Towards Robots that Learn from Humans

Dylan P. Losey
Department of Mechanical Engineering
Virginia Tech
losey@vt.edu

***Abstract*—How can we make robots that learn new tasks from human teachers? Most research attempts to answer this question by building and testing new imitation learning algorithms. But we believe that the *human* is equally important: whatever algorithms we develop should stem from how humans teach and interact with robotic systems. This writeup summarizes the insights our group has gained over the last five years by placing humans at the center of the learning problem. We organize our research along three themes, where each theme explores an underlying principle necessary to learn from humans. 1) *Learning as control*, where we inject structure to align robot learners with how humans teach, 2) *learning as representation*, where we enable man and machine to speak the same language, and 3) *learning as communication*, where we close the learning loop by providing feedback to the human teacher. Viewed together, these interconnected research directions at the intersection of human-robot interaction advance learning from humans in ways that go beyond learning algorithms. Our papers are available at: https://collab.me.vt.edu/**

I. Introduction

Robots should be able to learn new tasks from human teachers. At its core, this adaptability is the promise of robotics: machines not built for a single purpose, but able to intelligently assist humans throughout our everyday lives. We envision robots that observe how people behave, and then (based on our guidance, examples, and feedback) these robots learn how to make toast, drive us to work, or assemble new parts.

Unfortunately — after more than 30 years of focused research — learning from humans (i.e., imitation learning) still remains an unsolved question. So what's holding us back? Why haven't we achieved this functionality? At first glance, it might seem like the algorithm the robot uses to learn from the human is the key; i.e., if we can just find the right combination of data, architecture, and loss, then we will "solve" imitation learning. To be sure, the learning algorithm is important. But the human is equally important; whatever algorithms we develop need to inherently account for how humans teach and interact with robotic systems. This includes deficiencies the learner must overcome (e.g., limited amounts of imperfect human data), capabilities that the learner should tap into (e.g., the human's insight on how to complete the task), and discrepancies between man and machine (e.g., how the human and robot communicate).

Our work over the past five years has tried to place the human at the center of imitation learning problems. This formulation required a shift in approach. Instead of starting with learning algorithms — and then designing new architectures, losses, or heuristics to enhance their capabilities — we explored the underlying principles necessary to successfully learn from humans. For example: how can humans convey their desired task? How should robots represent and interpret the human's inputs? And what learning systems will converge towards the correct behavior when paired with a human teacher? If we can answer these and other fundamental questions, we can apply the resulting principles to build classes of algorithms and interfaces that successfully imitate everyday users. In what follows we summarize our progress across three main themes (see Figure 1).

Learning as Control (Section III). The way the robot learns should be aligned with how the human teaches. To reach this alignment we need structure; specifically, our work frames learning as a dynamical system. The robot's policy is the state of this system, the human's examples are inputs, and the learning rule governs how the robot's policy updates in response to those inputs. Analyzing learning with control-theoretic tools enables us to shape where the learning rule will converge (i.e., the equilibrium of the dynamical system). We find that this approach is invariant to the specific modality the human uses to teach their desired task. In addition, we can shape the learning system to converge towards control policies that have desirable properties: e.g., producing behaviors that align with human expectations. Our research outcomes include paradigms that modify a variety of existing learning algorithms so that they are robust to noisy human examples and generalize to situations not shown during training.

Learning as Representation (Section IV). When humans and robots interact, learning becomes a problem of understanding the other agent. How can robots enable humans to directly indicate what the robot should do, while also guiding the human towards contexts where the robot can more effectively collaborate? To facilitate this mutual, bidirectional understanding between man and machine we leverage representations; i.e., compact models that encode the other agent in a way that the ego agent can harness. We then equip both humans and robots with these representations. Our research shows that humans can interact with representations of the

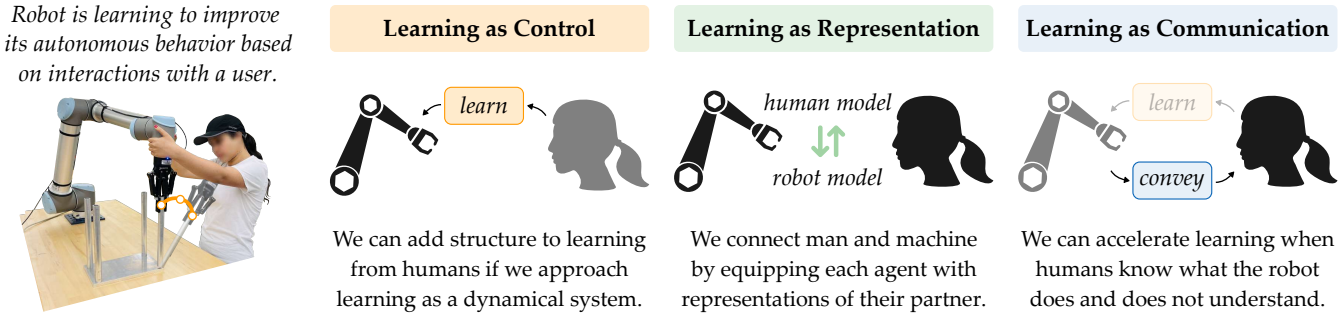

Fig. 1. (Left) Our group's research over the last five years seeks to place the human at the center of imitation learning problems. (Right) This human-centered approach leads to three interconnected formalisms for robot learning that span control, representation, and communication. (Bottom) Our resulting insights advance learning from humans in ways that go beyond simply improving the robot's learning algorithm.

robot's actions, policies, and features to intuitively control how they want the robot to behave. In parallel, robots can plan across learned representations of humans to anticipate human actions and guide co-adaptation towards mutually beneficial behaviors.

Learning as Communication (Section V). Finally, from the human's perspective we frame learning as communication. Humans can teach proficient policies if they understand how the robot learns. But today's robots are black boxes; as the human teaches, it is not clear what the robot has learned or when it is still uncertain. Communicating the robot's understanding closes this learning loop and fundamentally improves human teaching (i.e., the human can now focus on aspects of the task where the robot is struggling). To decide when, how, and what to communicate we develop an information-theoretic framework that correlates the human's actions with the robot's task understanding. Applying this framework results in robots that boil down complex, abstract learning parameters into personalized, understandable feedback. Our research outcomes include methods for mapping the robot's internal state (e.g., its learned policy) to both implicit and explicit signals that everyday human users can leverage to adjust their teaching.

The overall purpose of this paper is to summarize the interconnected work done by our group, and to organize our experiences and insights for creating robots that robustly, efficiently, and intuitively learn new tasks from human teachers. When viewed together, the three themes listed above provide complementary perspectives and underlying principles for learning from humans. Although we cannot claim to have "solved" imitation learning, we hypothesize that our insights across all three themes reach towards a solution in ways that methods which only consider (for example) the robot's learning algorithm cannot achieve. We emphasize that none of this research would be possible without prior scholarship and parallel efforts; please see related works in the cited papers for further details.

II. Learning from Humans

We broadly consider scenarios where a robot is learning to improve its autonomous behavior based on interactions with a human. Let x be the system *state*. Depending on the specific problem setting, this state could contain multiple components: e.g., the human's position, the robot's position, and relevant features in the environment. Let u be the system *action*. Again, this overall action could break down into the human's action $u_{\mathcal{H}}$ and the robot's action $u_{\mathcal{R}}$. Both inputs cause the system to transition according to its *state dynamics*: $x' = f(x, u)$, where x' is the new system state.

During interaction the robot decides which actions it should take based on a *control policy* π:

$$u_{\mathcal{R}} = \pi_\theta(x) \tag{1}$$

This controller is a mapping from states to actions, and is parameterized by θ. When learning from humans we often instantiate π_θ as a neural network — where θ forms the weights of that network.

Our goal is for the robot to learn the correct control policy; i.e., a control policy that will assist the human and autonomously perform desired tasks. More specifically, the robot seeks to identify a set of weights θ that maximizes its expected performance. Within imitation learning settings the robot extracts these weights by reasoning over examples from a human teacher. Let $\mathcal{D} = \{(x, u)\}$ be a *dataset* of state-action pairs where the human demonstrates how the robot should behave (offline), or provides feedback about the robot's current policy (online). Based on this dataset, the robot updates its control policy π_θ to imitate the human teacher.

III. Learning as Control

Control policies instantiated as neural networks are a powerful tool because of their ability to learn arbitrarily complex decisions. But one of the frustrating things about these neural control policies is that they are inherently open-ended. As designers, we do not know what the robot will learn, or what we can guarantee about the policy's performance. This obscurity leads to empirical

Fig. 2. **Learning as control**. (Left) We frame learning as a dynamical system, and shape the control policy π_θ to converge towards desirable equilibrium. (Middle) Robot playing air hockey. Each round the puck moves at different speeds and angles, but the robot can continuously hit that puck because our control-theoretic approach makes it robust to covariate shift. (Right) One of the ways human teachers can provide inputs to robot learners. The robot records a video of the human performing the task, and then matches object motions to complete the same task.

development cycles, where designers make changes to improve the algorithm's behavior, but it is not always clear *why* those changes are beneficial.

To help resolve this problem, our insight is that:

We can add structure to learning from humans if we approach learning as a dynamical system.

Under this re-framing the robot's control policy is treated as a "state," and the learning rule becomes its "dynamics." The equilibrium of the dynamical learning system should be the space of policies the human is trying to teach the robot. The human's inputs — of whatever modality, whether demonstrations, corrections, or preferences — ideally drive the state (control policy) towards its equilibrium (space of desired control policies). But learning from humans has inherent challenges: human teachers usually provide small amounts of data, and that data is generally noisy and often worse than the policy they want the robot to learn. Hence, we need to shape the dynamical learning system so that it (a) quickly converges to what the human meant (b) despite heterogeneous and suboptimal inputs.

To summarize: we are ultimately looking for performant learning algorithms. By treating learning as control, we gain structure and can apply control-theoretic tools to understand *why* specific learning methods work. See Figure 2 for examples of these concepts.

Control-Theoretic Analysis. Our recent works have explored two levels of dynamical learning systems. To explain these methods, we first need to take a step back and talk about how robots learn. Learning from humans often revolves around a *loss function*; this loss quantifies the performance of the control policy across its dataset. In practice, we train the control policy to reach weights θ that minimize the given loss function $\mathcal{L}(\theta)$. As a simple example, for behavior cloning the loss could be:

$$\mathcal{L}(\theta) = \sum_{(x,u)\in\mathcal{D}} \|u - \pi_\theta(x)\|^2 \tag{2}$$

Intuitively, Equation (2) encourages the robot to learn a policy that matches the demonstrated actions for each state in the dataset. But — as we will see — there are many reasonable choices for the loss function, and we can leverage control theory to shape each of these potential losses for more robust and stable learning.

We now return to our approaches that frame learning as a dynamical system. First, we consider the error in the control parameters θ [23]. If θ^* are the unknown parameters of the correct control policy, then let $e_\theta = \theta - \theta^*$ denote the difference between these true parameters and the weights the robot has actually learned. For learning algorithms that use gradient descent to update θ so that it minimizes loss $\mathcal{L}(\theta)$, the error evolves according to:

$$e'_\theta = e_\theta - \alpha\nabla_\theta\mathcal{L}(\theta), \quad e_\theta = \theta - \theta^* \tag{3}$$

where $\alpha > 0$ is the scalar learning rate. The equilibrium of the dynamical system in Equation (3) should ideally be $\|e_\theta\| = 0$, i.e., the robot should learn the desired control policy with weights $\theta = \theta^*$. By applying Lyapunov stability analysis, we find that a sufficient condition for convergence to this equilibrium is:

$$\alpha\|\nabla_\theta\mathcal{L}(\theta)\|^2 - 2e_\theta \cdot \nabla_\theta\mathcal{L}(\theta) < 0 \tag{4}$$

If Equation (4) is satisfied, then $\|e'_\theta\| < \|e_\theta\|$ and the error moves towards zero. This control-theoretic property tells us something fundamental about the design of the loss function. Specifically, we want to shape loss $\mathcal{L}$ so that Equation (4) holds for as wide a range of human inputs as possible. Doing so expands the basins of attraction, and enables the robot to learn the correct control policy from small amounts of noisy human data.

