Robotics: Science and Systems 2022
New York City, NY, USA, June 27-July 1, 2022

Gaze Complements Control Input for Goal Prediction During Assisted Teleoperation

Reuben M. Aronson
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA, USA
rmaronson@cmu.edu

Henny Admoni
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA, USA
henny@cmu.edu

***Abstract*—Shared control systems can make complex robot teleoperation tasks easier for users. These systems predict the user's goal, determine the motion required for the robot to reach that goal, and combine that motion with the user's input. Goal prediction is generally based on the user's control input (e.g., the joystick signal). In this paper, we show that this prediction method is especially effective when users follow standard noisily optimal behavior models. In tasks with input constraints like modal control, however, this effectiveness no longer holds, so additional sources for goal prediction can improve assistance. We implement a novel shared control system that combines natural eye gaze with joystick input to predict people's goals online, and we evaluate our system in a real-world, COVID-safe user study. We find that modal control reduces the efficiency of assistance according to our model, and when gaze provides a prediction earlier in the task, the system's performance improves. However, gaze on its own is unreliable and assistance using only gaze performs poorly. We conclude that control input and natural gaze serve different and complementary roles in goal prediction, and using them together leads to improved assistance.**

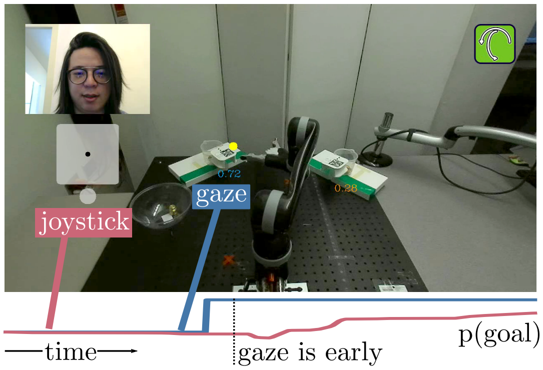


Fig. 1: A user controls a robot with a joystick to pick up a mug, while their eye gaze behavior is captured. Eye gaze gives information about the user's goal earlier than the joystick information does, which makes it appealing for incorporation into assistive systems.

I. Introduction

Teleoperation is often used to control robots, but performing complex tasks in this way is difficult. Limited interfaces, complex kinematics, and the lack of proprioception turns tasks easily performed by hand into exercises in frustration. Shared control can make the problem easier. These systems [12, 12, 22, 30, 34, 35, 47] often work by predicting the user's goal, planning to accomplish that goal, and combining the autonomous command with the user input.

Typically, shared control systems rely on the user's control input, like joystick motion, for goal inference [8, 24, 37, 44, 46]. When the system observes that the user is working towards a particular goal, the system can then assist towards that same goal. While this method does not necessarily provide the earliest predictions [4], user input works well for assistance, since accurate predictions arrive more often exactly when they are needed. When the user input differentiates between goals, the system has enough information to give goal-specific assistance. When all goals require the same motion, the user input does not help the system to predict the user's goal, but no goal prediction is actually needed. In fact, we can make a more formal claim: when a user controlling a shared autonomy system [19] provides control input given by $p(u|g) \propto \exp(Q_g(u))$, the expected regret over user actions stays bounded as the cost of taking a suboptimal action increases. We formalize and prove this result in Sec. III.

However, users often do not follow this optimal behavior. Specifically, the scenario itself can prevent the user from acting optimally. Consider a goal that can be split into multiple tasks that the robot can perform in parallel, e.g., splitting the goal of moving its end-effector to a desired pose into six independent dimensions of motion. The above analysis relies on the system knowing each task individually. However, the structure of the task itself may prevent the user from working on all of them: for example, modal control restricts users to giving only two directions of end-effector motion at a time. Then, a user working on one task and not another will not give sufficient information to enable full assistance.

For successful assistance in these cases, we must consider other sources of goal prediction: in this work, we incorporate the user's natural eye gaze. While people manipulate objects, they look at their goals before reaching them [14, 21] and look forward to next steps in their tasks [29]. These patterns also appear during teleoperated manipulation [2, 5, 11] and can be used for goal prediction globally through the task, whatever its current state [4, 33]. However, gaze is noisy and somewhat unreliable, making it a poor choice to use on its own. Thus, it is best used to provide the global information that complements predictions based on control input to increase the amount of assistance provided.

In this work, we implement an assistive teleoperation system that incorporates goal prediction using both the user's control input and their eye gaze behavior. We use this system to evaluate each prediction source in a real-world, COVID-safe user study. In the study, participants teleoperated a robot manipulator using modal control to pick up one of two cups while our system provided assistance. The scenario was designed so that the user could not act optimally, so their control input was unlikely to yield optimal assistance. During each trial, the assistance relied on goal prediction based on their joystick input, their gaze behavior, or both.

We find that for this experimental scenario, assistance based on joystick input alone is delayed relative to using both joystick and gaze, but only when the gaze prediction arrives sufficiently early. In the cases with early gaze predictions, trials finished more quickly and users supplied less control input. Specifically, early gaze leads to earlier assistance exactly on the axes for which the goal positions differ, and the assistance is the same otherwise, matching our theoretical analysis. However, gaze-based predictions are inherently less reliable, as many trials never gave sufficient information for accurate goal prediction, and feedback loops led to arbitrarily poor performance in some cases. This work explores a fundamental limitation of input-based goal prediction for assistance and shows that eye gaze provides the global information required for systems to provide as much assistance as possible.

II. Related Work

A. Assisted Teleoperation

Assisted teleoperation, in which a system predicts the user's intent, plans autonomously to achieve that intent, and combines its generated command with the user's direct input, has been widely studied [28]. Our work builds most directly on Javdani et al. [17, 18], which models assistance as a partially observable Markov decision process, with partial observability over the user's goal choice; this model has shown success in various iterations and applications [12, 22, 30, 34, 35, 47]. This structure enables the system to generate an assistance command even with no knowledge of the user's goal when the system can make progress towards all goals simultaneously. This work poses goal inference as an inverse reinforcement learning problem by assuming a noisily optimal human model, which is frequently built upon [8, 24, 37, 44, 46]. To make the joystick input more predictive of the goal, Gopinath and Argall [13] has the joystick start in a control mode such that the user can immediately perform goal-specific motion.

