, Allison M Okamura, Krishna V Shenoy, and Jaimie M Henderson. 2021. Effects of Peripheral Haptic Feedback on Intracortical Brain-Computer Interface Control and Associated Sensory Responses in Motor Cortex. *IEEE transactions on haptics* 14, 4 (2021), 762–775.

[11] Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. 2013. Legibility and predictability of robot motion. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 301–308.

[12] Nathan Dunkelberger, Jennifer L Sullivan, Joshua Bradley, Indu Manickam, Gautam Dasarathy, Richard Baraniuk, and Marcia K O'Malley. 2020. A multisensory approach to present phonemes as language through a wearable haptic device. *IEEE Transactions on Haptics* 14, 1 (2020), 188–199.

[13] Norina Gasteiger, Mehdi Hellou, and Ho Seok Ahn. 2021. Factors for personalization and localization to optimize human–robot interaction: A literature review. *International Journal of Social Robotics* (2021), 1–13.

[14] Soheil Habibian, Ananth Jonnavittula, and Dylan P Losey. 2022. Here's what I've learned: Asking questions that reveal reward learning. *ACM Transactions on Human-Robot Interaction* 11, 4 (2022), 1–28.

[15] Thomas Hellström and Suna Bensch. 2018. Understandable robots-what, why, and how. *Paladyn, Journal of Behavioral Robotics* 9, 1 (2018), 110–123.

[16] Sandy H Huang, David Held, Pieter Abbeel, and Anca D Dragan. 2019. Enabling robots to communicate their objectives. *Autonomous Robots* 43 (2019), 309–326.

[17] Shuhei Ikemoto, Heni Ben Amor, Takashi Minato, Bernhard Jung, and Hiroshi Ishiguro. 2012. Physical human-robot interaction: Mutual learning and adaptation. *IEEE Robotics & Automation Magazine* 19, 4 (2012), 24–35.

[18] Natasha Jaques, Angeliki Lazaridou, Edward Hughes, Caglar Gulcehre, Pedro Ortega, DJ Strouse, Joel Z Leibo, and Nando De Freitas. 2019. Social influence as intrinsic motivation for multi-agent deep reinforcement learning. In *International Conference on Machine Learning*. 3040–3049.

[19] Tobias Kaupp, Alexei Makarenko, and Hugh Durrant-Whyte. 2010. Human–robot communication for collaborative decision making — A probabilistic approach. *Robotics and Autonomous Systems* 58, 5 (2010), 444–456.

[20] Minae Kwon, Sandy H Huang, and Anca D Dragan. 2018. Expressing robot incapability. In *ACM/IEEE International Conference on Human-Robot Interaction*. 87–95.

[21] Kimin Lee, Laura M Smith, and Pieter Abbeel. 2021. PEBBLE: Feedback-Efficient Interactive Reinforcement Learning via Relabeling Experience and Unsupervised Pre-training. In *International Conference on Machine Learning*. 6152–6163.

[22] Mengxi Li, Dylan P Losey, Jeannette Bohg, and Dorsa Sadigh. 2020. Learning user-preferred mappings for intuitive robot control. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 10960–10967.

[23] James F Mullen, Josh Mosier, Sounak Chakrabarti, Anqi Chen, Tyler White, and Dylan P Losey. 2021. Communicating inferred goals with passive augmented reality and active haptic feedback. *IEEE Robotics and Automation Letters* 6, 4 (2021), 8522–8529.

[24] Stefanos Nikolaidis, David Hsu, and Siddhartha Srinivasa. 2017. Human-robot mutual adaptation in collaborative tasks: Models and experiments. *The International Journal of Robotics Research* 36 (2017), 618–634.

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

[26] Sagar Parekh and Dylan P Losey. 2022. Learning Latent Representations to Co-Adapt to Humans. *arXiv preprint arXiv:2212.09586* (2022).

[27] Ben Poole, Sherjil Ozair, Aaron Van Den Oord, Alex Alemi, and George Tucker. 2019. On variational bounds of mutual information. In *International Conference on Machine Learning*. 5171–5180.

[28] Siddharth Reddy, Sergey Levine, and Anca D Dragan. 2022. First Contact: Unsupervised Human-Machine Co-Adaptation via Mutual Information Maximization. In *Advances in Neural Information Processing Systems*.

[29] Fabio Rizzoglio, Maura Casadio, Dalia De Santis, and Ferdinando A Mussa-Ivaldi. 2021. Building an adaptive interface via unsupervised tracking of latent manifolds. *Neural Networks* 137 (2021), 174–187.