A second way we can frame learning as a dynamical system is to focus on errors in the state x [24]. During training, the human teacher shows the robot how to behave at states x in dataset $\mathcal{D}$ — the robot knows exactly what to do at these states. But at test time the robot will inevitably encounter new situations that are outside of the training dataset; it is critical that the robot's behavior remains safe and assistive across these novel scenarios. Let $\hat{x}$ denote a new, previously unseen state. The error (i.e., the *covariate shift*) between new and seen states is $e_x = \hat{x} - x$. Applying our system dynamics, this error evolves according to:

$$e'_x = f(\hat{x}, \hat{u}) - f(x, u), \quad e_x = \hat{x} - x \tag{5}$$

Substituting control policy π into Equation (5), and applying a first order Taylor Series approximation around $x \in \mathcal{D}$, we reach the locally linearized simplification:

$$e'_x = (\nabla_x f + \nabla_u \cdot \nabla_x \pi) e_x \tag{6}$$

What Equation (6) tells us is how a robot that starts at a new state $\hat{x}$ will behave as compared to a robot that starts at a known state x. Of course, these two robots should not act in the exact same way. But we want to prevent the robot starting at $\hat{x}$ from drifting completely out-of-distribution, and reaching scenarios where the control policy has no clue which actions to take. Hence, we will encourage the robot to remain close to examples the human teacher has demonstrated by making the error dynamics in Equation (6) stable about the equilibrium $e_x = 0$. This is achieved by — for example — shaping $\nabla_x f + \nabla_u \cdot \nabla_x \pi$ to be a stable matrix. Again, our control-theoretic finding has implications for the design of loss functions when learning from humans; we can select or modify $\mathcal{L}$ to enforce the stable error dynamics and make the learned policy robust to covariate shift.

Both of the examples listed above are just that: examples. Across our research we have tried different ways of framing learning as control [10, 31, 8]. The advantage of these formulations is that they allow for control-theoretic analysis, which can then be leveraged to derive learning algorithms (or classes of learning algorithms). Within this framework we often end up thinking about *equilibria* — i.e., what sorts of behaviors should the robot learner be converging to? Of course, we do not know the correct control policy *a priori*; but we do have some strong priors (e.g., the robot should not learn to collide with obstacles). Our recent works have translated these priors into desirable properties in θ, essentially reducing the range of equilibria $\theta \in \Theta$ the learning system can recover from the human teacher.

Various Types of Inputs. Whatever structure we impose must be able to assimilate the different ways in which humans can teach robots. In the context of our dynamical system, the loss function and associated learning rule shape how the robot learns. The human's actions — i.e., their teaching behaviors — become the *inputs* into this dynamical system. So how should the human teacher provide the examples which compose training dataset $\mathcal{D}$ (i.e., the inputs from which the robot learns)? Our goal here is to collect information-rich teaching data while making it as easy as possible for the human to input that data. If we can make human teaching seamless and holistic, then it is possible to collect large amounts of high-quality examples — and more human oversight translates to simpler robot learning problems.

We have therefore developed a suite of inputs that humans can use during teaching. Since treating learning as control is invariant to the type of input, designers can switch between these options to find the input method(s) that are best suited to their task, user, and scenario:

- *Demonstrations* [19]. Human teleoperates or physically guides the robot through the desired motion.
- *Corrections* [19]. Human intervenes online to modify a specific part of the robot's behavior.
- *Preferences* [19]. Human compares two or more robot trajectories and ranks these options.
- *Drawings* [25]. Human sketches and annotates the desired behavior on a 2D image of the scene.
- *Videos* [14]. Robot collects a video of the human performing the task with their own body.
- *Language* [4]. Human describes what the robot should be doing using natural language.

Effective learning often combines more than one input modality. Learning from these different modalities (e.g., both drawings and preferences) can be tricky since they are conveying different information. But we have found that robots can seamlessly reason over multiple input types by treating them all as *comparisons*. When giving an input the human is choosing a specific option $u_{\mathcal{H}}$. By comparing that choice to the counterfactuals (i.e., the alternatives $\tilde{u}_{\mathcal{H}}$ the human did not select), the robot gains an understanding of what the human is optimizing for, regardless of the input modality. For instance: the control policy learned by our dynamical system should satisfy the comparison $\|u_{\mathcal{H}} - \pi_\theta(x)\| < \|\tilde{u}_{\mathcal{H}} - \pi_\theta(x)\|$. We note that both drawings and videos can be grounded into demonstrations. For drawings, we project the 2D sketch into a 3D motion. For videos, the robot first tries to copy the behaviors the human showed. This inevitably fails — since the robot's kinematics are different from the human's — but the robot can iteratively explore around the initial waypoints to find a trajectory that matches how the human interacted with objects [22].

Outcomes. Overall, we have conducted research that frames learning as a dynamical system [23, 24, 8], research on the equilibrium the system should converge towards [10, 31], and research on how to input different types of human teaching into the system [19, 25, 14, 22].

IV. Learning as Representation

The learning paradigms described in Section III are based on inference. A human teacher shows instances of the task (e.g., drawings, videos), and the robot tries to guess the correct control policy from those examples. Of course, inference is challenging — so why not let the human directly specify what the robot should do? Imagine that we can enable humans and robots to *speak the same language*. If we find this language, then humans can leverage it to directly tell the robot its task, and robots can guide humans towards seamless collaboration.

To help construct this language, our insight is that:

We connect man and machine by equipping each agent with representations of their partner.