An assistive system can combine predictions from different sources, such as user input with gaze (proposed by Admoni and Srinivasa [1]). Jain and Argall [16] proposed combining multiple predictions by assuming each is independent conditioned on the goal, which we use. Structural challenges to effective shared autonomy have also been identified in Fontaine and Nikolaidis [10].

B. Gaze for Intent Prediction

During manual manipulation, people look at their targets before reaching towards them. Hayhoe [14] reports that 87% of reaching movements in a sandwich-making task were accompanied by target-directed fixations. These *directing* fixations [25, 29] indicate the actor's intention to interact with an object. A number of works have used gaze to predict people's goals and tasks [7, 9, 23, 43] during manual manipulation.

During teleoperation, however, gaze behavior changes. While gaze often predicts people's goals and tasks accurately [4, 11, 33, 42], the introduction of a robot causes challenges. The gaze signal can be noisy and difficult to align to the scene [3]. Unlike the largely goal-directed gaze during manual manipulation, people often look at the robot itself [2, 4, 5, 11, 33]. Worse yet, people can complete tasks without ever looking at their goal, especially when repeating the task [2, 4]. By analyzing offline data of gaze while operating a robot, Razin and Feigh [33] finds that task prediction using robot motion is more accurate than gaze alone, and adding gaze to the robot motion signal does not improve overall prediction performance. The difficulty of using gaze motivates our work to understand how to use gaze effectively.

C. Gaze in the Loop

Many systems use *intentional* eye gaze as a control input to a robotic manipulation system [6, 26, 27, 36, 40, 41]. Instead, we focus on people's *natural* gaze behavior, which emerges automatically while they execute a task. Huang and Mutlu [15] used people's natural eye gaze while selecting a menu item to anticipate their selection and move a serving robot, which improved performance. In Stolzenwald and Mayol-Cuevas [38], participants play a screen-based tile placing game using a robotic pointer; using natural eye gaze to predict people's targets so the robot can assist outperformed using the prediction to hinder the user, but it did not show any improvement over taking no action.

III. When Control Input Is Not Enough

Consider a user teleoperating a robot to pick up an object (Fig. 1). Grasping tasks like this are difficult, especially when using basic interfaces such as joysticks. To make the task easier, *shared control* [18] predicts the user's goal among a pre-specified set of goal candidates, plans to achieve the goal, and combines this autonomous command with the user command. Shared control systems [28] often use the joystick input itself to infer the user's goal. We explore the joystick signal and propose criteria for when another signal, such as eye gaze, will lead to better assistance.

A. Joystick-based Prediction and Assistance

In this section, we summarize the approach for goal prediction and assistance given in Javdani et al. [19]. This method uses the user's control input u to predict their goals, expressed as a probability distribution $p(G)$ over a pre-specified set of goal candidates. To do so, it frames goal inference as an inverse reinforcement learning problem [17, 20, 44, 45] and models the teleoperation problem as a family of Markov decision processes (MDPs) with different, pre-specified cost functions $C_g(x, u)$ for each goal candidate $g \in G$.[1] The system then assumes that the user is noisily optimizing the cost function corresponding to their true goal.

First, this method solves the Bellman equation for each goal MDP for a goal-specific action value function $Q_g(x, u)$. Then, it assumes that the user's action u at each state x is drawn from a distribution given as

$$p(u|x, g) \propto \exp(Q_g(x, u)). \tag{1}$$

Note that this is equivalent to the Boltzmann rational model with $\beta = 1$. Given a sequence of state-action pairs $\xi = (x_0, u_0, \cdots, x_n, u_n)$, the strategy assumes that the user's actions are conditionally independent given their goal. Since ξ is not a trajectory, as the robot will be acting simultaneously with the user, the method treats only the actions u_i as observations. Using Bayes' rule, it aggregates a goal prediction over time using

$$p(g|u_0, \cdots, u_i) = \frac{p(u_i|g)p(g|u_0, \cdots, u_{i-1})}{\sum_{g'} p(u_i|g')p(g'|u_0, \cdots, u_{i-1})}. \tag{2}$$

To generate an assistance signal from the goal prediction, this method represents the combined robot-human control problem as a partially observable Markov decision process (POMDP), with the user's goal a hidden parameter. The POMDP augments the system state x with a belief distribution over the user's goal given by $p(g)$ above. The action value function $Q(x, p, a)$ depends on the robot state, next action, and the belief state. Since solving the POMDP is generally computationally prohibitive, it adopts the hindsight optimization assumption, which assumes that the uncertainty expressed by $p(g)$ will resolve in the next step. From here, we can find the optimal assistance policy $\psi(x, p(g))$:

$$\psi(x, p(g)) = \arg\max_{a \in A} \sum_g p(g) Q_g(a). \tag{3}$$

This assumption replaces the overall value function of the POMDP with the expectation over the goal probabilities of the goal-specific value functions, and it reuses the goal-specific value functions $Q_g(u)$ used in Eqn. 1. (We use a here to represent that this action is selected by the robot, as opposed to u which is given by the user.) To compute the overall motion, sum a^* with the user command u directly: $a_{\text{exec}} = a^* + u$.

B. Evaluating Prediction Sources

While accuracy and forecast horizon are useful measures to evaluate a prediction of the user's goal, we want to evaluate the assistive system as a whole. Accurate predictions only matter when they improve the quality of the assistance provided. Whatever its metrics as a prediction, the user's control input is effective for assistance, since the signal is directly tied to the generation of the assistance command.

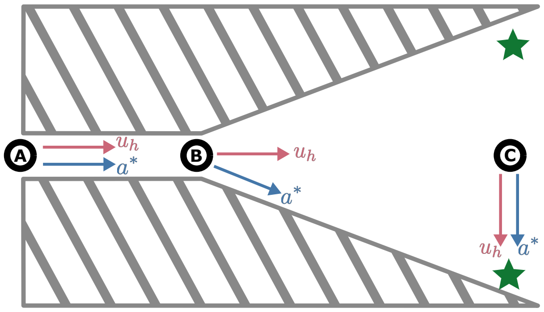


Fig. 2: Diagram of user input (u) and optimal robot motion (a^*) during an example task. The user moves a point robot to one of the green stars. At **A**, user input and optimal motion are both to the right. At **B**, user input is still directly to the right, but the optimal motion is diagonally towards the goal. At **C**, both the user input and the optimal motion point towards the goal. Early prediction improves task performance only at **B**.

To explore this coupling between assistance and goal prediction, we start with an example. A planar robot task is shown in Fig. 2. The user must move the point robot from **A** to one of the two goals (green stars). At **A**, the only way to make task progress for either goal is to move to the right. The user's expected input is the same for each goal, so it does not yield a goal prediction. However, no prediction is necessary: knowledge of the goal would not change the optimal motion. At **C**, the situation is reversed. The optimal motion is to move either up or down directly towards the user's goal. Here, the system requires a goal prediction to assist. As the user's input depends on the goal, though, the prediction is available.

Location **B** is different. Say the user continues moving to the right, which gives no goal information. However, the system can do better. If it knew the goal, it could move diagonally; without goal knowledge, however, it must wait until observing a goal-dependent user input (like at **C**) before it can assist along the vertical direction. Early, independent goal prediction only improves assistance at points like **B**, where goal information would change the motion but the user input does not provide it.

To formalize this analysis, consider an assistive system with two goal candidates $\{g_1, g_2\}$.

- Two goals require *different* motion at x if their optimal *robot* motion a^* depends on the goal: $a_1^*(x) \neq a_2^*(x)$; the motion is *identical* otherwise.
- Two goals are *distinguishable* at x if the observed *user* input generally differs based on the goal: $u(x|g_1) \neq u(x|g_2)$; they are *indistinguishable* otherwise.

Identical and different motion are properties of the robot's state, whereas distinguishable and indistinguishable goals are determined by the user's input. When the user is acting near-optimally, different motion likely leads to distinguishable goals. Next, we formalize this alignment between optimal users and effective assistance.

[1] Following Javdani et al. [19], the cost function was constant outside a radius of the goal and declined linearly to 0 at the goal location.

C. Noisily Optimal Input Bounds Regret

The above analysis suggests that when users give approximately optimal input, the system will likely receive the information needed to provide assistance. If we assume the user follows the model given in Eqn. 1, we can evaluate the expected performance of the shared autonomy policy given in Eqn. 3. We show that as the importance of taking the optimal robot action (measured by regret) increases, the user's probability of providing a distinguishing input increases faster, such that the overall system has bounded regret.

For simplicity, assume we only have two goal candidates with action value functions Q_1 and Q_2, and assume without loss of generality that the user's goal is g_1. We also assume that the set of actions A is finite and identify actions with the same $Q(a)$. At some state x (which we drop for ease of notation), let Q_1^* be the maximum value of $Q_1(a)$ attained at some action a_1^*. If we define the goal probability from control input u as above, the shared autonomy policy $\psi(p(g))$ is a function of u and we write $\psi(u)$. We can then compute the expected regret $R(\psi(u)) = Q_1^* - Q_1(\psi(u))$ of the assistance policy $\psi(u)$ over the user model.

We can measure the importance of taking a_1^* over any other action a' by letting R_{min} represent the minimum regret over all alternative actions:

$$R_{\text{min}} = \min_{a \neq a_1^*} R(a).$$

We want to understand the behavior of the system as R_{min} increases. Increasing R_{min} can be achieved by changing the selected state or the MDP itself. For example, consider an MDP with reward function $r(x)$. If we scale that reward function, $r'(x) = \lambda r(x), \lambda > 0$, the value function scales similarly, $Q'(x,a) = \lambda Q(x,a)$. Then, $R'_{\text{min}} = \lambda R_{\text{min}}$, and we can then consider the behavior as λ increases. Similar effects can also occur by changing x or $r(x)$ in other ways that are more complicated to formulate. However the change occurs, increasing values of R_{min} represent increased importance of taking the optimal action.

We can now determine the expected regret of the assistance policy under a user following Eqn. 1.

Proposition.

$$\lim_{R_{\text{min}} \to \infty} E_u[R(\psi(u))] = 0. \tag{4}$$

We sketch a proof in two parts. First, we show that as $R_{\text{min}} \to \infty$, the assistance action taken when observing the optimal action from the user, $\psi(a_1^*)$, becomes a_1^*:

$$\lim_{R_{\text{min}} \to \infty} \psi(a_1^*) = a_1^*.$$

By manipulating Eqn. 3 and collecting terms in $p(g)$, we find that for $\psi(a_1^*) = a_1^*$, we must have, for all $a' \in A$,

$$p(g_1|a_1^*)(Q_1(a_1^*) - Q_1(a')) \geq p(g_2|a_1^*)(Q_2(a') - Q_2(a_1^*)).$$

The left-hand side is greater than $p(g_1|a_1^*)R_{\text{min}}$ which increases as $R_{\text{min}} \to \infty$, while the right-hand side is nonincreasing through $p(g_2|a_1^*)$. Once the importance of taking the optimal action exceeds some threshold, the assistance will take that optimal action whenever it observes it from the user.