[30] Eric Rosen, David Whitney, Elizabeth Phillips, Gary Chien, James Tompkin, George Konidaris, and Stefanie Tellex. 2019. Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays. *The International Journal of Robotics Research* 38, 12-13 (2019), 1513–1526.

[31] Dorsa Sadigh, Nick Landolfi, Shankar S Sastry, Sanjit A Seshia, and Anca D Dragan. 2018. Planning for cars that coordinate with people: Leveraging effects on human actions for planning and active information gathering over human internal state. *Autonomous Robots* 42 (2018), 1405–1426.

[32] Hasti Seifi, Matthew Chun, Colin Gallacher, Oliver Schneider, and Karon E MacLean. 2020. How do novice hapticians design? A case study in creating haptic learning environments. *IEEE Transactions on Haptics* 13, 4 (2020), 791–805.

[33] Jiaming Song and Stefano Ermon. 2019. Understanding the limitations of variational mutual information estimators. In *International Conference on Learning Representations*.

[34] Ryo Suzuki, Adnan Karim, Tian Xia, Hooman Hedayati, and Nicolai Marquardt. 2022. Augmented reality and robotics: A survey and taxonomy for AR-enhanced human-robot interaction and robotic interfaces. In *CHI Conference on Human Factors in Computing Systems*. 1–33.

[35] Daniel Szafir and Danielle Albers Szafir. 2021. Connecting human-robot interaction and data visualization. In *ACM/IEEE International Conference on Human-Robot Interaction*. 281–292.

[36] Hong Z Tan, Charlotte M Reed, Yang Jiao, Zachary D Perez, E Courtenay Wilson, Jaehong Jung, Juan S Martinez, and Frederico M Severgnini. 2020. Acquisition of 500 English words through a TActile phonemic sleeve (TAPS). *IEEE Transactions on Haptics* 13, 4 (2020), 745–760.

[37] Maegan Tucker, Ellen Novoseller, Claudia Kann, Yanan Sui, Yisong Yue, Joel W Burdick, and Aaron D Ames. 2020. Preference-based learning for exoskeleton gait optimization. In *IEEE International Conference on Robotics and Automation*. 2351–2357.

[38] Antonio Alvarez Valdivia, Soheil Habibian, Carly A Mendenhall, Francesco Fuentes, Ritish Shailly, Dylan P Losey, and Laura H Blumenschein. 2023. Wrapping Haptic Displays Around Robot Arms to Communicate Learning. *IEEE Transactions on Haptics* (2023).

[39] Emma M Van Zoelen, Karel Van Den Bosch, and Mark Neerincx. 2021. Becoming team members: Identifying interaction patterns of mutual adaptation for human-robot co-learning. *Frontiers in Robotics and AI* 8 (2021), 692811.

[40] Michael Walker, Hooman Hedayati, Jennifer Lee, and Daniel Szafir. 2018. Communicating robot motion intent with augmented reality. In *ACM/IEEE International Conference on Human-Robot Interaction*. 316–324.

[41] Thomas Weng, Leah Perlmutter, Stefanos Nikolaidis, Siddhartha Srinivasa, and Maya Cakmak. 2019. Robot object referencing through legible situated projections. In *International Conference on Robotics and Automation*. 8004–8010.

Appendix

A Likert Survey Tables

A.1 Online User Study (Section 6)

This section lists the questions on the Likert scale survey from our online user study in Section 6. We organised questions into three scales (*Intuitive*, *Understand*, and *Improve*) and tested their reliability using Cronbach's $\alpha > 0.7$. We then grouped each multi-item scale into a combined score and performed paired t-tests to determine whether **LIMIT** received significantly higher scores than **Naive**. The results of the t-tests are reported in Section 6. Here we list the exact items and their corresponding scale.

Table 1. Questions from our online user study in Section 6

Questionnaire Item	Scale
– I thought the signals had a consistent pattern. – The signals seemed inconsistent or random.	*Intuitive*
– By the end I could accurately predict the phone location. – At the end I was still unsure what the interface was trying to convey.	*Understand*
– I felt like my performance improved over time. – It seemed like my performance stayed about the same.	*Improve*

A.2 In-Person User Study (Section 7)

This appendix lists the questions on the Likert scale survey from Section 7. We organized the items into four scales (*Intuitive*, *Understand*, *Consistent*, and *Improve*) and tested their reliability using Cronbach's $\alpha > 0.7$. We then grouped each multi-item scale into a combined score and performed pair t-tests to determine whether **LIMIT** received significantly higher scores than **Bayes**. The results of t-tests are reported in Section 7. Here we list the exact items and their scale.

B Network Architecture

In this section, we detail the specifics of the neural network architectures used throughout this paper. Nearly all networks in all experiments were multilayer-perceptrons (MLPs), but Table 3 lists the architecture in detail.

Table 2. Questions from our in-person user study in Section 7

Questionnaire Item	Scale
– If I used the interface more, I think I would understand what it was trying to say. – Even if I kept practicing with this interface, I still don't think I would get it.	*Intuitive*
– By the end I could understand what the interface was saying. – At the end I was still unsure what the interface was trying to convey.	*Understand*
– I thought the signals had a consistent pattern. – The signals seemed inconsistent or random.	*Consistent*
– I felt like my performance improved over time. – It seemed like my performance stayed about the same.	*Improve*

Table 3. Neural Network Architecture in Simulations (Section 5) and User Studies (Section 7)

Sim / User Study	Hidden Layers	Hidden Layer Size[1]	Activation Functions[2]	LR
1-DoF Sim	2	8, 8	Tanh	0.01
2-DoF Sim	2	16, 64	ReLU, Tanh	0.001
3-DoF Sim	3	36, 96	ReLU, Tanh	0.001
1-DoF User Study	2	8, 16	Tanh	0.01
2-DoF User Study	2	16, 32	Tanh	0.01
3-DoF User Study	3	36, 64	Tanh	0.01
Highway Sim	3	36, 64	ReLU, Tanh, Sigmoid	0.001
10-DoF Sim	2	64, 128	ReLU, Tanh	0.00025

Note that although the decoder networks used in this study were MLPs, other architectures could be used (such as gated recurrent units (GRUs)). Further, it is possible to use more exotic structures for LIMIT (such as Bayesian networks), but feed-forward networks were easier for us to use and demonstrate. For implementation specifics, see our Github repository.

C Additional Simulations

C.1 Scenarios Requiring Near-Immediate Signal Adaptation

LIMIT is intended for settings where the human repeatedly interacts with an interface, and the interface can tune its signal over time. This may result in the system being less efficient within situations where the user requires near-immediate understanding of the interface's signals (e.g., understanding the interface at the first interaction). To better understand how quickly LIMIT personalizes its signals, we have conducted additional simulations using the same environment as Section 5.2. We first pretrained an instance of LIMIT while interacting with a simulated human. We then sampled a random state-hidden information pair (s, θ) and plotted the signal x LIMIT learned with the baseline human. Figure 12 (Top) shows the initial signal.

We next paired two copies of the pretrained LIMIT with two different simulated humans. Each of these simulated humans had a different policy than the baseline; the new human policies were

[1] The first number listed refers to the human and interface networks, the second number listed refers to the size of the decoder's hidden layers; these are much larger than the human and interface models.

[2] Tanh was always used on the interface neural network to restrict the signal output. Whenever another activation function is listed, it was used for the decoder and human model networks. For more detail, see our Github repository.