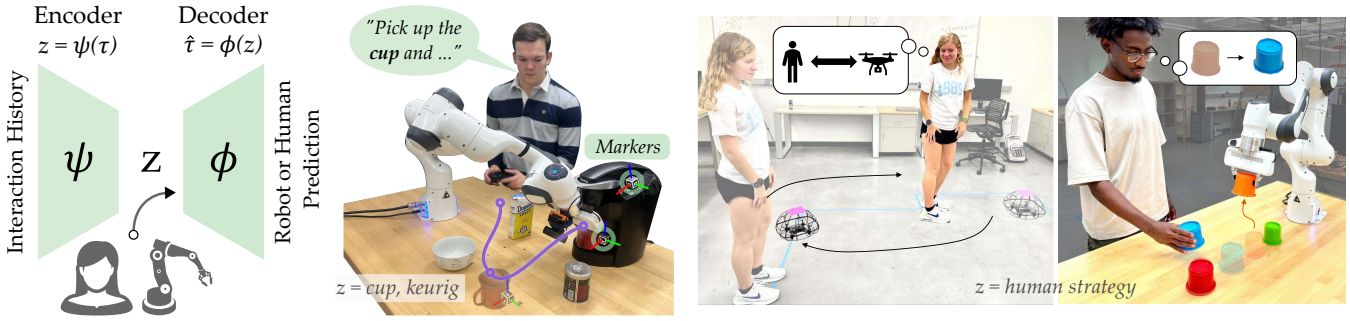

Fig. 3. Learning as representation. (Left) We use representations to enable humans and robots to directly interface with one another. The latent z encodes one agent, and then we equip the other agent with that variable. (Middle) *Robot representation*. The robot embeds its observations to a latent feature space, and the human teacher uses language and markers to indicate which features — in this case the cup and coffee maker — the robot should focus on in order to learn the current task. (Right) *Human representation*. Drones and robot arms predict how the human will behave, and then plan across that prediction to influence the human's future actions (e.g., changing the human's strategy).

Representations here refer to learned models. They embed some aspect of the other agent in a way that the ego agent can harness; e.g., we can use representations to embed the complex, interconnected motions of robot arms into a low-dimensional, intuitive action space. Framing learning as representation is really a two stage process. In the first stage, we extract a representation from data, and then in the second stage, we enable the ego agent to interact with the learned representation. So, for example, if we embed the high-dimensional motions of robot arms into joystick inputs, then humans can interface with that joystick to control the robot's behavior. Alternatively, if we embed the human's driving style into a latent vector, then autonomous cars can plan across these styles to influence the human's actions. Through representations humans and robots speak with one another and remove the need for inferring tasks from humans.

To summarize: we are looking for intuitive and expressive representations in human-robot interaction. If we view learning through these representations, we enable agents to directly specify *what* each other should do. See Figure 3 for examples of these concepts.

Formalizing Representations. Mathematically, our representations introduce an intermediate variable (also called a *latent* variable) to the control policies discussed in the previous sections. Instead of learning a mapping $\pi_\theta(x)$, we now instantiate the control policy as:

$$u_{\mathcal{R}} = \pi_\theta(x, z) \tag{7}$$

where z is the representation. For simplicity, let's say this representation is learned through an *encoder* and *decoder*. The encoder $\psi(\tau) \rightarrow z$ embeds observed data into the learned latent space, and the decoder $\phi(z) \rightarrow \hat{\tau}$ maps that latent variable back into observable data. What z represents depends on both what we are compressing, τ, and what we are trying to recover, $\hat{\tau}$. Across our research τ contains information that is mutually known by the human and robot, such as states, trajectories, and histories of interaction. Within $\hat{\tau}$ we then try to recover either the robot's next actions (to build a representation of the robot), or the human's next actions (to build a representation of the human). In either case, once representation z is learned via the encoder and decoder it provides additional context to the control policy in Equation (7). Put another way, based on z the robot has a better idea of what actions it should take.

Robot Representations. Let's start with our work that builds representations of the *robot* and provides these representations to the human teacher. The core idea here is to find latent spaces that enable humans to more directly convey their desired behavior to robots. Since the human will interface with the robot representations, these representations should be user-friendly: e.g., low-dimensional, intuitive, and consistent.

We begin by representing the robot's ***actions*** [18, 20]. Robot arms are composed of multiple connected joints, and when performing a task the robot must carefully synchronize all its joint positions. To give humans direct control over this coordinated motion, our work embeds the robot's high-dimensional actions $u_{\mathcal{R}}$ into a low-dimensional representation z. We train this representation to accurately reconstruct complex robot actions conditioned on the system state. Referring back to our encoder and decoder, here input τ is the action $u_{\mathcal{R}}$, output $\hat{\tau}$ is the reconstructed action $\hat{u}_{\mathcal{R}}$, and latent z is trained to minimize reconstruction error: $\|u_{\mathcal{R}} - \phi(\psi(u_{\mathcal{R}}, x), x)\|$. Once the latent variable is learned, humans can directly teleoperate the robot by selecting $z = u_{\mathcal{H}}$ through a joystick interface. This enables users to control robot arms in real-time; the robot converts the human's simple inputs (e.g., pressing right on the joystick) into coordinated joint motions (e.g., pulling open a drawer).

But if we limit our representation to individual actions, users have to constantly give inputs throughout the task. To facilitate more fundamental communication we have therefore explored multiple levels of abstraction for robot representations. For example, in [11, 13] we temporally extend actions into ***trajectories***, and use the human's

inputs to determine which task the robot should autonomously perform. In these works τ is a sequence of states and human inputs, and $\hat{\tau}$ are the future actions the robot arm should take. The system operates using the same principle as with latent actions: the robot maps the human's inputs into coordinated joint motions. But now — instead of only selecting a single action — the robot can recognize the larger task, and execute a series of actions to help complete that task.

Both action and trajectory representations offer a way for users to indicate what their robot should do. When learning from humans, it is also critical to know why the human is making each decision. For instance, to autonomously make coffee, the robot must learn to focus on the pose of the coffee cup and the coffee machine. Our recent works enable humans to convey the logic behind their decisions through ***feature*** representations [35, 4]. The robot embeds every environment feature it observes (e.g., the color of the cup, the clutter on the table) into a latent feature space. While providing demonstrations, the human teacher leverages natural language and physical markers to indicate which element(s) of this latent space are critical for the current task. Returning to the encoder and decoder: now τ is the robot's visual observation and its associated features, z is the latent feature space, and $\hat{\tau}$ is the language and marker data provided by the human teacher. Intuitively, latent features are a way for teachers to direct the learner's attention and remove confusion behind control decisions.

Our final type of robot representation is our most abstract — representing the robot's policy as a ***canonical space*** that is shared across multiple tasks [27]. Here we encode demonstrations along two axis: a discrete encoder $\psi_1(\tau)$ that embeds trajectories into different tasks, and a continuous encoder $\psi_2(\tau)$ that embeds those same trajectories into continuous styles. Combining these embeddings produces a latent manifold we refer to as the canonical space. Humans can interface with this space (e.g., clicking on a visualization) to select what task the robot should perform and the style with which the robot should complete that task. The decoder maps both human selections into robot actions $\hat{\tau}$ that are consistent with the user-chosen task and style.

Human Representations. Switching perspectives, we next look at works where robots build and reason over representations of the *human*. These representations are effectively human models: they enable robots to anticipate how humans will respond during interaction. The core idea here is that — if the robot can predict how the user will react to its actions — the robot can optimize its own behavior to guide the human towards synergies.

Our early works apply this paradigm to human and robot ***co-adaptation*** [29, 28]. The encoder $\psi(\tau) \rightarrow z$ inputs a history of states, actions, and rewards, and the decoder $\phi(z) \rightarrow \hat{\tau}$ predicts the rewards the robot will receive when executing the policy $\pi_\theta(x, z)$. Under this framework z captures how the human interacts with the robot: different values of z represent different strategies the human might follow (e.g., driving aggressively or defensively). At each new interaction the robot updates its estimate of z based on what happened during the previous interaction, and then reasons over z to select motions that are adapted to the human's new strategy.

Our recent works go beyond adaptation to ***influence*** humans [32, 34]. Here we use the same encoder structure as before, but now the decoder predicts the human's future actions. Robots plan across this prediction to find which robot trajectories will result in human responses that maximize the team's performance. Put another way, the robot leverages its human representation to influence the user towards collaborative behaviors (e.g., causing people to drive safely). When learning from humans this influence is particularly useful because the robot can lead the human towards interactive behaviors it knows how to respond to — thereby increasing its effectiveness. Of course, human representations are never perfect. To account for these modeling errors during interaction we have also explored robust safety approaches [2].

Outcomes. Overall, we have built robot representations which humans can interface with to directly control the robot's actions, trajectories, features, tasks, and styles [18, 20, 11, 13, 35, 4, 27]. We have also built human representations that robots can plan across to co-adapt or influence human-robot teams [32, 34, 29, 28, 2].

V. Learning as Communication

Sections III and IV focused on making the robot a better learner. In this final theme we now explore the opposite perspective, and focus on making the human a better teacher. This direction is critical for robots in-the-wild: everyday users will not know the tricks and insights that roboticists leverage when training systems (e.g., end-users may be unsure about why their robot is failing to learn). But if we can guide everyday humans to be expert teachers, then even robots with simple learning algorithms will extract the correct control policy.

To help humans become teachers, our insight is that:

We can accelerate learning when humans know what the robot does and does not understand.

Imagine a user training their robot to make coffee. If the human knows the robot is unsure about where to place coffee cups, then they can focus their teaching on that specific part of the task (e.g., giving more demonstrations that reach for cups). This seems simple enough: we just need robots to communicate their learned policy to the human. But control policies are parameterized by thousands of weights θ, and we cannot explain each of these weights — or what they mean — to the human teacher. Learning policies instantiated as neural networks effectively turns robots into black boxes. Framing learning as

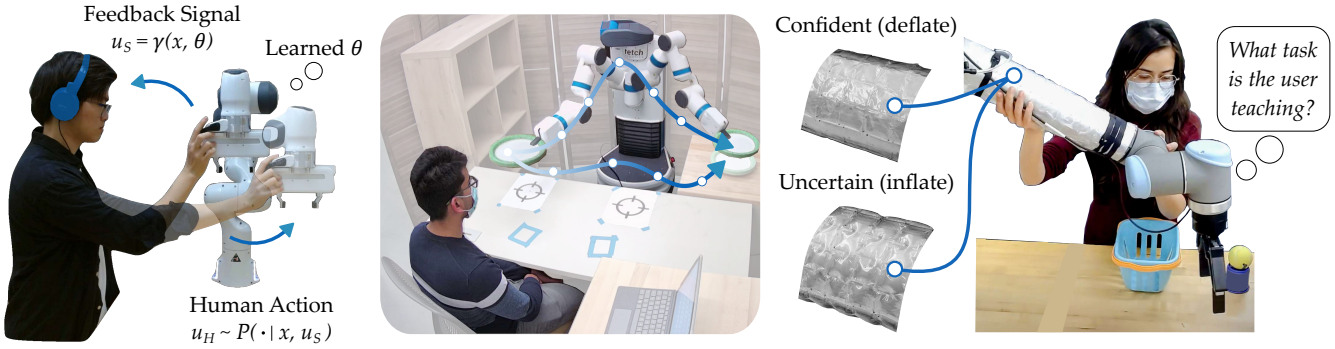

Fig. 4. **Learning as communication**. (Left) We communicate abstract and high-dimensional robot learning by selecting feedback signals u_S that correlate the human's actions u_H with the robot's learned weights θ. (Middle) Applying this information-theoretic framework, we can modulate the robot's motion to *implicitly* convey learning. Here a robot asks the human which trajectory they prefer: by exaggerating the height disparity, the robot makes it clear that it's uncertain about the correct height. (Right) Robots can also leverage our framework to provide *explicit* visual, language, or haptic feedback. Here a robot changes the pressure of haptic arrays to indicate its understanding of the human's teaching.

communication seeks to open these boxes: we develop feedback signals that *close the loop* and convey what the robot has learned back to the human teacher. Our core premise here is that — if we can determine when, where, and what feedback to provide — the human will align their teaching with the robot's learning, fundamentally simplifying learning from humans.

To summarize: we are looking for interfaces and algorithms that convey robot learning. By treating learning as communication, we reduce the burden on the robot learner and build a more seamless student-teacher team. See Figure 4 for examples of these concepts.

Mapping Learning to Communication. Let's extend the notation from Section II. In addition to taking actions $u_{\mathcal{R}}$, now the communicative robot provides *feedback signals* $u_{\mathcal{S}}$. Across our research these signals $u_{\mathcal{S}}$ include visual, haptic, or auditory cues (e.g., a wristband which vibrates at different frequencies). The robot chooses its signals (e.g., the vibration frequency) according to a *communication policy* $\gamma(x, \theta) \rightarrow u_{\mathcal{S}}$ that maps learned control weights θ into displayed cues $u_{\mathcal{S}}$. The standard approach for γ is to design a library of signals and then prescribe a meaning to each option (e.g., high-frequency vibrations mean the coffee-making robot is unsure about where to grasp the coffee mug). However, this pre-defined and static framework is incompatible with robot learning, where the types of information the robot needs to convey are inherently abstract and constantly changing.

Our works therefore adapt the communication policy γ during interaction to find cues which clearly convey robot learning back to the human teacher. For our running example: when teaching a robot to make coffee, perhaps high-frequency vibrations should indicate confusion about grasps. But when teaching a robot to make toast, that same signal could now indicate uncertainty about how to insert the bread. To identify and update the mapping between learning and signals, we leverage an information-theoretic approach [1, 3, 38]. Specifically, we train γ to maximize the correlation between the human's actions $u_{\mathcal{H}}$ and the robot's parameters θ:

$$I(u_{\mathcal{H}}; \theta \mid x) = H(u_{\mathcal{H}} \mid x) - H(u_{\mathcal{H}} \mid x, \theta) \tag{8}$$

Here I is the mutual information between two variables, and H is the Shannon entropy over a distribution. The hypothesis behind Equation (8) is that — even if we don't know what the human is trying to teach — we do know that the way the human teaches should vary in response to what the robot has learned. In practice, we find that training γ to maximize Equation (8) leads to robots that *personalize* when, how, and what they communicate. This is because I reasons over $P(u_{\mathcal{H}} \mid x, u_{\mathcal{S}})$, the probability of the current human teacher taking an action in response to the robot's feedback $u_{\mathcal{S}}$. Using this information-theoretic framework we can build communication mappings γ completely from scratch over repeated interactions [1], or we can apply existing human models P to bootstrap personalization [3].

Conveying Learning via Explicit and Implicit Signals. Equipped with the information-theoretic backbone from Equation (8), we next developed interfaces and algorithms that communicate robot learning. This communication has become an emerging area of research. For a review of current directions and open questions, see our survey paper on keeping humans in-the-loop [7].

We start with work that leverages ***explicit*** feedback signals (e.g., a text description of what the robot is thinking). Remember the user is trying to teach the robot; to be an effective teacher, the human must focus on their own demonstrations and the robot's reactions. As such, explicit feedback should not interrupt the human's attention. One approach for unobtrusive signals is *user-worn* devices — such as augmented reality headsets — that overlay the robot's plan onto the environment [26]. But our studies suggest that *robot-mounted* displays are better suited for learning from humans. For instance, in [36, 37] we wrap haptic arrays around robot arms:

Fig. 5. Applications. (Left) RISO grippers unify traditional rigid end-effectors with a novel class of soft adhesives. When grasping an object, RISOs can either pinch the item between non-deformable fingers or cause the object to stick to soft surfaces. We've used RISOs to grasp objects from 2 mg (a piece of lead) to 2 kg (a stack of papers). (Right) The Kiri-Spoon is a soft, shape-changing utensil specifically designed for robot-assisted feeding. Robots using Kiri-Spoons are able to more easily grasp and hold food items, simplifying acquisition and reducing spills.

when users teach the robot via physical demonstrations, they can feel these haptic displays inflating or deflating beneath their hands. We leverage the haptic arrays to render the robot's confidence. Specifically, we train an ensemble of k control policies $\pi_\theta(x)$, and then query these policies at the current state x. If all k policies output similar actions, the robot is confident and we deflate the haptic bags. But if the policies diverge, we inflate the array to notify the human about the learner's uncertainty. In [38] we extend this pneumatic display into modular hardware that users can reconfigure on-the-fly. Each configuration provides a different set of haptic cues, and — to determine which configuration is best suited for the current task and user — we maximize the expected information gain from Equation (8).

In parallel to explicit feedback, we have also advanced how robots ***implicitly*** reveal learning through natural motions. Take, for example, a robot performing kitchen tasks. By reaching for a ladle, that robot is implicitly communicating what sorts of behaviors it has learned to complete (e.g., scooping). Our works formalize this logic in the context of shared autonomy [12, 9], where the robot arm uses actions $u_\mathcal{R}$ to indicate a learned task, and in the context of preference learning [6], where the robot adjusts its questions to highlight regions of uncertainty. Implicit communication introduces a fundamental trade-off for robot motions. The robot is using one channel (i.e., its actions $u_\mathcal{R}$) to simultaneously perform the task and reveal its learning. To balance efficiency and communication, we accordingly propose a constrained optimization framework [12] where the robot maximizes its action transparency subject to performance limits.

With both explicit and implicit communication signals in place, we can now return to our original hypothesis. Does communication actually improve human teaching and robot learning? By and large, our experiments with ***human-robot teams*** support the role of communication, and suggest that this feedback enhances trust, team dynamics, and mutual understanding [5, 30]. Interestingly, we have also discovered that there are certain edge cases where communication is not helpful [33]. Conveying robot learning requires effort: both from the designer (who needs to create the communication protocol) and from the end-user (who needs to interpret and respond to the robot's signals). Our game-theoretic analysis shows that communication is not worth the cost when either (a) the potential improvements in performance are negligible or (b) the human struggles to understand what the robot is trying to convey.

Outcomes. Overall, we have developed a mathematical framework for mapping robot learning to communication [1, 3, 38]. We have then applied different instantiations of that framework to generate explicit and implicit feedback signals [26, 36, 37, 12, 9, 6], and researched their effect on the human-robot team [7, 5, 30, 33].

VI. Discussion

This writeup synthesizes how our research group is moving towards robots that learn from humans. Conventional wisdom starts with the learning algorithm. Instead, our research tries to place the *user* at the center of the learning problem by asking fundamental questions:

1) *How do we structure learning for humans?* This question led to our theme on **Learning as Control**. Our current answer is to model learning as a dynamical system, and apply control-theoretic tools to derive learning principles based on convergence, robustness, and heterogeneous inputs.
2) *How do we enable humans to clearly convey their task?* This question led to our theme on **Learning as Representation**. Our current answer is to build representations of the human and robot, and then enable agents to directly interface with one another through these latent representations.
3) *How do we make humans better teachers?* This question led to our theme on **Learning as Communication**. Our current answer is to close the loop on robot learning, and develop information-theoretic feedback signals that intuitively convey what the robot is learning back to the human teacher.

Taken together, these (and other) different perspectives across the *robot, interaction, and human* advance learning from humans in ways that a single perspective cannot achieve. For example: imitation learning methodologies should ideally improve both how the robot learns as well as how the human teaches. Ultimately, the goal of learning from humans remains the same — enabling multi-purpose robots to assist everyday users on new tasks. By placing the human at the center of the learning process, we maximize how much the robot learner assists while minimizing the human's teaching effort.

Looking Ahead. So what's still missing to learn from humans in-the-wild? Our experience is that practical *applications* reveal the core roadblocks behind widespread functionality. Beyond the proof-of-concept experiments and user studies within each cited paper, we have applied our frameworks to food processing and assistive eating (see Figure 5). Within the food industry, there is a need for robots that can autonomously manipulate diverse items (e.g., small and numerous foods) and break down large items (e.g., separating meat from fat). For these applications we have developed specialized RISO grippers that combine and decouple rigid and soft components to grasp items across a 1 million times range in weight [21, 16]. In addition, we have prototyped general-purpose meat processing robots [39] that can detect and cut meat safely alongside human workers.

At the other end of the food spectrum we are interested in robot-assisted feeding for users with mobility limitations. Within this application assistive robot arms reach for morsels on the user's plate, acquire bite-sized portions, and then carry that food to the user's mouth. In [15, 17] we facilitate the process by creating a mechanical utensil that robots can leverage to grasp and transport foods. This utensil — called the Kiri-Spoon — has the form of a traditional spoon and the function of a soft gripper: when actuated, its kirigami structure deforms into a 3D bowl that encapsulates food items. Our case studies with stakeholders indicate that the Kiri-Spoon improves the user's assistive eating experience.

From our applications we have gathered further insight into learning from humans. In Sections III–V we discussed algorithmic intelligence (e.g., controllers, representations, interfaces). In practice, however, we see that *mechanical intelligence* is often necessary to simplify the learning problem and reach everyday solutions. Consider autonomous grasping: many of our works explore how robots can learn to manipulate objects. But instead of the robot learning complex grasp patterns, why not just engineer more functional grippers? Mechanical intelligence — such as specialized utensils for robot-assisted feeding — makes tasks inherently easier, reducing what the robot needs to learn. Our future work will seek to more seamlessly integrate both mechanical and algorithmic intelligence within learning systems.

References

[1] Benjamin A Christie and Dylan P Losey. LIMIT: Learning interfaces to maximize information transfer. *ACM Transactions on Human-Robot Interaction*, 13(4):1–26, 2024.

[2] Benjamin A Christie and Dylan P Losey. Safe interaction via Monte Carlo linear-quadratic games. *arXiv preprint arXiv:2504.06124*, 2025.

[3] Benjamin A Christie, Heramb Nemlekar, and Dylan P Losey. Personalizing interfaces to humans with user-friendly priors. In *IEEE International Conference on Robotics and Automation*, 2025.

[4] Yinlong Dai, Robert Ramirez Sanchez, Ryan Jeronimus, Shahabedin Sagheb, Cara M Nunez, Heramb Nemlekar, and Dylan P Losey. CIVIL: Causal and intuitive visual imitation learning. *arXiv preprint arXiv:2504.17959*, 2025.

[5] Soheil Habibian and Dylan P Losey. Encouraging human interaction with robot teams: Legible and fair subtask allocations. *IEEE Robotics and Automation Letters*, 7(3): 6685–6692, 2022.

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

[7] Soheil Habibian, Antonio Alvarez Valdivia, Laura H Blumenschein, and Dylan P Losey. A survey of communicating robot learning during human-robot interaction. *The International Journal of Robotics Research*, 44(4):665–698, 2025.

[8] Joshua Hoegerman and Dylan Losey. Reward learning with intractable normalizing functions. *IEEE Robotics and Automation Letters*, 8(11):7511–7518, 2023.

[9] Joshua Hoegerman, Shahabedin Sagheb, Benjamin A Christie, and Dylan P Losey. Aligning learning with communication in shared autonomy. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 11530–11536, 2024.

[10] Ananth Jonnavittula and Dylan P Losey. I know what you meant: Learning human objectives by (under) estimating their choice set. In *IEEE International Conference on Robotics and Automation*, pages 2747–2753, 2021.

[11] Ananth Jonnavittula and Dylan P Losey. Learning to share autonomy across repeated interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1851–1858, 2021.

[12] Ananth Jonnavittula and Dylan P Losey. Communicating robot conventions through shared autonomy. In *IEEE International Conference on Robotics and Automation*, pages 7423–7429, 2022.

[13] Ananth Jonnavittula, Shaunak A Mehta, and Dylan P Losey. SARI: Shared autonomy across repeated interaction. *ACM Transactions on Human-Robot Interaction*, 13(2): 1–36, 2024.

[14] Ananth Jonnavittula, Sagar Parekh, and Dylan P Losey. VIEW: Visual imitation learning with waypoints. *Autonomous Robots*, 49(1):1–26, 2025.

[15] Maya Keely, Brandon Franco, Casey Grothoff, Rajat Kumar Jenamani, Tapomayukh Bhattacharjee, Dylan P Losey, and Heramb Nemlekar. Kiri-Spoon: A kirigami utensil for robot-assisted feeding. *arXiv preprint arXiv:2501.01323*, 2025.

[16] Maya Keely, Yeunhee Kim, Shaunak A Mehta, Joshua Hoegerman, Robert Ramirez Sanchez, Emily Paul, Camryn Mills, Dylan P Losey, and Michael D Bartlett. Combining and decoupling rigid and soft grippers to enhance robotic manipulation. *Soft Robotics*, 2025.

[17] Maya N Keely, Heramb Nemlekar, and Dylan P Losey. Kiri-Spoon: A soft shape-changing utensil for robot-

assisted feeding. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 121–128, 2024.

[18] Dylan P Losey, Hong Jun Jeon, Mengxi Li, Krishnan Srinivasan, Ajay Mandlekar, Animesh Garg, Jeannette Bohg, and Dorsa Sadigh. Learning latent actions to control assistive robots. *Autonomous Robots*, 46(1):115–147, 2022.

[19] Shaunak A Mehta and Dylan P Losey. Unified learning from demonstrations, corrections, and preferences during physical human–robot interaction. *ACM Transactions on Human-Robot Interaction*, 13(3):1–25, 2024.

[20] Shaunak A Mehta, Sagar Parekh, and Dylan P Losey. Learning latent actions without human demonstrations. In *IEEE International Conference on Robotics and Automation*, pages 7437–7443, 2022.

[21] Shaunak A Mehta, Yeunhee Kim, Joshua Hoegerman, Michael D Bartlett, and Dylan P Losey. RISO: Combining rigid grippers with soft switchable adhesives. In *IEEE International Conference on Soft Robotics*, 2023.

[22] Shaunak A Mehta, Soheil Habibian, and Dylan P Losey. Waypoint-based reinforcement learning for robot manipulation tasks. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 541–548, 2024.

[23] Shaunak A Mehta, Forrest Meng, Andrea Bajcsy, and Dylan P Losey. StROL: Stabilized and robust online learning from humans. *IEEE Robotics and Automation Letters*, 9(3): 2303–2310, 2024.

[24] Shaunak A Mehta, Yusuf Umut Ciftci, Balamurugan Ramachandran, Somil Bansal, and Dylan P Losey. Stable-BC: Controlling covariate shift with stable behavior cloning. *IEEE Robotics and Automation Letters*, 10(2):1952–1959, 2025.

[25] Shaunak A Mehta, Heramb Nemlekar, Hari Sumant, and Dylan P Losey. L2D2: Robot learning from 2D drawings. *arXiv preprint arXiv:2505.12072*, 2025.

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

[27] Heramb Nemlekar, Robert Ramirez Sanchez, and Dylan P Losey. PECAN: Personalizing robot behaviors through a learned canonical space. *ACM Transactions on Human-Robot Interaction*, 2025.

[28] Sagar Parekh and Dylan P Losey. Learning latent representations to co-adapt to humans. *Autonomous Robots*, 47 (6):771–796, 2023.

[29] Sagar Parekh, Soheil Habibian, and Dylan P Losey. RILI: Robustly influencing latent intent. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2022.

[30] Sagar Parekh, Lauren Bramblett, Nicola Bezzo, and Dylan P Losey. Using high-level patterns to estimate how humans predict a robot will behave. *arXiv preprint arXiv:2409.13533*, 2024.

[31] Shahabedin Sagheb and Dylan P Losey. Counterfactual behavior cloning: Offline imitation learning from imperfect human demonstrations. *arXiv preprint arXiv:2505.10760*, 2025.

[32] Shahabedin Sagheb, Ye-Ji Mun, Neema Ahmadian, Benjamin A Christie, Andrea Bajcsy, Katherine Driggs-Campbell, and Dylan P Losey. Towards robots that influence humans over long-term interaction. In *IEEE International Conference on Robotics and Automation*, pages 7490–7496, 2023.

[33] Shahabedin Sagheb, Soham Gandhi, and Dylan P Losey. Should collaborative robots be transparent? *International Journal of Social Robotics*, pages 1–17, 2025.

[34] Shahabedin Sagheb, Sagar Parekh, Ravi Pandya, Ye-Ji Mun, Katherine Driggs-Campbell, Andrea Bajcsy, and Dylan P Losey. A unified framework for robots that influence humans over long-term interaction. *arXiv preprint arXiv:2503.14633*, 2025.

[35] Robert Ramirez Sanchez, Heramb Nemlekar, Shahabedin Sagheb, Cara M Nunez, and Dylan P Losey. RECON: Reducing causal confusion with human-placed markers. *arXiv preprint arXiv:2409.13607*, 2024.

[36] Antonio Alvarez Valdivia, Ritish Shailly, Naman Seth, Francesco Fuentes, Dylan P Losey, and Laura H Blumenschein. Wrapped haptic display for communicating physical robot learning. In *IEEE International Conference on Soft Robotics*, pages 823–830, 2022.

[37] Antonio Alvarez Valdivia, Soheil Habibian, Carly A Mendenhall, Francesco Fuentes, Ritish Shailly, Dylan P Losey, and Laura H Blumenschein. Wrapping haptic displays around robot arms to communicate learning. *IEEE Transactions on Haptics*, 16(1):57–72, 2023.

[38] Antonio Alvarez Valdivia, Benjamin A Christie, Dylan P Losey, and Laura H Blumenschein. A modular haptic display with reconfigurable signals for personalized information transfer. *arXiv preprint arXiv:2506.05648*, 2025.

[39] Ryan Wright, Sagar Parekh, Robin White, and Dylan P Losey. Safely and autonomously cutting meat with a collaborative robot arm. *Scientific Reports*, 14(1):299, 2024.