The expected regret is given by

$$E_u[R(\psi(u))] = \sum_u R(\psi(u))p(u|g_1).$$

From above, once R_{min} is sufficiently large, $R(\psi(a_1^*)) = R(a_1^*) = 0$. We can therefore break a_1^* out of the sum. If we define $R_{\text{max}} = \max_a R(a)$ analogously, we have

$$\begin{aligned} E_u[R(\psi(u))] &= R(\psi(a_1^*))p(a_1^*|g_1) + \sum_{u \neq a_1^*} R(\psi(u))p(u|g_1) \\ &= \sum_{u \neq a_1^*} R(\psi(u))p(u|g_1) \\ &\leq R_{\text{max}} p(u \neq a_1^*|g_1). \end{aligned}$$

Finally, we bound the probability of the user giving an action other than the optimal action based on our model of user behavior,

$$\begin{aligned} p(u \neq a_1^*|g_1) &= \frac{\sum_{u \neq a_1^*} \exp Q_1(u)}{\exp Q_1^* + \sum_{u \neq a_1^*} \exp Q_1(u)} \\ &= \frac{\sum_{u \neq a_1^*} \exp(-R(u))}{1 + \sum_{u \neq a_1^*} \exp(-R(u))} \\ &\leq \frac{(|A|-1)\exp(-R_{\text{min}})}{1 + (|A|-1)\exp(-R_{\text{min}})}. \end{aligned}$$

Putting it all together,

$$E_u[R(\psi(u))] \leq \frac{R_{\text{max}}}{1 + \frac{1}{|A|-1}\exp(R_{\text{min}})}.$$

As long as R_{max} increases less than exponentially with R_{min}, the result goes to 0 as $R_{\text{min}} \to \infty$ and the regret is bounded. This condition is met by uniformly scaling the reward as described earlier. □

As the importance of taking the optimal action increases, the chance of the user performing that optimal action under the model increases exponentially faster, so the system is more likely to receive the information it needs.

D. Control Input Restrictions Require New Prediction Sources

We see from the previous result that noisily-optimal users are particularly easy to assist using input-based goal prediction. If we remove the assumption of optimality — by assuming, e.g., that the user acts randomly, mistakenly, or adversarially — we no longer have guarantees that the assistance will behave well. However, there is a large class of problems for which the user still acts optimally but the assistance can be arbitrarily ineffective: when the user's action are limited to only a subset of the actions that the system can take.

> *It is not the user's suboptimality that limits the effectiveness of the system, but the constraints that the system itself puts on the user's behavior.*

One common example of this problem in teleoperation is the use of modal control. In this scheme, the robot can control its end-effector simultaneously in all directions. However, the user has only a 2-D joystick with which to control the robot. The user can fully control the robot by cycling through modes with the joystick controlling x/y, z/yaw, and pitch/roll in turn. If the optimal action does not align with a single control mode, the user cannot perform it. The best the user can do is to provide input in the single most useful mode. And when the robot motion is different but the control input *within the optimal mode* is not distinguishing, assistance does not have enough information to be optimal.

We can return to Fig. 2 to explore this limitation further. At **B**, we observe the user giving indistinguishable motion, though the assistance requires different motion per goal. In the noisily rational model, this user action occurs at a lower probability than a distinguishable input. However, if we add the additional restriction that the user can only provide axis-aligned commands, the user's input at **B** is optimal. Even with an optimal user, the assistance does not receive enough information to provide full assistance. In these situations, the system benefits from an alternative, global method for goal prediction that is less reliant on the user's local behavior. While an alternative information source will not remove the direct restrictions of modal control, it can bypass the limitations in goal information forced by the control restriction and improve overall system performance.

IV. Gaze-based Goal Prediction

To provide goal prediction when motion differs but input is indistinguishable, we use natural eye gaze. Gaze provides a *global* goal prediction which is less dependent on the state of the task, and people's gaze often anticipates future tasks while their actions focus on the current one.

Systems using *intentional* gaze behavior typically select the goal closest to the user's gaze location and implicitly rely on the user to adjust their gaze to provide accurate information [6, 26, 27, 36, 40, 41]. However, natural gaze is not so reliable. While gaze relates to the user's intentions, most gaze is directed towards the robot end-effector, and people can complete robotic manipulation tasks without ever looking at their goals [2, 5, 33]. These complications require more complex prediction strategies.

To predict the user's goals from their natural gaze, we adapt the sequential method given in Aronson et al. [4]. This method has two stages of gaze processing: (1) semantic gaze labeling, which segments the raw gaze into individual fixations and labels each fixation with its corresponding scene keypoint; and (2) sequential goal prediction, which uses a pre-trained hidden Markov model to yield goal probabilities from this sequence.

A. Semantic Gaze Labeling

Raw gaze data is captured as a 90Hz time series of pixel locations. This signal is segmented into individual fixations, during which the user's object of focus remains fixed[2], using a variant of the I-BMM algorithm [39]. Next, each fixation is matched with an object in the scene based on proximity. In this task, candidate objects included one for each goal, one for each robot joint, and one representing the displayed mode indicator. This timed, labeled sequence of fixations is then used for goal prediction.

B. Sequential Goal Prediction

The sequence is next passed into a pre-trained hidden Markov model for processing. We obtain an observation probability of each sequence by relabeling each goal candidate in turn as the true goal with a function f_g, evaluating the HMM likelihood, and marginalizing over all goal candidates assuming a uniform prior:

$$p(g|\ell_0, \cdots, \ell_n) = \frac{p_{\text{HMM}}(f_g(\ell_0), \cdots, f_g(\ell_n))}{\sum_{g' \in G} p_{\text{HMM}}(f_{g'}(\ell_0), \cdots, f_{g'}(\ell_n))}.$$

To train the model, we use the HARMONIC data set [32], which consists of natural gaze behavior of people performing a similar task with a similar robot. While this method differs from the method described in Aronson et al. [4], it produces comparable results: 57.8% accuracy (vs. 33% chance), 63.2% mean probability assigned to the correct goal at the end of the trial, and 92.0% median probability assigned to the correct goal at the end of the trial.

C. Combined Prediction

To combine the joystick and gaze predictions, we follow Jain and Argall [16] and assume that each prediction is independent conditioned on the goal. Assuming a uniform prior, we compute

$$p(g|\text{gaze}, \text{joystick}) = \frac{p(g|\text{gaze})p(g|\text{joystick})}{\sum_{g' \in G} p(g'|\text{gaze})p(g'|\text{joystick})},$$

with $p(g|\text{gaze})$ and $p(g|\text{joystick})$ given as above. Combining the probabilities ensures that the assistance command is always providing the maximum effort based on the system knowledge, so conflicting information between the signals leads to full movement to a neutral position.

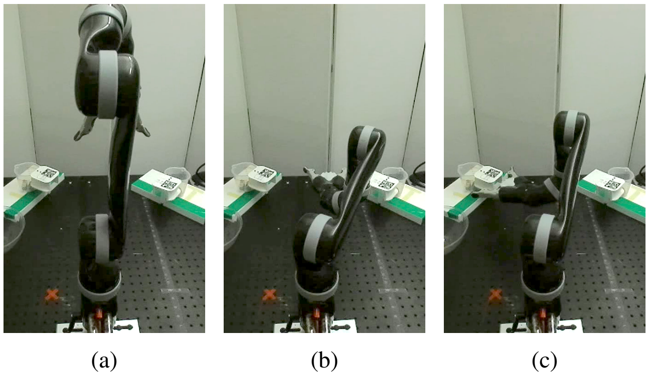
V. User Study

We hypothesize that gaze-based prediction will improve assistance when the user is unable to make progress on all parts of the task simultaneously, but the system could act in parallel with sufficient information. In this situation, task metrics will improve and goal-specific assistance will appear earlier than without the use of gaze. To evaluate this claim, we design an appropriate task and conduct a COVID-safe user study.

A. Task Development

As discussed in Sec. III, we expect that only some tasks benefit from early prediction. We design a task such that at some state typically reached, the assistance required is *different* but the user's command is *indistinguishable*. The task is a 6-dimensional, 3-mode analogue for the example in Fig. 2, in which the user can control only one axis at a time.

[2] Traditional gaze analysis distinguishes between *fixations* towards stationary objects and *smooth pursuits* towards moving objects. We only require that the object of regard remain the same, so we elide the difference.


(a)

(b)

(c)

Fig. 3: Evolution of mug grasping task. First (a), users generally reorient the robot so that the gripper is coplanar with the grasp points of the cups (b). Next, the user translates and rotates the robot to align with their specific goal (c). If the robot knows the user's goal in stage (a), it can provide goal-specific motion in x and roll.

We start from an object spearing task used in our prior assisted manipulation work [5, 17, 32] but modify it into a cup grasping task. The robot starts at a neutral position, and the user must teleoperate it with modal control to grasp one of the two cups. From prior work with this task, we observe that users generally start by moving the robot forward ($+y$) to close the distance to all goals and reorienting the end-effector to face forward (pitch) before performing goal-specific motion. Therefore, we change the initial robot position to start midway between the goals in the x axis, so initial left-right motion is different based on the goal. We add an additional, goal-specific constraint along the roll axis by orienting the cup handles differently; to grasp a cup, the user must rotate the end-effector to align with its handle, another motion that depends on the goal. The stages of the new task appear in Fig. 3. While the user is moving the robot in y and pitch, the system does not get any information about their goal from their control input; early, gaze-based goal prediction enables assistance in x and roll before the user begins providing goal-specific input.

B. User Study

We conducted a user study in which participants performed this cup grasping task. The study was performed within subjects and fully counterbalanced, with three conditions {*joystick*, *gaze*, *merged*} corresponding to which prediction strategy was used for the assistance.

Because of the COVID-19 pandemic, the user study was performed in a hybrid remote-local fashion. The robot and a stationary camera were set up in the lab. Each participant received a laptop, eye gaze sensor (Tobii Eye Tracker 4C, a screen-based tracker), joystick, webcam, and computer paraphernalia at their home. Participants assembled the equipment with remote experimenter supervision. They then connected the laptop to the lab via OpenVPN. Using ROS and a custom interface, the laptop displayed a live video feed of the robot and transmitted the user's joystick command, gaze data, and face video (which was used only for communication). In this way, participants controlled the robot without indoor contact.

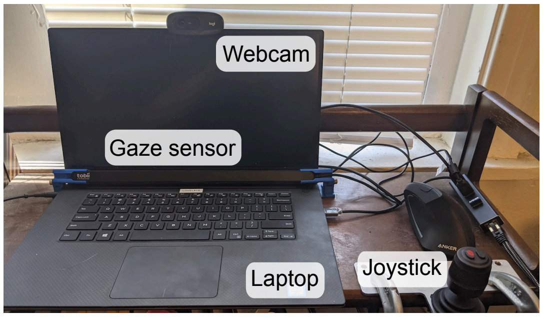


Fig. 4: Study setup that participants prepared at home.

C. Procedure

After filling out a consent form and reporting demographic information, participants received an explanation of the task while observing an autonomous grasp by the robot. Next, participants were instructed on how to control the robot and practiced for approximately five minutes. During this time, camera parameters were adjusted to compensate for latency; the resulting delay was typically $50 - 70$ ms. In addition, the fixation segmentation algorithm [39] was trained on their eye gaze data. Next, the participant performed four trials with no assistance. Finally, the participant performed four trials each of the three conditions listed above, fully counterbalanced. To accustom participants to the assistance, they performed an additional trial in their first assisted condition which was omitted from analysis. Participants filled out a questionnaire after each condition and another questionnaire at the end (see supplementary material).

D. Participants

The study was conducted with 12 participants (6 male, 6 female, 0 other). Ages of participants were 6 aged 18-24, 4 aged 25-30, and 2 aged 30-40. For familiarity with operating robots, 2 reported lots of familiarity, 6 reported some familiarity, and 4 reported no familiarity. Participants received \$20 compensation for their participation, which took approximately 1.5-2 hours including setup and teardown. The study was approved by the university IRB office. Since the study required lending materials to participants, recruitment was limited to university posting and word of mouth.

E. Evaluation Metrics

a) Algorithmic metrics: Within each trial, we compute the *prediction strength*, which is the probability assigned to the correct goal during the course of the trial.

b) Trial metrics: For each trial, we compute the *trial duration* and the *active fraction*. *Trial duration* refers how long it took the user to complete the task, and *active fraction* refers to what fraction of the trial the joystick command was non-zero; i.e., the user was explicitly providing input. Shorter trials and trials with less joystick input were considered better.

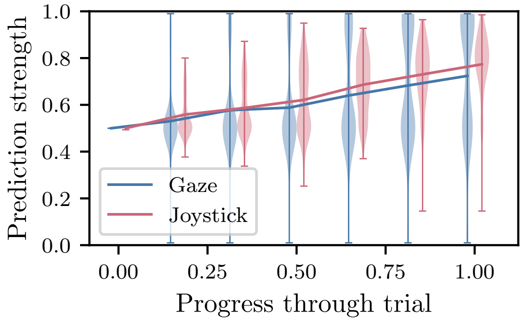


Fig. 5: Distributions of prediction strength given by gaze and joystick methods over all trials, normalized by trial duration. While the median prediction strength over time is similar between the two, the distributions are different. The joystick prediction for each trial smoothly increases over time. The gaze prediction, however, is bimodal, and the median gaze prediction strength increases as more trials transition from the $p \approx 0.5$ to $p \approx 1$ at different times. The bimodal nature of gaze means that many trials provide accurate goal predictions substantially earlier than the joystick method does, despite the two signals' similar median performance.

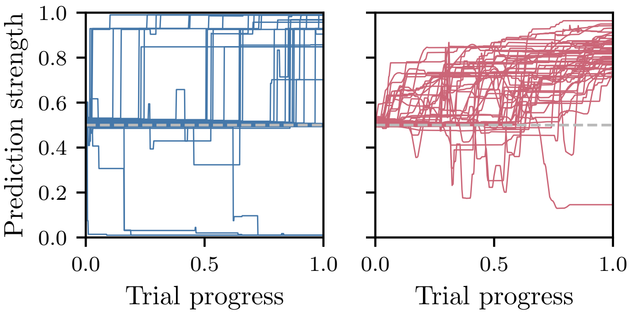


Fig. 6: Prediction strength for each condition over all trials, normalized by trial duration. The gaze predictions (left) generally transition sharply between $p \approx 0.5$ (no prediction) and $p \approx 1$ (confident, accurate prediction). The joystick predictions (right) smoothly increase over time.

c) Subjective metrics: See supplementary material.

F. Hypotheses

H1: Eye gaze is capable of predicting the user's goal earlier than joystick input can. This hypothesis follows the observation in Aronson et al. [4] that gaze can give an earlier prediction horizon, which underlies our model for task improvement. We do not require (or expect) the gaze prediction to *consistently* precede joystick prediction; rather, we only need it to do so sufficiently often to evaluate its impact on the assistance.

H2: When the assistance system receives a prediction from gaze before a distinguishable state, trial metrics will improve and goal-specific assistance will appear earlier. By considering only trials in which gaze yielded a prediction and analyzing when the prediction was received, we evaluate the model of when joystick-based assistance is improved.

VI. Results

A. Gaze Gives Early Predictions

Our model for gaze improving assistance requires that it gives earlier predictions than the joystick input does. Figure 5 shows the prediction strength of gaze and joystick over the course of each trial. While gaze and joystick prediction medians behave similarly, they follow different distributions. Gaze-based prediction is bimodal, which agrees with Aronson et al. [4]. While the joystick prediction strength steadily increases throughout each trial, the gaze prediction strength increases by shifting probability mass from $p = 0.5$ to $p \approx 1$. Fig. 6 shows traces of all runs in the gaze and joystick conditions. The gaze prediction generally starts at 0.5 and jumps to $p \approx 1$ at some point. This jump occurs at the first identified fixation on one of the goals. While the effect is not consistent, we do find that gaze is capable of providing earlier predictions than the joystick can, so *H1* is supported.

B. Early Gaze Improves Trial Performance

Next, we assess how early goal prediction from gaze affects trial performance. First, we consider only trials in which the gaze gave a prediction at all. We divide this set into those that gave an *early* prediction and those that gave a *late* prediction. Early trials predicted a goal before a threshold time T_c. Specifically, we require:

$$\forall t, t \geq T_c : |p(g|\text{data}_0, \cdots, \text{data}_t) - 0.5| \geq 0.1.$$

Since there are only two goals, either goal can be used for this calculation. These criteria mirror the ones given in Sec. III: the gaze must give a prediction when the optimal motion is different for each goal, but the user's command is still indistinguishable. To choose this threshold, we observe that the goal-independent assistance generally finishes about $T_c = 20$ seconds into the task. The remaining trials that gave a prediction were labeled late. Of the 47 trials in the merged condition, 21 (45%) were early and 9 (19%) were late. (The remaining 17 (37%) did not give a prediction.) We compare the early and late gaze prediction strength with the joystick prediction strength in Fig. 7 to confirm that this threshold generally aligns with when the joystick gives a goal prediction.

We now consider how the timing of the prediction affects trial metrics. Fig. 8 show task metrics for early and late trials compared to trials in the joystick condition. A one-way ANOVA evaluated on the log of the data shows significance for both trial duration ($F(2, 76) = 6.78, p < 0.002$) and active fraction ($F(2, 76) = 4.32, p < 0.013$). Post-hoc analysis with the Tukey HSD test shows that early gaze has shorter trials than both late gaze ($p < 0.006, 95\%$ CI $= [0.14, 0.93]$) and joystick alone ($p < 0.008, 95\%$ CI $= [0.077, 0.60]$). In addition, early gaze takes less joystick effort than does joystick

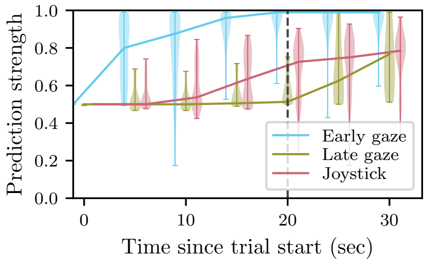


Fig. 7: Distributions of prediction strength over all trials for early gaze, late gaze, and joystick. The x-axis here is not normalized by trial time. The dashed line at 20s indicates the cutoff time T_c for early gaze prediction.

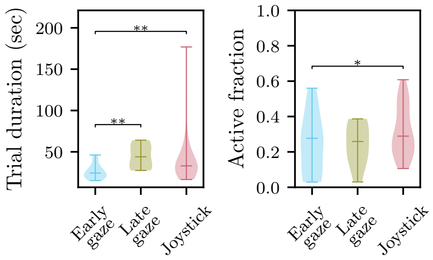


Fig. 8: Trial metrics for early gaze, late gaze, and joystick. * indicates significance at $p < 0.05$ and ** at $p < 0.01$. Early gaze trials are shorter than both other conditions and require less input than the joystick.

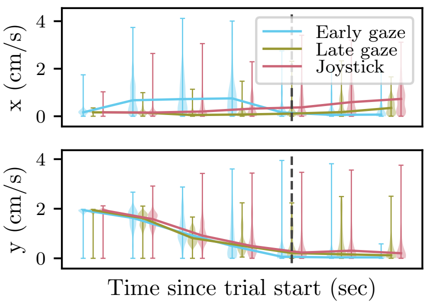


Fig. 9: Robot assistance over time in x (top) and y (bottom). Early gaze assists in x before the $T_c = 20$ sec. cutoff, while late gaze and joystick do not assist in this axis until after T_c. In y, the assistance is the same for all conditions.

alone ($p < 0.02, 95\%$ CI $= [0.077, 0.93]$). The benefit of early gaze specifically relative to both late gaze and joystick show that *H2* is supported.

We also consider the magnitude of the assistance over time, shown in Fig. 9. As described in Sec. V-A, the task is designed such that the optimal motion is different depending on the user's goal along the x axis throughout the task, but it is identical along the y axis. We see that the early gaze allows earlier assistance in x than late gaze or joystick do, since the latter conditions can only assist once the user input becomes distinguishing. In contrast, the assistance along the y axis is similar for all cases; receiving a goal prediction does not change the assistance. This observation aligns with the reasoning given in Sec. III.

VII. Study Circumstances

A. User Gaze is Natural, Not Intentional

This study proposed to evaluate *natural* gaze for goal prediction. Unlike during passive data collection, the system responded actively to participants' gaze behavior. Therefore, participants may have noticed that the system responded to their gaze and chosen to use their gaze as an explicit input. To determine if the gaze was indeed natural, participants were asked after each condition if they used any particular strategies to control the robot. In addition, in the final questionnaire, they were asked to select trials in which the robot was responsive to their gaze. Of the 12 participants, 8 reported that they did not notice gaze responsiveness in any system, 2 incorrectly labeled the joystick condition as gaze-responsive, 1 identified the merged condition but not the gaze condition, and 1 labeled the conditions correctly. Several participants expressed surprise at the question and during the subsequent debrief, saying they had forgotten about the gaze collection entirely or assumed that it was only for passive collection. Therefore, much of the gaze captured seems to be natural rather than intentional.

B. Remote Robot Control

As described in Sec. V-B above, the study was performed in a hybrid manner, in which a participant at their home controlled a robot in the lab, which led to some challenges. The primary challenge mentioned by participants was using a single, stationary camera to judge the robot's position. Participants often reported struggling with depth perception, particularly during the first, unassisted trials and when aligning the robot gripper with the goal handle. When the assistance was available, depth perception was less of a problem. Few participants reported latency problems; when they did, modifying the video streaming resolution mitigated the problem. In addition, using a stationary viewpoint made the gaze detection problem significantly easier, as it eliminated head motion, 3D gaze detection, and parallax. Ultimately, the remote study seemed to validate our system on a physical robot and using eye tracking in the loop despite the restrictions imposed by the COVID-19 pandemic.

VIII. Discussion

The results above demonstrate a particular example of when goal prediction using control input falls short. Even when the user acts optimally, the constraints of the task cause assistance using only input-based prediction to underperform. When another source provides an earlier goal prediction, the assistance can help more, earlier. This finding matches the model for the success and limitation of input-based prediction discussed in Sec. III.

In addition, we find that natural gaze can provide the early goal prediction that the input cannot. However, the gaze pipeline used here, and the gaze signal itself, does not provide the information consistently. Only 21/47 (45%) of trials using gaze alongside the joystick gave accurate predictions sufficiently early to outperform trials with only joystick-based assistance. These findings suggest that an appropriate use of gaze-based prediction is as a *signal of opportunity*. While gaze

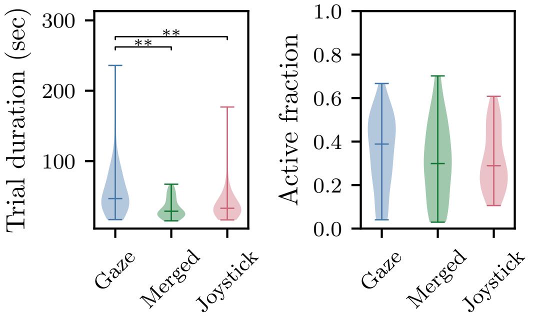


Fig. 10: Trial metrics per condition. * indicate significance at $p < 0.05$, and ** at $p < 0.01$. Gaze takes significantly longer than either condition, and there is no distinction within active fraction.

can improve task performance for certain tasks, its unreliability makes it a poor signal on its own. Though better interpretation pipelines can improve performance, the lack of any goal-directed fixations during some trials fundamentally limits its predictive ability.

Alternate strategies for merging the two prediction methods may make gaze more useful. Since we find that gaze only helps when it appears before the joystick prediction does, we can use gaze for an initial prediction, but switch to the joystick method and entirely omit gaze once distinguishing input becomes available. In addition, other tasks that are more sensitive to early prediction may show greater improvement using gaze. By analyzing the specific role and of each prediction source, we can combine multiple signals in a more nuanced way and achieve better overall performance.

A. Gaze Alone Performs Poorly

To further explore the usefulness of natural gaze for goal prediction, we measure how effective the gaze signal is for assistance on its own. We report overall trial metrics in Fig. 10 for each condition. A one-way ANOVA evaluated on the log of the data shows significance only for trial duration ($F(2, 142) = 12.7, p < 10^{-5}$). Post-hoc analysis using the Tukey HSD test on the log shows that the gaze condition alone takes longer than both the merged condition ($p = 0.001, 95\%$ CI $= [-0.71, -0.25]$) and the joystick condition ($p < 0.002, 95\%$ CI $= [-0.59, -0.12]$). In addition, people generally rated the gaze-alone condition worse than either of the others (see supplementary material).

Gaze suffers because goal-directed gaze does not occur in every trial. Familiarity with the scene from previous trials, adjusting goal-independent factors such as robot rotation, and peripheral vision all contribute to the unreliability of distinguishing gaze behavior [2, 5]. In fact, 33/95 (35%) of trials exhibited no goal-directed fixations at all. In these cases, assistance was provided for the first part of the trial (when it is identical for each goal), but subsequent motion is unassisted.

Incorrect predictions were even worse than no predictions at all. If the gaze prediction selects the incorrect goal early in the trial, it was nearly impossible for users to correct it. For example, if the user glances at one goal while trying to navigate to the other (due to, e.g., wandering attention or an error in gaze detection), the gaze-based assistance moves the robot directly to that goal. When the user attempts to maneuver the robot arm away from that goal, they look at the robot end-effector and at the incorrect goal to avoid collision, reinforcing the incorrect prediction. This self-reinforcing behavior was nearly impossible for participants to correct. Participants described this condition as “adversarial” and “like trying to hold onto a slimy eel while it attempts to wriggle away,” and even changed their goals to “accept its whimsy ways.” This behavior is analogous to the adversarial conditions in Newman et al. [31] and Stolzenwald and Mayol-Cuevas [38]. While this issue can arise when a system using control input approaches collinear goals [10], when gaze is the only prediction source, even maximum input to the other goal does not fix the problem. The simplicity of the gaze model, and the focus on object identification without an understanding of object role, illustrates the fragility of this method for goal prediction in even a simple task.

B. Adding Gaze Does Not Provide Overall Improvement

While adding gaze improves on tasks metrics when the gaze provides an early prediction, we consider the overall impact of adding gaze. The merged condition, which uses both gaze and joystick predictions, does not show improvement over using joystick alone in trial metrics (Fig. 10) or subjective metrics (see supplementary material). While 45% of merged trials contained early gaze and thus better performance, the effect may not have been sufficiently large or occur frequently enough to make an overall difference. In addition, the downsides of poor gaze may have led to frustrating behavior that counteracted the benefit gained from early gaze.

C. Extension to More Complex Tasks

The gaze-based method can be extended to include additional goals, with the caveat that gaze discrimination becomes noisier as the goals get closer together. For more complex tasks, however, gaze prediction will require more sophisticated analysis. In particular, it is difficult using gaze itself to determine the *role* that any particular object has in a task: users can look at one object since it is a goal, and another since it is an obstacle. More detailed analysis such as stronger task models [9] or analysis of gaze locations within an object [2, 21] may help for more general tasks.

In addition, this work assumes that a grasp is the only possible interaction with a goal. However, both control input [22] and natural gaze [42] can be used to infer information about the intended task of the user. We believe that task inference may follow similar patterns as goal inference, with task-specific control input restricted in time if the interface can only support particular interactions and gaze possibly providing earlier task information. Extending this work to more varied tasks is an important aim of future work.

Finally, this work assumes that the user's goal is one of a pre-specified set of objects already known to the assistance system. While this assumption is standard [18], it represents a significant gap between the experimental conditions and a full, deployed system. We look forward to expanding the goal inference process to more general settings.

IX. Conclusion

In this work, we explore the strengths and limitations of goal prediction based on control input for assisted robot teleoperation, and we explore natural gaze as a prediction method to mitigate some of those problems. We demonstrate that particular task constraints can arbitrarily limit assistance even if the user acts optimally. In a user study, we demonstrate this suboptimality in joystick-based prediction. Using natural eye gaze for the prediction as well does improve task metrics when the gaze information comes sufficiently early, which it does often. However, it does not give this information reliably, as people will often never produce goal-distinguishing gaze during a trial, and using gaze alone can lead to problematic feedback loops. Further work will focus on developing this complementarity between gaze-based prediction and joystick-based prediction, specifically by exploring nuanced ways to combine the signals for effective assistance.

Acknowledgement

This work was supported by the National Science Foundation (IIS-1943072), the Tang Family Foundation Innovation Fund, and an Okawa Foundation Research Grant. Thanks to Carolyn Youstra for assistance in fabricating the experimental equipment.

References

[1] Henny Admoni and Siddhartha Srinivasa. Predicting user intent through eye gaze for shared autonomy. In *AAAI Fall Symposium - Technical Report*, volume FS-16-01 -, pages 298–303, 2016. ISBN 9781577357759. URL http://hennyadmoni.com/documents/admoni2016aaaifs.pdf.

[2] Sai Krishna Allani, Brendan John, Javier Ruiz, Saurabh Dixit, Jackson Carter, Cindy Grimm, and Ravi Balasubramanian. Evaluating human gaze patterns during grasping tasks: Robot versus human hand. In *Proceedings of the ACM Symposium on Applied Perception, SAP 2016*, pages 45–52, 2016. ISBN 9781450343831. doi: 10.1145/2931002.2931007. URL http://dx.doi.org/10.1145/2931002.2931007.

[3] Reuben M Aronson and Henny Admoni. Semantic gaze labeling for human-robot shared manipulation. In *Eye Tracking Research and Applications Symposium (ETRA)*. Association for Computing Machinery, 2019. ISBN 9781450367097. doi: 10.1145/3314111.3319840.

[4] Reuben M Aronson, Nadia Almutlak, and Henny Admoni. Inferring Goals with Gaze during Teleoperated Manipulation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021.

[5] R.M. Aronson, T. Santini, T.C. Kübler, E. Kasneci, S. Srinivasa, and H. