Belief-Aware Influence and Trust (BAIT): Shaping Human Belief During Repeated Human-Robot Interaction

***Abstract*—Repeated human-robot interaction (HRI) requires proactively accounting for humans who continually adapt to evolving beliefs about the robot. Prior frameworks often treat encounters as isolated events, suffering cumulative task performance decay as human perception drifts, or maintain long-term influence through erratic, unpredictable behavior that erodes perceived human trust and relies on computationally unscalable formulations. To address these gaps, we introduce the Belief-Aware Influence and Trust (BAIT) controller. BAIT integrates a hierarchical particle filter, which infers both fast human strategic shifts and slow perceptual belief updates, with a belief-aware Model Predictive Path Integral planner. BAIT explicitly optimizes the trade-off between long-horizon influence and human trust, while enforcing immediate task performance as a strict constraint. Across simulations, a human-subject study, and a real-world GEM vehicle deployments in repeated lane-merging scenarios, BAIT achieves task performance comparable to baselines that optimize long-term influence through unpredictability while yielding significantly higher user trust.**

***Index Terms*—Long Term Interaction, Acceptability and Trust, Autonomous Vehicle Navigation.**

I. Introduction

As robots become integrated into human spaces, they must account for an unavoidable dynamic: human behavior continually adapts in response to robot actions, especially across repeated encounters. Consider a delivery robot crossing paths with the same office workers every morning, an autonomous vehicle merging past the same commuters, or an assistive robot navigating around a household. In each of these settings, human behavior is neither static nor purely repetitive but rather evolves as humans observe the robot. When robots are predictably defensive, humans often exploit this predictability (e.g., human drivers refusing to let a hesitant vehicle merge or pedestrians asserting right-of-way), degrading task efficiency and efficacy. Sustained task performance across repeated encounters therefore depends on how a robot's immediate actions shape the human's evolving perception and future behavior.

Existing human-aware robot planning frameworks frequently overlook this human dynamic, treating each encounter as an isolated event and modeling the human as a fixed reward-maximizing agent [1], a static behavior distribution [2], [3], or an offline-trained trajectory predictor [4]. These stationary assumptions may be sufficient for one-shot interactions but fundamentally inadequate in the long-term settings described above. A planner that ignores how its current robot behavior alters future human behaviors optimizes only the immediate interaction, and empirically suffers from measurable task-performance decay as the human's perception drifts.

A few recent works instead model the evolution of human internal state to actively guide this dynamics to maintain cumulative task performance [5], [6], but face two primary

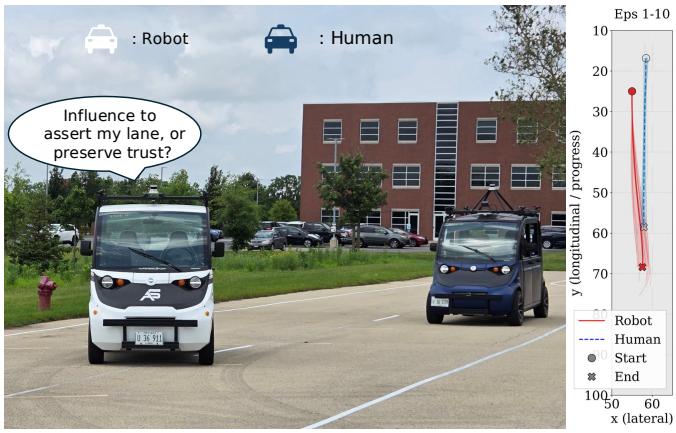

Fig. 1. **Repeated lane-merging encounters deployed on Polaris GEM vehicles.** Our BAIT-adapt robot (white) explicitly navigates the internal trade-off between influence to assert its lane right-of-way and preserving human trust. As shown in the corresponding trajectory rollouts (right), the BAIT controller (solid red) successfully executes the merge across repeated 10 episodes while consistently prompting the human driver (dashed blue) to yield.

limitations that constrain their deployment as practical controllers. First, these approaches maintain long-horizon influence by deliberately generating unpredictable robot behavior. Such unpredictability provides a reasonable initial mechanism for preventing humans from settling into a lazy or low-effort response pattern, thereby preserving the robot's influence over repeated interactions. However, this mechanism comes at a cost: the resulting robot behavior may be erratic, which degrades behavioral transparency and reduces human trust [7], [8]. We show that the unpredictability of previous long-term influence approaches narrows safety margins and erodes perceived human trust and comfort. Second, these long-term influence controllers face severe scalability bottlenecks relying on either strict leader-follower Stackelberg assumptions [6] or computationally heavy Mixed Observability Markov Decision Process (MOMDP) solvers [5].

To address these scalability and unpredictability limitations, we make the following contributions:

- **The BAIT framework.** We propose the **Belief-Aware Influence and Trust (BAIT)** controller, which explicitly trades off long-horizon influence (unpredictability) against human trust (transparency). BAIT structures the human's latent state into a two-timescale hierarchy: a short-term strategy variable (z^t) to predict immediate human behaviors and a long-term perception variable (ϕ^t) to capture the human's evolving beliefs of the robot.
- **Online Influence-Trust Management.** We track the human's evolving belief in real time using a two-layered particle filter that fuses action and geometric evidence. This inferred latent state is propagated forward as a deterministic surrogate into a belief-aware Model Predictive

Path Integral (MPPI) planner. This integration enables BAIT to uniquely optimize the influence-trust trade-off as a primary design axis, rather than optimizing either objective in isolation, while enforcing immediate task performance as a hard constraint.

- **Sustained Influence and Improved Human Trust.** Through simulation and real-world hardware experiments, we demonstrate that BAIT sustains higher cumulative task performance than influence-only baselines. A user study further shows that BAIT significantly improves human trust over influence-only controllers while preserving the task performance that trust-only methods, which prioritize transparent robot behavior, sacrifice.

II. Related Work

A. Influence in HRI: Short-Term to Long-Term

Traditional frameworks for HRI often treat encounters as an interdependent game where the robot acts proactively to shape human behavior [9], [10], [11], [12]. These proactive influence approaches are primarily optimized for single or short-term interactions, frequently employing Stackelberg formulation to maximize immediate task efficiency. However, these static models suffer from performance decay during repeated interactions because humans continually adapt over time, eventually exploiting or ignoring predictable robot behaviors [13], [14]. There have been recent attempts to model this long-term adaptation. Sagheb et al. [6] address long-horizon influence by deliberately generating unpredictable robot behaviors to prevent humans from exploiting predictable robot policies rather than coordinating with the robot. However, their method relies on the Stackelberg formulation, which is based on best-response human and leader-follower assumptions that are not realistic in the real world. Sagheb et al. [5] propose a framework that unifies these repeated interactions through a MOMDP formulation and show how prior methods are simplifications of this unified framework. However, these approaches rely on computationally intensive solvers. Standard online MOMDP solvers (e.g., POMCPOW) evaluate full belief trees, creating computational bottlenecks as state and action dimensions scale. We extend this lineage by maintaining long-horizon influence while ensuring real-time computational tractability via a conditionally decoupled hierarchical formulation.

B. Transparent Actions and Human Trust

Foundational human factors research establishes that legible and predictable robot trajectories are crucial for fostering user trust and comfort [8], [7]. However, this requirement creates a fundamental tension with recent proactive influence methods, which deliberately generate unpredictable robot behavior to force human compliance [5], [6] over repeated interactions. To mitigate this erosion of trust, Peng et al. [15] employ Bayesian Persuasion to derive a theoretical minimum trust level required to influence humans in a transparent way. However, this approach optimizes for a static human model, which may not fully capture scenarios where human perception continuously evolves. In this work, we address this influence-trust tension for dynamic human adaptation by explicitly treating the trade-off between unpredictability-driven influence and trust as our primary design axis, maintaining the robot's necessary influence while preserving human trust.

C. Human Modeling with Latent Representations

To capture the hidden intent of human partners, a prominent line of work represents their strategies using continuous or discrete latent variables [16], [17]. These methods infer continuous or discrete latent embeddings online to adapt robot policies within an interaction. However, most existing latent representation frameworks operate under the assumption that the underlying dynamics of human adaptation are stationary, often reducing the problem to a one-step Q-function for Markov Decision Process that ignores exploratory information gathering [5]. In this work, we structure the human's hidden state into a conditionally coupled, two-timescale hierarchy that separates high-frequency tactical reactions from gradual perceptual drift. This hierarchical separation bypasses the inefficiency of standard flat filters and allows the robot to actively plan over the human's evolving belief [18].

III. Problem Statement

We study robot decision-making in repeated human-robot interactions, focusing specifically on interactive lane changing in autonomous driving. When vehicles repeatedly negotiate shared lane space and right-of-way, humans do not act passively; they continuously observe the robot, update their internal perception of the robot, and adapt their strategy accordingly. Unlike one-shot encounters where human policies can be approximated as stationary [19], [20], repeated interactions induce non-stationary behavior across distinct timescales as the human adapts with their beliefs about the robot's behavior. Consequently, this adaptation requires a dual objective for the robot: (i) accurately inferring the underlying latent states, such as human's belief and behavioral strategy, and (ii) leveraging the inference for belief-aware planning. Specifically, the planner must steer the evolution of the human's perception toward configurations that secure favorable long-term outcomes without violating immediate task constraints. Crucially, in traffic navigation, securing these outcomes means safeguarding the robot's influence to prevent behavioral exploitation, ensuring the robot retains the leverage necessary to induce human cooperation, such as yielding during a lane change.

To capture human adaptation, we model the human's internal state using a hierarchy of two tightly coupled latent variables: a fast-changing short-term strategy z^t and a slow-evolving long-term belief ϕ^t. The *short-term strategy* $z^t \in \mathbb{R}^d$ encodes the human's instantaneous weighting vector across d competing task-related costs, such as forward progress (z_{progress}) versus collision avoidance (z_{safety}), enabling rapid reactions to the immediate physical context. Conversely, the *long-term belief* $\phi^t \in [0, 1]$ encodes the human's overarching perception of the robot, ranging from assertive ($\phi \to 0$) to defensive ($\phi \to 1$), and evolves gradually based on accumulated evidence across multiple interactions. Crucially, ϕ^t acts as a prior that modulates z^t formally captured by the conditional

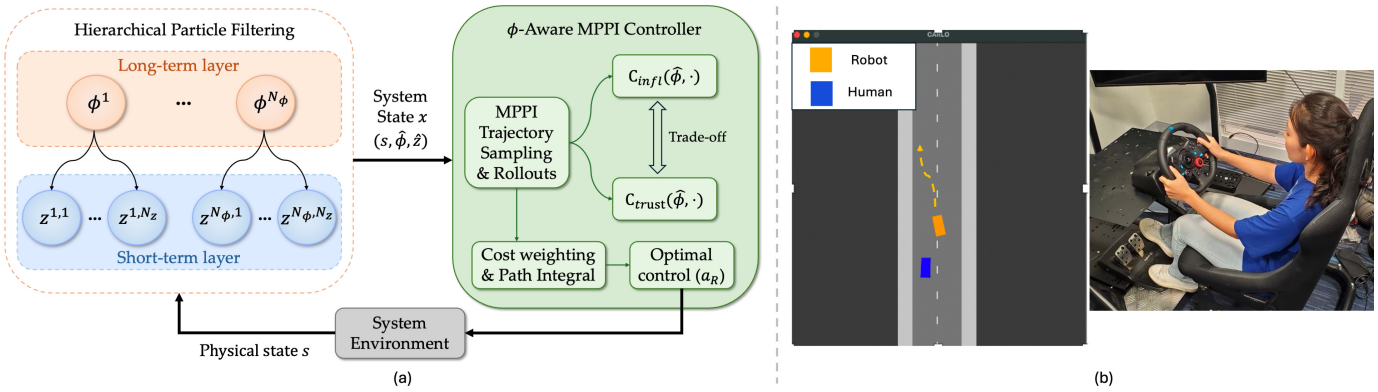

Fig. 2. **Overview of our BAIT controller.** (a) Our BAIT controller has two tightly coupled modules: a hierarchical particle filter that tracks the joint posterior over (ϕ, z) from observed interaction history, and an MPPI planner that rolls this posterior forward through a deterministic belief-propagation using the posterior mean $(\hat{\phi}, \hat{z})$. This planner optimizes an explicit influence-trust trade-off while enforcing a hard task-cost constraint. (b) The 2D CARLO lane-merging simulation environment.

probability $p(z^t|\phi^t)$. For example, during a lane merge, perceiving the robot as defensive ($\phi \to 1$) induces an aggressive momentary strategy ($z_{\text{progress}} > z_{\text{safety}}$), whereas perceiving an aggressive robot ($\phi \to 0$) prompts conservative yielding ($z_{\text{safety}} > z_{\text{progress}}$) to prioritize safety.

Building upon the Unified Framework for long-term influence proposed by [5], we formulate the problem as a MOMDP governed by the two-timescale latent hierarchy. However, unlike prior work that optimizes the MOMDP solely for cumulative influence, we introduce a scalable, real-time architecture that explicitly balances influence against human trust. Let $s^t = [s^t_{\mathcal{R}}, s^t_{\mathcal{H}}]$ denote the fully observable physical state of the robot $\mathcal{R}$ and the human $\mathcal{H}$. The unified MOMDP state is defined as $x^t = [s^t, z^t, \phi^t]$ with transitions defined by the joint dynamics:

$$x^{t+1} = \begin{bmatrix} s^{t+1} \\ z^{t+1} \\ \phi^{t+1} \end{bmatrix} = \begin{bmatrix} f_p(s^t, a^t) \\ f_s(s^t, a^t, z^t, \phi^t) \\ f_l(s^t, a^t, \phi^t) \end{bmatrix} = F(x^t, a^t_{\mathcal{R}}) \quad (1)$$

where $a^t = [a^t_{\mathcal{R}}, a^t_{\mathcal{H}}]$ represents the joint robot and human action, f_p defines the physical dynamics, f_s is the short-term strategy transition, and f_l is the long-term, belief transition. Substituting the human policy into equation 1 yields the underactuated dynamics $F(x^t, a^t_{\mathcal{R}})$ from the robot's perspective, where the human action $a^t_{\mathcal{H}} \sim \pi_{\mathcal{H}}(\cdot|s^t, z^t)$ is drawn from a maximum-entropy Boltzmann policy:

$$\pi_{\mathcal{H}}(a \mid s, z) \propto \exp\big(-\eta \cdot z^\top \bar{r}(s, a_{\mathcal{R}}, a_{\mathcal{H}})\big) \quad (2)$$

where $\bar{r} \in \mathbb{R}^d$ is a vector of instantaneous physical reward features parameterized linearly by the strategy weights z, and η is a rationality parameter where $\eta \to \infty$ recovers full rational behavior. A crucial structural property of these dynamics is that ϕ^t only influences actions indirectly through z^t, creating a partial-evidence problem where the human's true perception is masked by their immediate strategy.

The robot's objective is to solve this MOMDP over a receding planning horizon H by actively shaping ϕ^t to maximize cumulative task performance. Formally, the robot seeks an action sequence $\mathbf{a}_{\mathcal{R}} = a^{t:t+H}_{\mathcal{R}}$ that optimizes a joint objective:

$$\min_{\mathbf{a}_{\mathcal{R}}} \quad \mathbb{E}\left[\sum_{k=t}^{t+H} \mathcal{C}_{\text{task}}(s^k, a^k_{\mathcal{R}}) + \mathcal{C}_{\text{belief}}(\phi^{k+1}, s^k, a^k_{\mathcal{R}})\right] \quad (3)$$
$$\text{s.t.} \quad x^{k+1} = F(x^k, a^k_{\mathcal{R}})$$

To maintain influence while preserving human trust, the latent cost $\mathcal{C}_{\text{belief}}$ explicitly balances robot's behavioral unpredictability against transparency. Specifically, $\mathcal{C}_{\text{belief}}$ penalizes states where the human perceives the robot as easily exploitable ($\phi \to 1$), such as when the human perceives the robot will act passively defensive. Simultaneously, this cost regularizes against erratic or overly unpredictable actions, ensuring that the robot's overarching strategy remains sufficiently transparent to preserve human trust during the traffic negotiation.

IV. Belief-Aware Influence and Trust

To solve the Unified MOMDP formalism in Section III, we introduce the Belief-Aware Influence and Trust (BAIT) controller. Whereas prior long-term influence approaches focus on maximizing cumulative influence, often generating unpredictable behaviors that degrade human trust, BAIT explicitly optimizes the trade-off between proactive influence and trust (operationalized through behavioral transparency). To actively evaluate this dual objective in real time, BAIT overcomes the computational bottlenecks inherent to continuous-space MOMDPs by decoupling the problem into two tightly integrated modules: a hierarchical particle filter that tracks the joint posterior over (ϕ^t, z^t) from observed interactions, and an MPPI planner that rolls this inferred posterior forward using deterministic belief-propagation as shown in Fig. 2.

A. Hierarchical Inference of Human Latent States

We track the human's hidden state using a two-layered particle filter designed to preserve the timescale separation and conditional structure $p(z|\phi)$ of the latent hierarchy formalized in Section III. While a standard flat filter samples (ϕ, z) jointly, such a uniform structure scales inefficiently by enforcing redundant belief evaluations during high-frequency strategy z updates and causing a computational bottleneck. To resolve this, we structure the posterior as a hierarchical ensemble:

$$\mathcal{B}^t = \{\phi^{(i)}, w^{(i)}_\phi, \{z^{(i,j)}, w^{(i,j)}_z\}_{j=1}^{N_z}\}_{i=1}^{N_\phi} \quad (4)$$

where each of the N_ϕ long-term particles carries its own conditional ensemble of N_z short-term particles.

Prediction Step: The prediction step reflects this hierarchy by applying distinct transition kernels matched to each variable's timescale. The long-term belief variable ϕ propagates via a

mixture Beta kernel encoding bounded, gradual drift alongside a slow reset probability ϵ to prevent particle collapse:

$$\phi^{t+1} \sim (1-\epsilon)\text{Beta}\big(\kappa\phi^t, \kappa(1-\phi^t)\big) + \epsilon\text{Beta}(c,c) \quad (5)$$

where κ is a concentration parameter that controls the magnitude of the step-to-step drift variance (i.e., a higher κ yields a tighter concentration and lower variance), while c shapes the distribution of the random reset particles (e.g., c=1 for a uniform reset). Conditioned on this predicted belief ϕ^{t+1}, the short-term strategy z evolves via an Ornstein-Uhlenbeck (OU) process that mean-reverts toward an attractor z_{tgt} determined by its parent ϕ^{t+1}:

$$z^{t+1} = z^t + \alpha_{\text{ou}}\big(z_{\text{tgt}}(\phi^{t+1}) - z^t\big) + \sigma_z\xi^t, \quad \xi^t \sim \mathcal{N}(0,I) \quad (6)$$

where $\alpha_{\text{ou}} \in (0,1]$ is the mean-revision rate that dictates how rapidly the human's short-term strategy adapts toward the target attractor $z_{tgt}(\phi^{t+1})$, and σ_z scales the standard normal variable ξ^t to model the stochasticity and natural execution variance inherent to human driving behavior. This coupling structurally enforces the behavioral prior established in Section III (e.g., drawing z towards an aggressive momentary strategy when $\phi \to 1$ and conservative strategy $\phi \to 0$).

Update Step: To robustly update these particles despite the indirect observability, the filter's update step applies distinct evidence to each hierarchical layer. At the bottom layer, the short-term z-particles are reweighted using *action evidence* via the Boltzmann log-likelihood of the observed human actions in Eq. (2) over a decaying history length W:

$$\ell_z^{(i,j)} = \sum_{\tau=t-W+1}^{t} \rho^{\,t-\tau}\, \log \pi_{\mathcal{H}}\big(a_{\mathcal{H}}^{\tau} \,\big|\, s^{\tau},\, z^{(i,j)}\big) \quad (7)$$

where $\rho \in (0,1]$ is an exponential decay factor. To prevent premature weight collapse under low-signal observations, we update the normalized weight using a soft inertia blend:

$$w_z^{(i,j)} \leftarrow (1-\alpha_z) w_z^{(i,j)} + \alpha_z \frac{\exp(\ell_z^{(i,j)})}{\sum_{m=1}^{N_z} \exp(\ell_z^{(i,m)})} \quad (8)$$

where α_z is the update rate for the new evidence.

At the top layer, the long-term ϕ-particles are updated by fusing this indirect *action evidence* with a direct *geometric evidence* channel (e.g., lateral gap and closing rate for lane merging). First, the action evidence is aggregated for each parent $\phi^{(i)}$ by marginalizing the log-likelihood of its z-ensemble in Eq. 7:

$$\ell_\phi^{(i)} = \log \sum_{j=1}^{N_z} \exp\big(\ell^{(i,j)}\big). \quad (9)$$

Simultaneously, a geometric proxy infers the human's belief based on recent physical interactions; for example, if the robot merges aggressively, it naturally induces the human to believe the robot will not yield. We extract an instantaneous yield-belief target y^t from these kinematic features and maintain a smooth yield-belief b_y^t via an exponential moving average:

$$b_y^t \leftarrow (1-\alpha_\phi) b_y^{t-1} + \alpha_\phi y^t \quad (10)$$

where α_ϕ is the smoothing rate. The filter fuses both evidence channels in log-space to update the intermediate ϕ^t weights:

$$\log \tilde{w}_\phi^{(i)} = \log w_\phi^{(i)} + \frac{1}{T_\phi}\ell_\phi^{(i)} + \frac{\mathbb{I}_{\text{int}}}{T_\phi} \log p\big(b_y^t \mid \phi^{(i)}\big) \quad (11)$$

where T_ϕ is a scaling temperature, $p(b_y^t|\phi^{(i)})$ represents the likelihood of the particle given the geometric evidence, and $\mathbb{I}_{\text{int}}$ is an indicator function that activates the geometric channel only when the human enters a critical interaction zone defined by spatial threshold d_{interact} and the robot is longitudinally ahead of the human driver. In the context of lane merging, this threshold prevents nominal driving maneuvers from miscalibrating the belief ϕ^t by ensuring geometric evidence is only fused when vehicles are close enough to actively negotiate the right-of-way. Normalizing these intermediate weights $\tilde{w}_\phi^{(i)}$ yields the final posterior weights $w_\phi^{(i)}$. This dual-channel fusion yields a joint posterior that remains well-calibrated across varying interaction intensities, effectively resolving the partial-evidence bottleneck.

Resampling Step: Finally, resampling is triggered independently for each layer based on an effective-sample-size (ESS) criterion, respecting the timescale hierarchy. The long-term ϕ-layer employs Liu-West kernel smoothing to maintain stable continuous moments. Conversely, the short-term z-ensembles are resampled with a partial injection of new particles drawn from the prior $p(z|\phi^{(i)})$ to ensure the strategy remains agile to rapid behavioral shifts.

B. Belief-Aware MPPI Planning

Leveraging the inferred posterior $(\hat{\phi}, \hat{z})$, our belief-aware MPPI planner actively optimizes the robot's control sequence by explicitly trading off long-term influence against human trust. At each time step, the controller samples K candidate action sequences and evaluates them against a belief cost, $\mathcal{C}_{\text{belief}}$. This cost dynamically balances two opposing terms using an arbitration parameter $\beta \in [0,1]$ that shifts the controller from pure trust (β=0) to pure influence (β=1):

$$\mathcal{C}_{\text{belief}} = \big(\beta\mathcal{C}_{\text{infl}} + (1-\beta)\mathcal{C}_{\text{trust}}\big)\mathbb{I}_{\text{int}} \quad (12)$$

where $\mathbb{I}_{\text{int}}$ is the same spatial interaction indicator defined in Eq. (11). Through this gating, the robot restricts its belief-shaping maneuvers to active lane-changing negotiations, avoiding unnecessary adjustments in open traffic and acting only when the human driver's internal state actively evolves.

The two belief-aware terms in Eq. (12) are anchored on propagated entropy of the human belief regarding the robot's yielding intent $\mathcal{H}(\hat{\phi}^{k+1})$. The influence cost $\mathcal{C}_{\text{infl}}$ promotes unpredictable trajectories ($\mathcal{H} \uparrow$) to prevent the human from settling into an assertive policy, thereby inducing conservative yielding. However, optimizing for entropy in isolation could produce erratic, task-irrelevant behavior (e.g., randomly oscillating across lanes). To prevent this, we introduce a directional regularizer. Weighted heavily below the entropy ($w_d \ll w_e$), this regularizer biases the unpredictability toward assertive outcome ($\phi \to 0$), ensuring the robot's actions actively claim right-of-way:

$$\mathcal{C}_{\text{infl}} = w_e\big(1 - \mathcal{H}(\hat{\phi}^{k+1})\big)^2 + w_d(\hat{\phi}^{k+1})^2. \quad (13)$$

In contrast, the trust cost $\mathcal{C}_{\text{trust}}$ promotes *transparent* trajectories whose induced perception is predictable ($\mathcal{H}(\hat{\phi}^{k+1})\downarrow$), signaling an accommodating intent ($\phi \to 1$) via a legible merge:

$$\mathcal{C}_{\text{trust}} = w_e\mathcal{H}(\hat{\phi}^{k+1})^2 + w_d(1-\hat{\phi}^{k+1})^2. \quad (14)$$

Since $\mathcal{H}(\hat{\phi})$ is symmetric, $w_d \ll w_e$ only sets belief direction, not transparency (ablated in Section V). To evaluate $\mathcal{C}_{\text{infl}}$ and $\mathcal{C}_{\text{trust}}$ across K candidate sequences and H horizon steps, running the full sample-based filter is computationally intractable. Instead, we ensure real-time scalability by propagating a noise-free *deterministic surrogate* of the posterior moments. We initialize each rollout with the current posterior means, $\hat{\phi}^t = \mathbb{E}[\phi^t|\mathcal{B}^t]$ and $\hat{z}^t = \mathbb{E}[z^t|\mathcal{B}^t]$, alongside the current yield-belief b_y^t. At each horizon step $k \in [t, t+H-1]$, we propagate these posteriors forward via three operations: (i) predicting the human action using the Boltzmann policy $a_{\mathcal{H}}^k \sim \pi_{\mathcal{H}}(\cdot|s^k, \hat{z}^k)$; (ii) advancing physical state through the joint dynamics from Eq. (1); and (iii) updating the surrogate state using the geometric kernel from Eq. (10) and the noise-free strategy dynamics from Eq. (6). Our BAIT approach is explicitly reliant on this structured human model, and this efficient forward pass tracks the expected latent state, providing the necessary signals to compute the belief-aware costs without the computational burden of a full ensemble simulation.

Finally, to ensure that belief-shaping maneuvers do not compromise physical safety or primary task objectives, the planner treats the task cost ($\mathcal{C}_{\text{task}}$ in Eq. 3), which encompasses collision avoidance and lane merging, as a strict conditional value-at-risk (CVaR) constraint:

$$\mathcal{C}_{\text{task}} = M\, \mathbb{I}\big[\mu_{\text{task}}^{(k)} + \eta_r\, \sigma_{\text{task}}^{(k)} > \tau\big] \tag{15}$$

where M is a sufficiently large scalar penalty, $\mu_{\text{task}}^{(k)}$ and $\sigma_{\text{task}}^{(k)}$ denote the mean and standard deviation of the task cost for the k-th rollout, η_r is a risk-aversion parameter, and τ is the acceptable cost threshold. Any rollout whose upper-bound risk assessment exceeds this threshold receives the penalty M, driving its corresponding MPPI sampling weight to near-zero and functionally excluding the unsafe trajectory from the control update. This architecture ensures that BAIT exercises its influence-trust trade-off strictly within the set of physically safe and task-feasible trajectories.

V. Simulation Experiment

We evaluate BAIT in a repeated lane-merging scenario with simulated humans. The robot starts ahead and merges into the human's lane. If the robot influences the human to yield (low lane progress), the merge succeeds; otherwise, the human maintain or increases speed, blocking the merge. This experiment aims to assess two core algorithm capabilities: (i) the invariant estimation accuracy of the two-layered particle filter against ground-truth human parameters across different robot control modes, and (ii) the impact of the influence–trust arbitration on cumulative task performance against an adapting human agent. While this section strictly assesses objective algorithmic metrics, subjective evaluations of human trust are addressed through a user study in Section VI.

Hypothesis 1: Influence and Task Success. *In repeated encounters, static and trust policies are easily exploitable, whereas influence policies maintain influence. Adaptive arbitration dynamically shifts between these strategies.*

A. Simulator and Adaptive Human Model

We instantiate the MOMDP as a two-lane merging task in a 2D driving simulator based on CARLO [21] across $N = 20$ consecutive episodes and conduct 30 independent trials using random seeds. Each episode consists of 75 steps with a step size of $\Delta t = 1$ s. While the physical state resets each episode, the human's latent state (ϕ^{GT}, z^{GT}) persists to model continuous adaptation across interactions. These ground-truth parameters remain inaccessible to all controllers.

Following recent findings that humans aggressively exploit predictable policies but become conservative under uncertainty [5], [6], our generative human model simulates more complex, non-linear dynamics than the robot's internal filter. Specifically, the simulated human updates its ground-truth belief ϕ^{GT} using the same kinematic moving-average formulation defined for the robot's geometric proxy (b_y^t in Eq. (10)). Furthermore, while the robot's filter models the strategy attractor (z_{tgt} in Eq. (6)) as a simple interpolation of ϕ, the actual human strategy z^{GT} is drawn toward a highly non-linear ground-truth attractor z_{tgt}^{GT}.To rigorously test our filter's tracking capability, z_{tgt}^{GT} augments the baseline interpolation with two adversarial, rule-based overlays: a *cautious bias* that smoothly shifts the human toward a conservative prototype when belief entropy is high ($\mathcal{H}(\phi^{GT}) > 0.75$), and an *aggressive spike* that instantly snaps to an aggressive prototype if low entropy ($\mathcal{H}(\phi^{GT}) < 0.70$) persists over a moving window of 6 steps. Without access to these underlying non-linear reactive thresholds, the robot's particle filter must robustly infer these rapid, strategic shifts using its simplified surrogate dynamics.

B. Implementation Details.

We instantiate the full BAIT pipeline using $K = 100$ rollout sequences, $N_\phi = 20$ long-term particles, and $N_z = 50$ short-term particles. Specifically, each long-term particle tracks a single belief dimension $\phi \in [0, 1]$, while each short-term particle tracks the 2-dimensional short-term strategy trade-off between forward progress (z_{progress}) and collision avoidance (z_{safety}) discussed in Section III. These two adapted dimensions operate within 5-dimensional instantaneous cost vector ($d_z = 5$) that also accounts for lane center deviation, heading, and steering smoothness. For the hierarchical particle filter, we maintain a history window $W = 5$ with an exponential decay $\rho = 0.7$. The short-term strategy employs an update rate $\alpha_z = 0.8$, a mean-reverting rate $\alpha_{ou} = 0.5$, and an Ornstein-Uhlenbeck noise scale $\sigma_z = 0.1$. The long-term belief utilizes a geometric smoothing rate $\alpha_\phi = 0.25$ and a scaling temperature $T_\phi = 3.0$. The beta kernel is configured with a concentration $\kappa = 50.0$, a slow reset probability $\epsilon = 0.02$, and a uniform reset shape $c = 1.0$. The critical interaction zone is triggered at a spatial threshold $d_{\text{interact}} = 25.0$m. For the MPPI planner, the influence cost uses an entropy weight $w_e = 0.8$ and directional weight $w_d = 0.2$ while trust cost shifts these to $w_e = 0.9$ and $w_d = 0.1$. The CVaR safety constraint utilizes a risk-aversion parameter $\eta_r = 1.0$, a per-step acceptable cost threshold $\tau = 2.5$, and an effective infinite scalar penalty $M = 10^9$.

TABLE I
ROBUSTNESS OF THE HIERARCHICAL PARTICLE FILTER ESTIMATION ACROSS VARYING ROBOT CONTROL MODES.

		Latent State Estimation				Trajectory Prediction		
Environment	**Method**	ϕ **BCE** ($\downarrow$)	ϕ **MSE** ($\downarrow$)	z **MSE** ($\downarrow$)	z **Cos.** ($\uparrow$)	**1-ADE** ($\downarrow$)	**H-ADE** ($\downarrow$)	**H-FDE** ($\downarrow$)
Simulated Human	BAIT-trust	0.44(0.25)	0.04(0.03)	0.09(0.10)	0.95(0.04)	0.050(0.010)	0.530(0.060)	1.100(0.140)
	BAIT-infl	0.69(0.25)	0.05(0.02)	0.04(0.03)	0.97(0.02)	0.050(0.010)	0.510(0.070)	1.070(0.140)
	BAIT-adapt	0.52(0.31)	0.05(0.03)	0.07(0.09)	0.96(0.04)	0.050(0.010)	0.520(0.060)	1.090(0.130)
In-Person User Study	BAIT-trust	0.32(0.17)	0.02(0.02)	0.01(0.01)	1.00(0.00)	0.046(0.006)	0.584(0.136)	1.301(0.325)
	BAIT-infl	0.72(0.18)	0.05(0.02)	0.02(0.01)	0.99(0.01)	0.046(0.005)	0.559(0.118)	1.241(0.283)
	BAIT-adapt	0.52(0.26)	0.04(0.02)	0.02(0.01)	0.99(0.01)	0.045(0.005)	0.554(0.116)	1.230(0.278)

Note: Results show posterior mean (std. dev.) evaluated step-wise across 600 interactions (30 subjects×20 consecutive episodes). BCE/MSE: binary cross-entropy/mean-squared error; Cos.: cosine similarity; ADE/FDE: average/final displacement error at the 1- and H-step horizon.

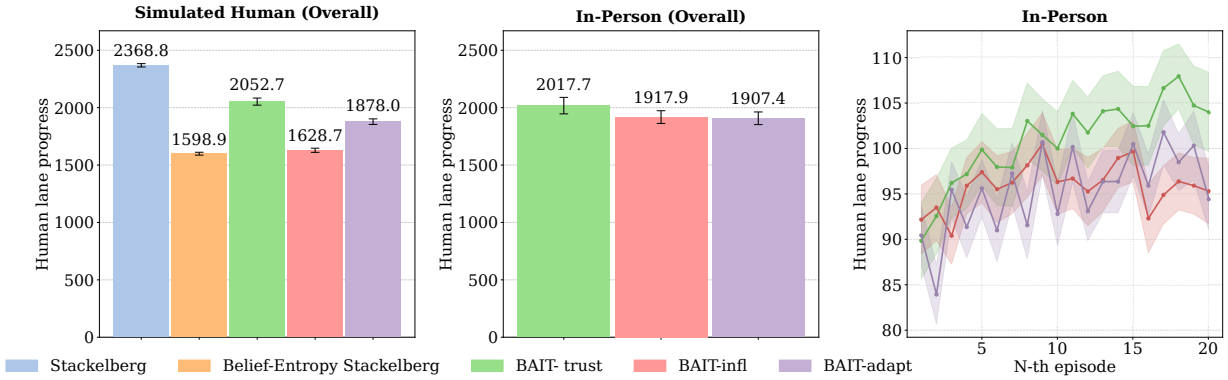

Fig. 3. **Comparison of Human Lane Progress.** Overall lane progress for the BAIT variants (BAIT-trust, BAIT-infl, and BAIT-adapt) against Stackelberg baselines (Stackelberg, Belief-Entropy Stackelberg) when interacting with a belief-aware simulated human (left), alongside the overall (center) and episode-by-episode (right) progress of in-person users interacting with the three BAIT variants. Both the simulated and in-person data consists of 600 total interactions (30 simulated/human subjects×20 consecutive episodes). Error bars denote standard error. In simulation (left), both Belief-Entropy Stackelberg and BAIT-infl maintain low human lane progress by successfully influencing the human to yield. In contrast, the standard Stackelberg and BAIT-trust modes resulted in high human lane progress, as the human becomes assertive upon perceiving that the robot will yield. The BAIT-adapt mode balances these two extremes by strategically switching between influence and trust based on the recent interaction history. Furthermore, the in-person results (center, right) confirm that this strategic switching allows BAIT-adapt to consistently regulate human lane progress across repeated physical interactions.

C. Baselines.

We evaluate our BAIT framework against two baselines: **Stackelberg**, which assumes a stationary best-response human; and **Belief-Entropy Stackelberg**, which augments the **Stackelberg** objective with a belief-entropy reward (similar to our influence cost in Eq. (13)) to optimize strictly for latent influence. To ensure fair comparison, all controllers share the exact planning horizon and resolution (H=5, Δt=1), physical parameters, and reward configurations as BAIT. We also test two ablations to evaluate the extremes of our cost formulation: pure trust **BAIT-trust** (β=0) and pure influence **BAIT-infl** (β=1). Finally, **BAIT-adapt** arbitrates between episodes via binary, history-dependent switching ($\beta\in\{0,1\}$), shifting to influence mode (β=1) if the robot's success rate falls below 60% over a sliding 5-episode window or the human's belief indicates expected yielding ($\hat{\phi}$>0.7), and reverts to trust once both conditions clear. We evaluate the task and inference fidelity jointly to prevent the emergence of degenerate policies.

D. Result and Discussions

Estimation Robustness. Table I demonstrates the policy-invariant tracking accuracy of BAIT's hierarchical particle filter. Because the Stackelberg baselines lack complete hierarchical latent state estimation, we compare the three BAIT variants to validate our filter robustly tracks human adaptation regardless of the robot policy, rather than serving as an ablation. Furthermore, presenting in-person tracking alongside simulated data validates our simulated human as a reliable proxy prior to discussion task outcomes.

Because ground-truth internal states are inaccessible for real human drivers, we extract offline pseudo-labels $(\hat{\phi}^{GT}, \hat{z}^{GT})$ over the recorded dataset by joint trajectory optimization. Because the belief ϕ only influences human actions indirectly, minimizing the negative log-likelihood (NLL) of actions is insufficient to uniquely recover the latent hierarchy. Thus, we extract the pseudo-labels by minimizing the action NLL subject to the structural priors of our generative hierarchy:

$$\min_{\hat{\phi}^{1:T},\hat{z}^{1:T}} \sum_{t=1}^{T} \Big[-\log \pi_{\mathcal{H}}(a_{\mathcal{H}}^t \mid s^t, \hat{z}^t) + \lambda_z \|\hat{z}^t - z_{\text{tgt}}(\hat{\phi}^t)\|^2 + \lambda_\phi (\hat{\phi}^t - b_y^t)^2 + \mathcal{C}_{\text{smooth}} \Big] \tag{16}$$

where the first term fits the short-term strategy to the observed physical driving, the second term enforces the hierarchical coupling by anchoring $\hat{z}^t$ to its $\hat{\phi}^t$-conditioned prototype, the third term anchors the extracted belief to the geometric yield proxy b_y^t (Eq. (10)), and $\mathcal{C}_{\text{smooth}}$ applies standard temporal smoothing to prevent erratic, step-to-step fluctuations in both variables. Evaluating against these optimized pseudo-labels, the filter maintains tight convergence and low error across both environments, accurately capturing internal state transitions.