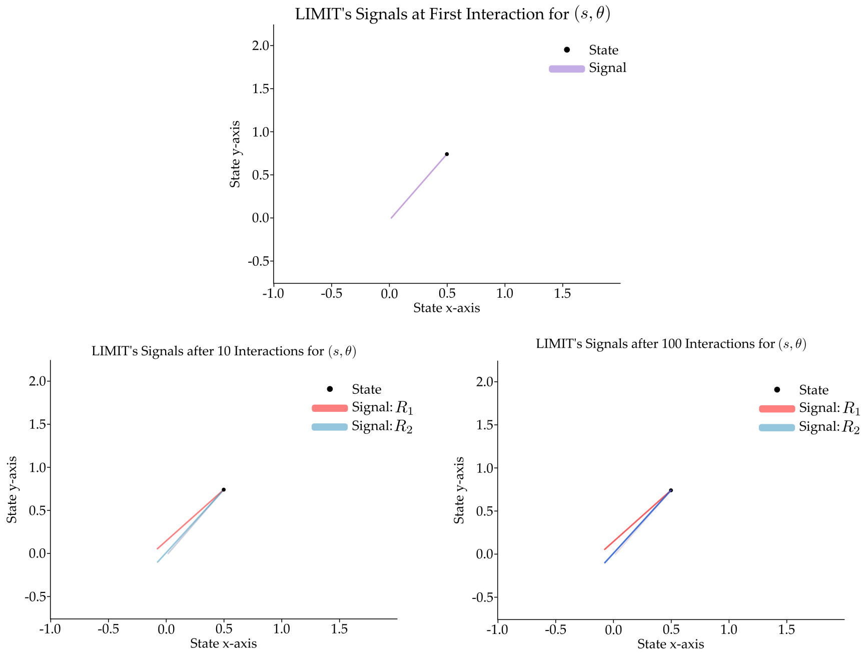

Fig. 12. The change in LIMIT's signals over repeated interaction for different human policies in a 2-DoF environment. (Top) The initial signal for the first human model $\mathcal{H}_0$ for (s, θ), pretrained for one interaction. (Bottom Left) The signals *plotted for each interaction* for the same instance of LIMIT, trained with new human models $\mathcal{H}_1$ and $\mathcal{H}_2$ for 10 interactions. Note that the signals are different than the first figure, and that the signals change slightly over repeated interaction as LIMIT adjusts to the new users' behavior. (Bottom Right) Signals produced by LIMIT for $\mathcal{H}_1, \mathcal{H}_2$ at (s, θ) after 100 interactions. Note that after pretraining, $\mathcal{R}_1$ and $\mathcal{R}_2$ are separate instances of LIMIT *trained with different replay memory buffers and different optimizers.*

randomly sampled. We then plotted the adapted signals that LIMIT learned after 10 interactions and 100 interactions with both of the new simulated humans. See Figure 12 (Bottom). Comparing the plots, we conclude that the interface's learned signals have converged within 10 interactions: the signals the interface sends at 10 interactions are the same signals the interface uses after 100 interactions. Overall, Figure 12 suggests that LIMIT can personalize to new users in less than 10 interactions.

C.2 Variance in LIMIT's Signals Between Users

To investigate whether LIMIT produced different signals for different users, we conducted an additional simulation, similar to that of Appendix C.1. Here, we pretrained an instance of LIMIT $\mathcal{R}_0$ on a human model $\mathcal{H}_0$ in our standard 2-DoF environment described in Section 5.2 (matching the "Lights" task from our in-person user studies). After pretraining LIMIT for several interactions with $\mathcal{H}_0$, we duplicated the instance of LIMIT and paired each with two new human agents with distinct policies $\mathcal{H}_1$ and $\mathcal{H}_2$. Then, after each interaction, we plotted the signals observed for each instance of LIMIT ($\mathcal{R}_1$ and $\mathcal{R}_2$) for a particular (s, θ). Figure 13 shows this clearly: LIMIT adjusts its signals for new users over repeated interaction, accommodating their distinct behaviors.

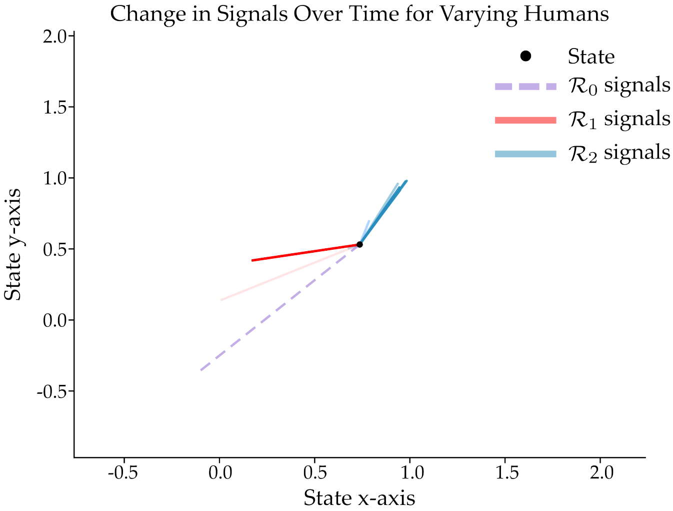

Fig. 13. The change in signals over time for new humans $\mathcal{H}_1$ and $\mathcal{H}_2$ is clearly shown for an arbitrary yet particular (s, θ). Here, the signals are shown as vectors pointing away from the "state" point. Note that the signals' x- and y-axis correspond with the axis of the plot, i.e. a vector pointing along the x-axis with one unit of length would correspond to a signal of $\begin{bmatrix}1 & 0\end{bmatrix}^T$.