Exploring Active Human Goal Inference in Shared Autonomy and Autonomous Driving

Nick Landolfi and Anca D. Dragan

***Abstract*—Many human-robot interaction tasks require inferring human goals. Shared autonomy relies on the robot inferring what the person wants in order to help them achieve it. Autonomous driving relies on the robot inferring what the person wants in order to effectively share the road with them. Goal inference is typically *passive*: the robot looks to human actions for observations about what they want, updating its estimate at every step, and planning its actions using the current estimate. In this paper we explore the pros and cons of *active* inference, where the robot leverages its actions to trigger informative responses from the human and converge to the correct estimate more quickly.**

I. Introduction

Human goal inference is a crucial part of human-robot interaction. Imagine operating a robot arm via a joystick to pick up an object – if the robot can infer what you want to do, it can take actions that help you achieve it [1, 3, 7, 8, 10, 12]. Or imagine driving next to an autonomous car – if the autonomous car knows you want to merge, it can better prepare to accommodate that.

Work related to human goal inference falls under two broad categories. Prior work in *shared autonomy* has explored how the robot can recover a probability distribution over human operator goals by treating their inputs as observations about what they want to achieve [6, 7, 10]. The robot then plans its actions using the most probable goal [3], or in expectation over the possible goals [7, 10].

This inference is *passive*: the robot always plans with the information it currently has, assuming it will not get more information in the future. But in reality, the person will continue to provide control input, and thus information. This input that they will provide in the future will depend on what the robot does now, because the robot changes the state of the world, e.g., moving further away from the correct goal when its current estimate is wrong, triggering input in the correct direction. Because of the dependency on the robot's actions, this future human input is an opportunity for the robot, because the robot can take actions that will result in informative, disambiguating observations. This is called *active goal* inference, and has only recently been explored in shared autonomy by considering the operation *modes* that the robot is in [5], or implicitly, by computing the exact solution to a POMDP by discretizing the state space [9].

Active inference has been explicitly explored, however, in the *driving* domain [11]. But there, it has been restricted to inferring the human's *driving style* in order to better anticipate what they will do, *assuming their goal is known*. Actively inferring goals is yet to be explored in driving as well.

This paper provides an analysis of inference that is both *active* and over the space of *goals*, in both the shared autonomy and the autonomous driving domains.

We simulate different types of human input, and explore its effects on the overall system performance. One of our most interesting findings is in shared autonomy. There, under the standard observation model of human actions as Boltzmann-rational [2] – i.e., human takes approximately optimal actions given the correct goal – passive inference leads to actions that are already good with respect to information gain. However, this performance degrades under a different observation model. We explore a model in which humans actually *react* to the robot's actions, as opposed to solely providing an approximately optimal input regardless of what the robot does. We model people as providing input only when the robot is highly suboptimal, intervening to correct the robot (a model supported by findings in [4]). With this new model, we find that actively gathering information outperforms passive inference.

Overall, we find that active goal inference can sometimes be useful, but its utility depends on how people actually behave, how we model them, and what prior we start off with.

II. Active Goal Inference

We consider problems in which a robot $\mathcal{R}$ can take actions $u_\mathcal{R}$ that affect the state of the world via some dynamics model $x' = f(x, u_\mathcal{R})$. The robot wants to optimize a reward function

$$r_\mathcal{R}(x, u_\mathcal{R}; \theta_\mathcal{H}) \tag{1}$$

that can depend on a desired human's goal $\theta_\mathcal{H}$. In shared autonomy, the reward will incentivize the robot to reach the desired goal. In driving, the reward will incentivize the robot to reach its own destination and avoid collisions with the human-driven car.

The robot does not observe $\theta_\mathcal{H}$ directly, but gets observations about it in the form of human actions $u_\mathcal{H}$ (control input by the operator in shared autonomy, controls to the human-driven car in driving). The robot knows an observation model

$$P(u_\mathcal{H} | x, \hat{u_\mathcal{R}}, \theta_\mathcal{H}) \tag{2}$$

with $\hat{u_\mathcal{R}}$ the human's estimate of what the robot's action will be (which is $u_\mathcal{R}$ if the human gets to first observe $u_\mathcal{R}$ and then react). We assume that the robot also has a model of $\hat{u_\mathcal{R}}$.

This is a POMDP over x and $\theta_\mathcal{H}$ as states. Solving it would result in a policy that effectively trades off between exploiting the current information and taking actions that gather more, depending on how useful that new information would be. We are particularly interested in systems where it is not tractable to compute the exact solution to the POMDP, e.g. when we are dealing with continuous domains for x and $u_\mathcal{R}$.

A natural solution is to separate estimation and control. This leads to *passive* estimation. At every step, the robot has a current belief $b(\theta_{\mathcal{H}})$ for what $\theta_{\mathcal{H}}$ is. It takes the action that is optimal in expectation over its current estimate $b(\theta_{\mathcal{H}})$:

$$\arg\max_{u_{\mathcal{R}}} \mathbb{E}_{b(\theta_{\mathcal{H}})} Q(x, u_{\mathcal{R}}; \theta_{\mathcal{H}}) \tag{3}$$

This is only optimal if the robot will no longer receive any more observations. But in fact, after executing a control, the robot receives an observation $u_{\mathcal{H}}$ and updates its belief over $\theta_{\mathcal{H}}$:

$$b'(\theta_{\mathcal{H}}) \propto b(\theta_{\mathcal{H}}) P(u_{\mathcal{H}} | x, \hat{u_{\mathcal{R}}}, \theta_{\mathcal{H}}) \tag{4}$$

It repeats these two steps at every time step.

In contrast, we explore *active* estimation, which leverages the robot's action to gain information. Active estimation takes the action that optimizes a trade-off between exploiting and exploring:

$$\arg\max_{u_{\mathcal{R}}} \mathbb{E}_{b(\theta_{\mathcal{H}})} \left[Q(x, u_{\mathcal{R}}; \theta_{\mathcal{H}}) + \lambda(H(b) - H(b'))\right] \tag{5}$$

with $H(b)$ denoting the Shannon entropy of the distribution.

In what follows, we specialize these to shared autonomy and driving, analyzing whether active estimation is useful in these domains.

III. Shared Autonomy with Boltzmann-Rational Operators

A. Shared Autonomy Formulation

In shared autonomy, the human operates the robot to reach the goal $\theta_{\mathcal{H}}$. The human actions $u_{\mathcal{H}}$ are control inputs to the robot, which the robot can treat as observations about what the human wants [7]. The reward function incentivizes the robot to reach the human's desired goal as efficiently as possible.

A simple example of a reward, which we explore in our experiments, is to penalize the robot for actions it is taking every time it is not at the goal:

$$r_{\mathcal{R}}(x, u_{\mathcal{R}}; \theta_{\mathcal{H}}) = -\min(\|u_{max}\|, d(f(x, u_{\mathcal{R}}), \theta_{\mathcal{H}})) \tag{6}$$

where u_{max} is the robot control of maximum norm, and $d(x, y)$ is the Euclidean distance between x and y. The robot pays $\|u_{max}\|$ every time step is spends too far from the goal. When it's closer than two maximum actions away, it pays for the distance between the state it will end up in and the goal. This assumes $f(x, u_{\mathcal{R}}) = x + u_{\mathcal{R}}$, i.e. $u_{\mathcal{R}}$ is a velocity.

This reward function has a very simple value function:

$$V(x; \theta_{\mathcal{H}}) = d(x, \theta_{\mathcal{H}}) \tag{7}$$

The robot takes the maximum action towards the goal until it gets close enough, and then the action that takes it exactly to the goal, traveling the straight line to the goal in as few steps as possible.

If the robot knew $\theta_{\mathcal{H}}$, decision making would be trivial:

$$\arg\max_{u_{\mathcal{R}}} Q(x, u_{\mathcal{R}}; \theta_{\mathcal{H}}) = \begin{cases} \|u_{max}\| \frac{\theta_{\mathcal{H}} - x}{\|\theta_{\mathcal{H}} - x\|} & \text{if } d(x, \theta_{\mathcal{H}}) < \|u_{max}\| \\ \theta_{\mathcal{H}} - x & \text{otherwise} \end{cases} \tag{8}$$

Direct teleoperation addresses not knowing $\theta_{\mathcal{H}}$ by directly following the human control input $u_{\mathcal{H}}$, which works well if the person provides near-optimal input:

$$u_{\mathcal{R}}^{direct} = u_{\mathcal{H}} \tag{9}$$

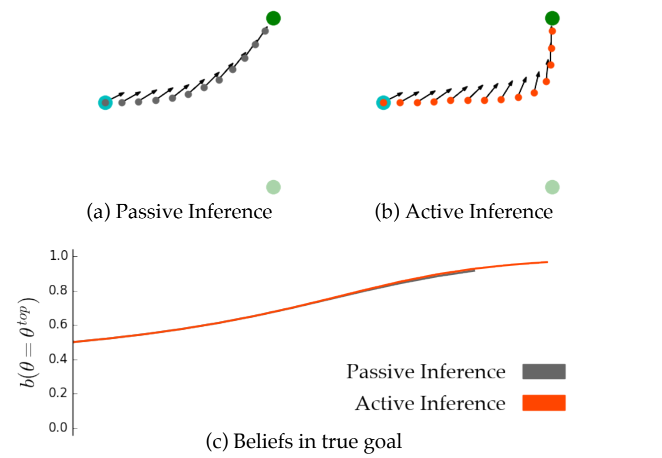


Fig. 1. **Boltzmann observation model, perfect human input.** A comparison of passive (a) and active (b) estimation with a Boltzmann-rational observation model. The beliefs (c) are very similar. Passive inference does a good job of accidentally gathering information. Active inference pulls the human toward the center of the two goals to provoke more informative actions.