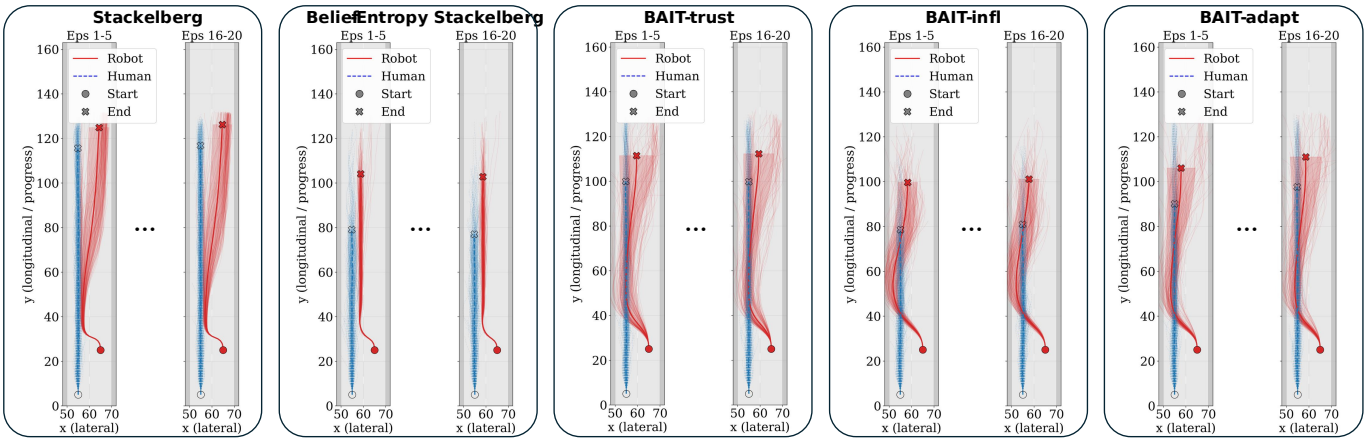

Fig. 4. **Evolution of Interaction trajectories.** Comparison of robot (solid red) and simulated human (dashed blue) trajectories during early (Eps $1-5$) and late (Eps $16-20$) interactions. Results are aggregated over 30 independent trials across five control modes. While standard Stackelberg and BAIT-trust initially secure the lane merge, they fail to prevent the human from becoming aggressive over repeated interactions. Conversely, the Belief-Entropy Stackelberg, BAIT-infl, and BAIT-adapt controllers successfully maintain the human's yielding behavior across the majority of trials.

Trajectory prediction remains flat across variants (H-ADE $\approx$0.5), confirming the expected posterior moments provide a reliable input to our deterministic surrogate regardless of β.

Task performance. To evaluate the robot's influence on human yielding, we report human lane progress analyzed via repeated-measures ANOVA (F=250.51, p<0.001) and Bonferroni-corrected post-hoc test (α=0.05). Supporting **H1**, static **Stackelberg** and **BAIT-trust** allow significantly greater human lane progress than the influence-optimizing models, **Belief-Entropy Stackelberg** and **BAIT-infl**, (all pairwise p<0.001), as the human agent exploits their predictable yielding (Fig. 3). Ablating the directional regularizer, a transparent-assertive variant ($w_e\mathcal{H}(\hat{\phi}^{k+1})^2 + w_d(\hat{\phi}^{k+1})^2$) yields total lane progress 1962.64 ± 36.58, matching **BAIT-trust** (1957.66), not **BAIT-infl** (1533.74), so predictability, not belief direction, governs influence. Furthermore, **BAIT-adapt** successfully arbitrates between influence and trust, achieving a statistically significant reduction in lane progress compared to **Stackelberg** and **BAIT-trust** (both p<0.001) yet significantly higher than **BAIT-infl** and **Belief-Entropy Stackelberg** (both p<0.001). Crucially, BAIT achieves this effective behavioral arbitration without sacrificing real-time scalability: by propagating a noise-free deterministic surrogate rather than branching full trees like online MOMDP solvers (e.g., POMCPOW) or nested best-response MPCs, BAIT bypasses continuous-space branching bottlenecks to achieve real-time execution at $\approx$0.11 s/step.

VI. User Study

We conducted an in-person user study to assess subjective trust and perceived comfort across three BAIT variants (**BAIT-trust**, **BAIT-infl**, and **BAIT-adapt**). Because our simulation result in Section V confirmed equivalent efficacy between **Belief-Entropy Stackelberg** and **BAIT-infl** ($p = 1.000$), we focus our experimental evaluation on three BAIT controllers. Participants interacted with the merging robot within the environment described in Section V, controlling the human vehicle using a Logitech G29 steering wheel and pedal set (Fig. 2). To ensure consistency, all implementation parameters remained identical to those used in the simulation experiments.

Hypothesis 2: Human Trust and Comfort. *Transparency builds human trust and comfort, whereas unpredictability degrades them. Adaptive arbitration exerts necessary influence without sacrificing these subjective metrics.*

A. Participants

We recruited 30 participants (11 female, age 28$\pm$6.2), 14 of whom reported prior experience riding in or driving alongside autonomous vehicles. This study was conducted in strict accordance with Institutional Review Board guidelines (#[Anonymized]). Following a within-subject design, each participant completed 20 consecutive merging episodes against each of the three controllers. Participants rated four custom-designed measures on a 7-point Likert scale: comfort ("I felt comfortable driving alongside this autonomous vehicle (AV)"), discomfort ("Interacting with this AV made me feel tense or uneasy," reverse-coded as a consistency check), trust ("I trusted this AV to behave safely during our interactions"), and anticipation ("I could anticipate what AV was going to do next"). The presentation order was fully counter-balanced to mitigate ordering effects.

B. Results and Discussions

We report subjective outcomes evaluated using non-parametric Friedman tests and Bonferroni-corrected Wilcoxon post-hoc analyses (α=0.0167). The in-person experimental results closely mirror our simulation results (Fig. 3). **BAIT-trust** allowed the highest human lane progress, which increased over time as participants exploited the robot's transparent yielding. Conversely, both **BAIT-infl** and **BAIT-adapt** successfully asserted merging to maintain lower human lane progress over the episodes. Notably, **BAIT-adapt** achieved a statistically significant reduction compared to **BAIT-trust** (p<0.01), while demonstrating equivalent efficacy to **BAIT-infl** (p=0.811).

The user study results comparing the human trust, comfort, and anticipation for the three controller variants also supported **H2**. The participants felt significantly more comfortable with the predictable yielding of **BAIT-trust** (comfort median 5, discomfort median 3) compared to **BAIT-infl** (comfort median

2, discomfort median 5) or **BAIT-adapt** (comfort median 4, discomfort median 5) significantly degrading overall comfort and discomfort ($p<0.005$ across all four pairwise comparisons). However, the trust variance was highly significant ($\chi^2(2)=14.04, p<0.001$). While **BAIT-trust** maximized trust (median 5), **BAIT-adapt** (median 3) preserved user trust at level significantly higher than the pure influence baseline **BAIT-infl** (median 2, $p= 0.0128$). Importantly, this trust preservation coexists with effective influence maintenance. Because **BAIT-adapt** switches between modes across episodes via hysteresis, participants perceived it as less predictable (anticipation median 4 vs. 6 for **BAIT-trust**, $p= 0.003$). These results reveal that while controlled adaptation must inherently sacrifice the human comfort to maintain long-term influence for task efficiency adversarial exploitation, allowing **BAIT-adapt** to make the human yield as effectively, it successfully mitigates degradation of human trust caused by pure influence.

VII. Real-World Experiment

To validate real-time deployment, we implemented the proposed **BAIT-adapt** controller on a physical Polaris GEM e2. The ego vehicle interacted with a human-driven GEM e4 in a scaled lane-merging scenario (Similar to Section V and Section VI). State estimation was achieved using a Septentrio AsteRx SBi3 Pro+ GNSS receiver with a ProPak 6 RTK base station, eliminating localization uncertainty while low-level steering and throttle commands were executed via ROS PACMod. As shown in Fig. 1, the **BAIT-adapt** runs in real time onboard the ego vehicle and successfully secures the merge. The robot smoothly asserts its right-of-way and consistently prompts the human driver to yield across repeated interactions, demonstrating the practical viability of the BAIT.

VIII. Conclusion

We introduce BAIT, a real-time controller for repeated human-robot interactions that integrates a hierarchical particle filter–tracking the human's short-term strategy and long-term perception–with a deterministic surrogate MPPI planner. BAIT explicitly arbitrates between long-horizon influence and human trust, while enforcing immediate task performance as a strict CVaR constraint. Across simulations, a user study, and real-world GEM vehicle deployments, BAIT maintains influence by consistently prompting humans to yield while mitigating the severe trust degradation of pure influence controllers. Although mode switching inherently sacrifices the human comfort of a trust policy, it achieves controlled unpredictability that preserves influence. Future work will replace the current binary-mode switching by optimizing the arbitration variable continuously to enable smoother adaptation.

References

[1] Y.-J. Mun, M. Itkina, S. Liu, and K. Driggs-Campbell, "Occlusion-aware crowd navigation using people as sensors," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 12 031–12 037.

[2] A. Xie, D. Losey, R. Tolsma, C. Finn, and D. Sadigh, "Learning latent representations to influence multi-agent interaction," in *Conference on Robot Learning (CoRL)*, 2021, pp. 575–588.

[3] M. Muchen Sun, F. Baldini, K. Hughes, P. Trautman, and T. Murphey, "Mixed strategy nash equilibrium for crowd navigation," *The International Journal of Robotics Research (IJRR)*, vol. 44, no. 7, pp. 1156–1185, 2025.

[4] Y. Huang, J. Du, Z. Yang, Z. Zhou, L. Zhang, and H. Chen, "A survey on trajectory-prediction methods for autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 652–674, 2022.

[5] S. Sagheb, S. Parekh, R. Pandya, Y.-J. Mun, K. Driggs-Campbell, A. Bajcsy, and D. P. Losey, "A unified framework for robots that influence humans over long-term interaction," *arXiv preprint arXiv:2503.14633*, 2025.

[6] S. Sagheb, Y.-J. Mun, N. Ahmadian, B. A. Christie, A. Bajcsy, K. Driggs-Campbell, and D. P. Losey, "Towards robots that influence humans over long-term interaction," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 7490–7496.

[7] J.-L. Bastarache, C. Nielsen, and S. L. Smith, "On legible and predictable robot navigation in multi-agent environments," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 5508–5514.

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

[9] D. Sadigh, S. Sastry, S. A. Seshia, and A. D. Dragan, "Planning for autonomous cars that leverage effects on human actions," in *Robotics: Science and Systems (RSS)*, 2016.

[10] J. F. Fisac, E. Bronstein, E. Stefansson, D. Sadigh, S. S. Sastry, and A. D. Dragan, "Hierarchical game-theoretic planning for autonomous vehicles," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2019, pp. 9590–9596.

[11] W. Schwarting, A. Pierson, J. Alonso-Mora, S. Karaman, and D. Rus, "Social behavior for autonomous vehicles," *Proceedings of the National Academy of Sciences*, vol. 116, no. 50, pp. 24 972–24 978, 2019.

[12] S. Parekh and D. P. Losey, "Learning latent representations to co-adapt to humans," *Autonomous Robots*, vol. 47, no. 6, pp. 771–796, 2023.

[13] M. Cooper, J. K. Lee, J. Beck, J. D. Fishman, M. Gillett, Z. Papakipos, A. Zhang, J. Ramos, J. A. Shah, and M. L. Littman, "Stackelberg punishment and bully-proofing autonomous vehicles," in *International Conference on Social Robotics (ICSR)*, 2019, pp. 368–377.

[14] A. Ayub, Z. De Francesco, P. Holthaus, C. L. Nehaniv, and K. Dautenhahn, "Continual learning through human-robot interaction: Human perceptions of a continual learning robot in repeated interactions," *International Journal of Social Robotics*, vol. 17, no. 2, pp. 277–296, 2025.

[15] S. Peng, K. Driggs-Campbell, and R. Dong, "Trust-aware embodied bayesian persuasion for mixed-autonomy," *arXiv preprint arXiv:2509.15404*, 2025.

[16] A. Xie, D. P. Losey, R. Tolsma, C. Finn, and D. Sadigh, "Learning latent representations to influence multi-agent interaction," in *Conference on Robot Learning (CoRL)*, 2021, pp. 575–588.

[17] A. Bajcsy, A. Loquercio, A. Kumar, and J. Malik, "Learning vision-based pursuit-evasion robot policies," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 9197–9204.

[18] Z. Huang, Y.-J. Mun, F. C. Pouria, and K. Driggs-Campbell, "Hierarchical intention tracking with switching trees for real-time adaptation to dynamic human intentions during collaboration," *arXiv preprint arXiv:2506.07004*, 2025.

[19] L. J. Ratliff, R. Dong, S. Sekar, and T. Fiez, "A perspective on incentive design: Challenges and opportunities," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 2, no. 1, pp. 1–34, 2018.

[20] J. Foerster, R. Y. Chen, M. Al-Shedivat, S. Whiteson, P. Abbeel, and I. Mordatch, "Learning with opponent-learning awareness," in *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2018, pp. 122–130.

[21] Z. Cao, E. Biyik, W. Z. Wang, A. Raventos, A. Gaidon, G. Rosman, and D. Sadigh, "Reinforcement learning based control of imitative policies for near-accident driving," in *Robotics: Science and Systems (RSS)*, 2020.