Shared autonomy is useful when the human control input is noisy. In that case, rather than directly executing it, the robot aggregates it over time to figure out what the person actually wants. It uses an observation model to estimate $\theta_{\mathcal{H}}$. A classic model of human action from cognitive science [2] that has been used in shared autonomy as well [7] is that the person is Boltzmann-rational in their actions:

$$P(u_{\mathcal{H}} \mid x, \hat{u_{\mathcal{R}}}, \theta_{\mathcal{H}}) = P(u_{\mathcal{H}} \mid x, \theta_{\mathcal{H}}) \propto e^{\beta Q(x, u_{\mathcal{R}} = u_{\mathcal{H}}; \theta_{\mathcal{H}})} \tag{10}$$

This model assumes that:

- the person's input only depends on the current state and not on what they think the robot will do
- the person assumes that the robot will execute their control input directly

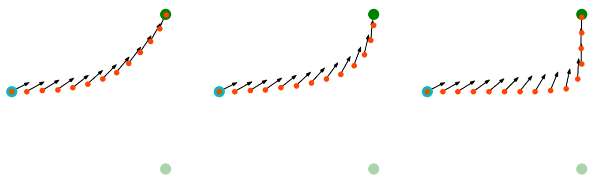
β is a temperature parameter controlling how close to optimal the human is. $\beta = 0$ assumes human control input is random, and $\beta \to \infty$ assumes a perfectly rational human.

B. Analysis

We explored active estimation in a simple simulator. The robot was a 2D point, and we had 2 possible goals as in [7]. We used the Boltzmann-rational observation model introduced above, and simulated human input that is optimal, as well as noisy human input to analyze its effects on the results.

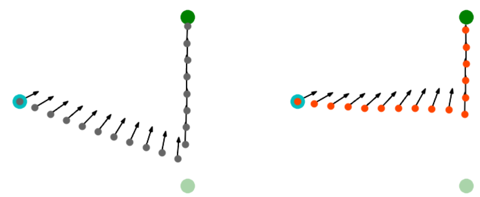
In the following experiments, the human is always trying to reach the top of two goals, indicated in green. The robot starts in the configuration indicated in blue, and its trajectory is traced out by gray (passive) or orange (active) points. The human's input is shown as black arrows. We terminate the experiment once the robot is within $\|u_{max}\|$ of the goal.

a) Passive vs. Active Estimation: We begin by comparing passive and active inference in this simple setup in Fig.1. In both cases, the robot assumes a reasonably optimal human ($\beta = 2.0$). In the active case, we set λ in (5) to 50.0, which results in a visibly different trajectory than the passive case ($\lambda = 0$). The simulated human provides the optimal input at each step.

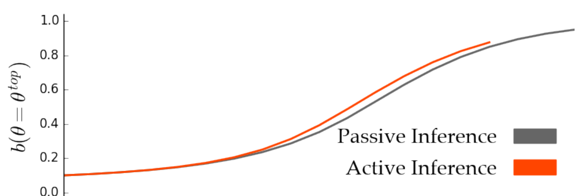


(a) $\lambda = 1$ (b) $\lambda = 10$ (c) $\lambda = 100$

Fig. 2. **Boltzmann observation model, different λ.** As λ increases, the robot cares more and more about information gain, and less about achieving the goal as quickly as possible. It pulls the human in between the two goals, where there is the most opportunity for distinguishing actions.



(a) Passive Inference (b) Active Inference



(c) Beliefs in true goal

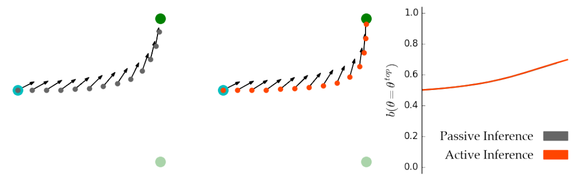
Fig. 3. **Active is useful when the prior is incorrect.** We examine starting with an *incorrect* prior, 0.1 on the true goal, 0.9 on the incorrect goal, and perform inference according to the Boltzmann-rational observation model. Active is less sensitive to the incorrect prior, and recovers from it faster than passive.

Active information gathering *pulling the human* towards the center of the two goals, at which point an optimal human's input would be most differentiated between the two goals. But passive estimation has a similar effect anyway. Planning in expectation solely over the Q-value is effective at accidentally gathering information – when the robot plans in expectation over what the goal might be, it selects a direction that works as best as possible in either case. This leads to states from which the human input is more dissimilar depending on the goal, which helps estimation as well. The belief (Fig.1.c) shows that the belief is not really impacted by not doing active estimation explicitly. Further, active estimation here results in a longer path.

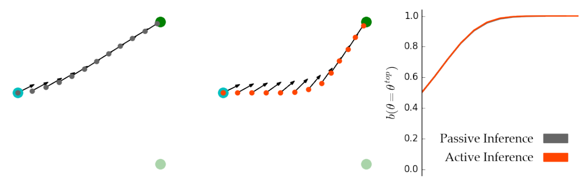
Fig.2 shows the effect of different trade-off scalars λ on the resulting behavior, where information gathering changes the resulting path more and more towards the center between the two goals.

b) Starting with a Wrong Prior: Fig.1.c, shows that the majority of the belief is on the correct goal for the entire trajectory. This might help explain the effectiveness of passive inference. Even at the first planning step, passive has a correctly biased belief (toward the top goal), because it has already observed the first human action, which breaks the 50-50 symmetry between goals. In Fig.3 we ask: "what happens with an incorrect prior?"

Here, passive over-commits to the wrong goal, before ac-



(a) Passive trajectory, active trajectory and beliefs for $\beta = 0.5$



(b) Passive trajectory, active trajectory and beliefs for $\beta = 10.0$

Fig. 4. **Assuming different levels of human noise.** We vary the temperature, β, of the Boltzmann-rational observation model, giving the robot a model of a more (a) and less (b) noisy human. The actual human input, shown as black arrows, is optimal. In the case that there is very little information to be gained, passive and active are similar. When each observation provides more information, both algorithms benefit from beliefs that collapse faster. Although beliefs are similar in (b), passive beats active for similar reasons as Fig.1

cumulating enough evidence to the contrary. How long that lasts will depend on the observation model: the more noisy we allow the human to be, the longer it will take. In contrast, active seeks out information and estimates the correct goal more quickly. This then also results in achieving the task more quickly, suggesting active has better worst-case performance.

c) Varying Human Rationality (Model): When varying the prior, we notice that it takes some time, perhaps too much time, for the beliefs to collapse. Intuition suggests that after several repeated, consistent jerks upward on the joystick, the robot should have high confidence in the top goal. This suggests that we should explore effects of modifying the temperature of the Boltzmann observation model, perhaps increasing β (model of a more optimal human).

In Fig.4 we explore decreasing and increasing the temperature constant, modeling a more noisy and more rational human than in Fig.1, respectively. The beliefs between passive and active remain similar, but reflect less confidence in the case of a lower β (Fig.4.a) and more confidence in the case of a higher β (Fig.4.b). In the case that the robot receives more informative observations, the beliefs collapse faster and active information gathering is able to more quickly commit to the goal, shortening trajectory length and task completion time. Still, passive outperforms active with respect to these two metrics in both cases.

d) Varying Human Rationality (Actual): Modeling the human as nearly-optimal is reasonable, because the simulated human is *actually* optimal. This causes passive's success in Fig.4.b, and allows active to successfully plan to gather information in all the examples so far. But what if we simulated noisier human input, and there was a mismatch between the human's actual noisiness and the model's assumed noisiness?

In Fig.5, we simulate an actually noisy human. The mismatch between the robot's model (we keep $\beta = 2$, the same as earlier in Fig.1) causes instability in the trajectories generated while performing active inference. This variance is a result of

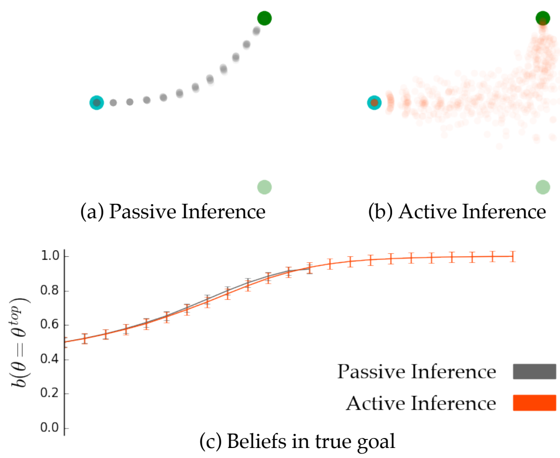


Fig. 5. **Actual human noise.** We simulate a human that is actually approximately optimal, by sampling the human control from a Gaussian with variance 0.01, centered at the optimal control. For reference, $\|u_{max}\| = 0.1$. We plot the results of 30 replications. The bars in (c) indicate the standard error of the plotted (mean) belief. Passive's trajectory is stable in the face of this noise, whereas active's varies greatly.

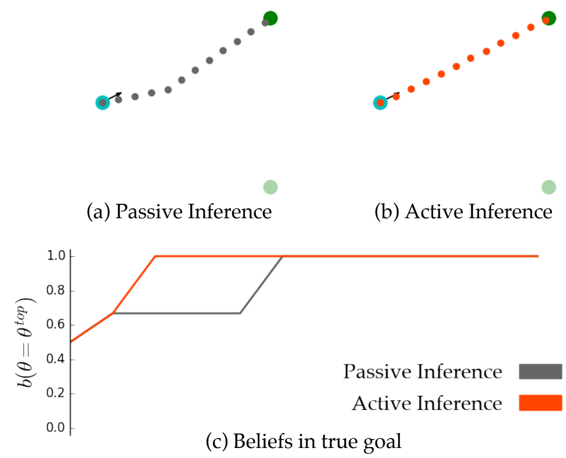


Fig. 6. **Corrective observation model.** A comparison of passive (a) and active (b) estimation with a corrective observation model. The beliefs (c) diverge between the two inference algorithms for several time steps. Active inference commits to a goal early, testing the human for feedback, whereas passive hedges. We set a modest operator threshold, which is why passive can collapse the belief at time step 4, when it is not corrected for taking an action that is out of the threshold for optimality with respect to the bottom goal.

active's dependence on the observation model in planning. We need an accurate model of the human in order to adequately plan, especially when trying to actively gather information. Passive inference is less affected by the truly noisy human inputs, and by the mismatch between assumed and actual rationality of the human.

Overall, shared autonomy with a Boltzmann observation model works well so long as it does not start off with a wrong prior. Actively gathering information is rarely necessary in such a system, and usually decreases the robot's reward in the task.

IV. Shared Autonomy with Operators that *React* to the Robot's Actions

A. A Corrective Observation Model

So far we have used a Boltzmann-rational observation model of the human's actions. Although common, this model implicitly assumes that the operator is unaffected by the robot's actions – they are simply providing the optimal input regardless of where the robot is, assuming the robot will follow that input exactly, despite the fact that it hasn't been doing this so far.

Of course, in reality people notice the robot is not following their input, and thus do not expect it to all of the sudden start doing that at the next time step [4]. They provide different inputs depending on what the robot does.

If the robot is moving toward the correct goal, people sometimes let off the control, providing no input, assuming the robot knows what it's doing. If instead the robot is moving toward an incorrect goal, the human pulls back and toward the correct goal, as if to counteract the robot's action and indicate the correct action.

To account for this behavior, we formulate a new observation model with two regimes: (1) when the human is applying no input because the robot's action looks almost optimal according to $\theta_{\mathcal{H}}$, and (2) when the human is applying a correction because the robot's action was suboptimal:

$$P(u_{\mathcal{H}} \mid x, \hat{u_{\mathcal{R}}}, \theta_{\mathcal{H}}) = \begin{cases} P_{\text{watch}} & \text{if } Q(x, \hat{u_{\mathcal{R}}}, \theta_{\mathcal{H}}) \geq Q_{max} - \delta \\ P_{\text{fix}} & \text{otherwise} \end{cases} \tag{11}$$

where $Q_{max} = \max_{u_{\mathcal{R}}} Q(x, u_{\mathcal{R}}, \theta_{\mathcal{H}})$ and δ is some threshold for deciding how close to optimal the robot's predicted action is. We assume people predict the robot's action to be the same as its previous actions, i.e. $\hat{u_{\mathcal{R}}}^t = u_{\mathcal{R}}^{t-1}$.

P_{watch} and P_{fix} define the observation model in the two regimes.

$$P_{\text{watch}}(u_{\mathcal{H}} \mid x, \hat{u_{\mathcal{R}}}, \theta_{\mathcal{H}}) = \begin{cases} 1 - \epsilon & \text{if } u_{\mathcal{H}} = 0 \\ \epsilon & \text{otherwise} \end{cases} \tag{12}$$

where ϵ is selected to be small, but $\neq 0$ so that we do not annihilate a belief. We model the human as almost certainly applying no input if the last robot control was approximately optimal.

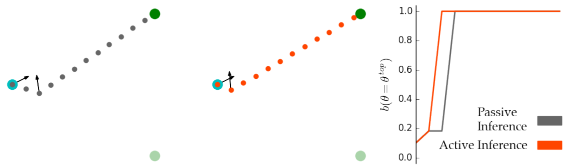
In the case that the robot's last action was suboptimal according to δ and $\theta_{\mathcal{H}}$, we model them as both counteracting the last robot's control and applying the optimal control otherwise (noisily).

$$P_{\text{fix}}(u_{\mathcal{H}} \mid x, \hat{u_{\mathcal{R}}}, \theta_{\mathcal{H}}) \propto \mathcal{N}(-\hat{u_{\mathcal{R}}} + u_{\mathcal{H}}^{t,*}, \sigma^2)$$

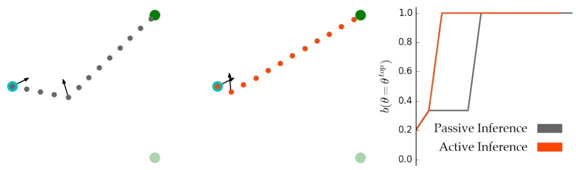
In one regime, when the robot is issuing correct controls, the human is "content" and only "watching." In the other regime, they are "attentive" and "fixing" the robot's action.

B. Analysis

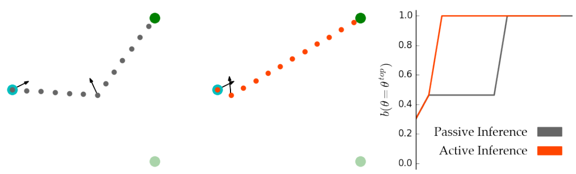
As earlier, we simulate a point robot in a 2D plane, with two goals. The correct goal remains the top one.



(a) Passive, active trajectory and beliefs for $b^0(\theta) = (0.1, 0.9)$



(b) Passive, active trajectory and beliefs for $b^0(\theta) = (0.2, 0.8)$



(c) Passive, active trajectory and beliefs for $b^0(\theta) = (0.3, 0.7)$

Fig. 7. **Corrective observation model with wrong prior.** A comparison between varying degrees of incorrectness in the prior under a corrective observation model. Passive's unwarranted confidence leads it to prompt corrections at different points in its trajectory. A large incorrect prior in the passive case mimics active's strategy of committing to a goal early on. The resulting information gain is a *happy coincidence* for passive inference in (a), whereas active estimation *explicitly plans* for it in every case.

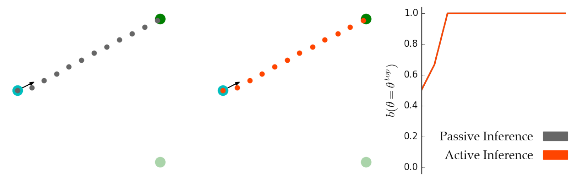
a) Passive vs Active Estimation: In Fig.6, we compare the inference algorithms under the corrective observation model. The ability of the robot to *affect the human's action* differentiates active and passive inference. Active *commits* to the top goal, after its initial belief is biased by the human's first (and only) control. This is because when the human does not correct, the belief distribution collapses.

Passive, because it does not plan to gather information, does not realize that taking an action toward one of the goals will completely collapse the belief, and therefore hedges for a few time steps before happening into a state where its action is suboptimal with respect to the bottom goal, receiving a correction, and switching course.

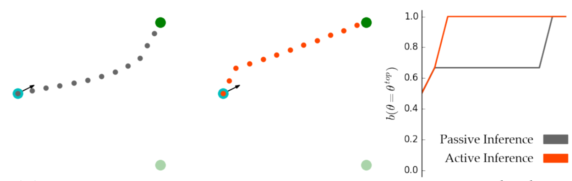
Active information thus achieves the task optimally, despite initial uncertainty.

b) Starting with a Wrong Prior: We also explore the effect of varying the prior. Fig.7 demonstrates how passive's trajectory changes significantly with the degree of error in initial beliefs, whereas active's remains constant. The passive robot's confidence causes it to stumble, after some number of time steps, into being suboptimal with respect to the true goal, warranting a correction from the human. Counterintuitvely, the *more incorrect* the prior, the *earlier* passive discerns its initial mistake in beliefs, because the earlier it gets a correction.

Active's trajectory is consistent across these priors, and in our experiments, all incorrect priors. As before, it commits to the goal holding a majority belief, to an extent that if this goal were wrong, the human would supply a correction. In each case of wrong priors it receives the correction, and the beliefs collapse in one step.



(a) Passive trajectory, active trajectory and beliefs for low δ



(b) Passive trajectory, active trajectory and beliefs for high δ

Fig. 8. **Human threshold for intervening.** A comparison of a more strict (a) and more liberal (b) operator threshold for optimality (δ in (12)). In the case of a more corrective human (a), passive and active are identical. When planning for a less strict human (b), active estimation goes out of its way looking for a correction. Active is able to plan well, *regardless* of operator threshold.

c) Varying Operator Threshold for Optimality: The operator's threshold for optimality clearly affects the trajectories of Fig.6 and Fig.7. In Fig.8 we explore the result of both decreasing and increasing the threshold. A lower threshold, modeling a human more apt to correct the robot, helps passive inference, and beliefs are indistinguishable between the two algorithms.

For a less attentive, lazier human with a larger threshold, passive suffers. Active acts as we saw earlier, except now it exaggerates its preference toward one goal, deliberately choosing an action which is suboptimal with respect to the bottom goal. Again, when it receives no corrections from the human, its beliefs collapse. A larger threshold does not prevent active from collapsing the beliefs as fast as with the lower thresholds we have seen so far. Active *compensates* for the threshold.

Overall, when the person is providing corrective input to the robot, active estimation leverages this and commits to a goal, knowing that if it is wrong it will get a correction. This leads to more efficient behavior.

V. Autonomous Driving

In the driving domain, the car interacts not with an operator, but with a human-driven vehicle. The actions $u_{\mathcal{H}}$ are no longer just observations, they actually affect physical state.

Autonomous cars usually plan using Model Predictive Control (MPC), i.e. use a short time horizon and replan at every step. Thus, we can treat $u_{\mathcal{R}}$ as a sequence of controls of length T, with T the MPC time horizon.

Similarly, $u_{\mathcal{H}}$ is a sequence of human controls of length T. The robot's reward is to reach its own destination and avoid collisions with the human. Critically, this means that $r_{\mathcal{R}}$ is a function of $u_{\mathcal{H}}$, because the robot needs to know the person's trajectory in order to evaluate whether it stayed far enough away. As in [11], we model the human as optimizing their

(a) Merge Example (b) Simulation

Fig. 9. An example of a car desiring to merge (a) and our discretization of the goal space: hold lane or merge. In our simulation (b), the human (white car) desires to merge in front of the robot (red car), but the robot does not know this.

own reward in response to $u_{\mathcal{R}}$:

$$u_{\mathcal{H}}(x, u_{\mathcal{R}}; \theta_{\mathcal{H}}) = \arg\max_{u} r_{\mathcal{H}}(x, u_{\mathcal{R}}, u; \theta_{\mathcal{H}}) \tag{13}$$

This takes us back to the robot's reward only depending on x, $u_{\mathcal{R}}$, and $\theta_{\mathcal{H}}$. Note that because actions are trajectories for the time horizon, $r_{\mathcal{H}}$ is actually cumulative reward over that horizon.

We use a Boltzmann-rational observation model again, looking at the reward that the human accumulates:

$$P(u_{\mathcal{H}} \mid x, u_{\mathcal{R}}, \theta_{\mathcal{H}}) \propto e^{\beta r_{\mathcal{H}}(x, u_{\mathcal{R}}, u_{\mathcal{H}}; \theta_{\mathcal{H}})} \tag{14}$$

Since the human has to keep driving, the corrective observation model we explored in shared autonomy no longer applies.

Thus, $u_{\mathcal{H}}$ serves two roles: it changes the physical state of the world, and it provides observations about $\theta_{\mathcal{H}}$. The robot uses a deterministic model to plan accounting for the former, and a probabilistic model to estimate the latter.

A. Analysis

We consider a merge scenario, as it captures the disambiguation task and exemplifies an everyday situation forcing drivers to perform inference. The robot thinks the human has one of two goals: merge or hold lane (see Fig.9). The trajectories and beliefs for the case when the human wants to merge, and the robot starts with a uniform prior, are shown in Fig.10.

The beliefs for the active and passive algorithms separate considerably. In the passive case, the robot slowly begins to uncover the humans intent from the human's slight nudging. It takes a good deal of time, however, for these beliefs to converge. It certain contexts, such as allowing somebody to merge for an exit, it could take too long. Active inference, on the other hand, pushes the beliefs to collapse earlier. It gathers information by *planning to influence* the human through its model of the human's response. Even more interesting is the *behavior which emerges out of the optimization*, and its surprising resemblance to typical human behavior.

Specifically, the robot takes actions which *probe the human*. It brakes and slows, behavior communicating: "would you like to merge?" In the example shown in Fig.10, the human does wish to merge, and completes this goal successfully. If the human responded by not moving into the lane, the robot would interpret the slight nudging over as noise, and the belief distribution would collapse the other direction.

(a)

(b)

(c)

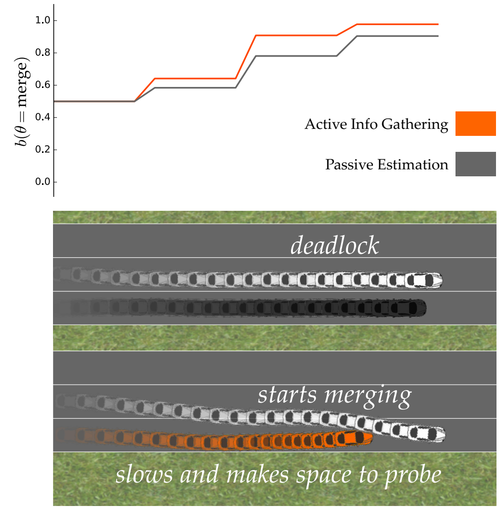


Fig. 10. The belief distributions (a) and corresponding behavior (b, c) for the merge scenario where the human (white car) does wish to merge. In the first case (b) the robot is performing passive inference. In the second case (c) the robot is performing active inference.

VI. Discussion

This paper summarized our exploration into the effects of actively estimating human goals in shared autonomy and driving. We found that active estimation can be especially useful when we start with the wrong prior over the human goals, or in shared autonomy when the operator only provides corrections if the robot is too suboptimal. In such cases, active estimation means that the robot can put the operator in a situation where they either correct the robot, which is highly informative, or the robot is already doing the correct task.

Our work is really a step in this analysis, since our experiments were for particular and rather simple scenarios. We also opted for simulated models of people in order to tease out effects such as the amount of user noise in deciding user controls, but in future work we must test active estimation with real users.

Overall, we believe that robots should not always passively update their beliefs about human goals, but they should sometimes leverage their actions to probe for more information. Much work remains to understand when and to what extent this should happen, e.g., an autonomous car should not always slow down when someone passes. Active probing should ideally be based on other criteria, such as the value of information, rather than optimizing for information gain for the sake of it.

Acknowledgments

We thank Davis Foote and Dorsa Sadigh for their helpful advice. This work is supported by Berkeley Deep Drive and NSF CAREER #1652083.

References

[1] Henny Admoni and Siddhartha Srinivasa. Predicting user intent through eye gaze for shared autonomy. In *2016 AAAI Fall Symposium Series*, 2016.

[2] Chris L Baker, Rebecca Saxe, and Joshua B Tenenbaum. Action understanding as inverse planning. *Cognition*, 113(3):329–349, 2009.

[3] Anca Dragan and Siddhartha Srinivasa. Formalizing assistive teleoperation. In *Robotics: Science and Systems*, Pittsburgh, PA, July 2012.

[4] Anca Dragan and Siddhartha Srinivasa. A policy blending formalism for shared control. May 2013.

[5] Deepak Edakkattil Gopinath and Brenna Argall. Mode switch assistance to maximize human intent disambiguation. In *Robotics: Science and Systems*, July 2017.

[6] Kris Hauser. Recognition, prediction, and planning for assisted teleoperation of freeform tasks. *Autonomous Robots*, 35(4):241–254, 2013.

[7] Shervin Javdani, Siddhartha Srinivasa, and J. Andrew (Drew) Bagnell. Shared autonomy via hindsight optimization. In *Proceedings of Robotics: Science and Systems*, Rome, Italy, July 2015.

[8] Hema S Koppula and Ashutosh Saxena. Anticipating human activities using object affordances for reactive robotic response. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):14–29, 2016.

[9] Stefanos Nikolaidis, Yu Xiang Zhu, David Hsu, and Siddhartha Srinivasa. Human-robot mutual adaptation in shared autonomy. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '17, pages 294–302, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4336-7. doi: 10.1145/2909824.3020252. URL http://doi.acm.org/10.1145/2909824.3020252.

[10] Stefania Pellegrinelli, Henny Admoni, Shervin Javdani, and Siddhartha Srinivasa. Human-robot shared workspace collaboration via hindsight optimization. In *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, pages 831–838. IEEE, 2016.

[11] Dorsa Sadigh, S Shankar Sastry, Sanjit A Seshia, and Anca Dragan. Information gathering actions over human internal state. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 66–73. IEEE, 2016.

[12] Brian D Ziebart, Nathan Ratliff, Garratt Gallagher, Christoph Mertz, Kevin Peterson, J Andrew Bagnell, Martial Hebert, Anind K Dey, and Siddhartha Srinivasa. Planning-based prediction for pedestrians. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 3931–3936. IEEE, 2009.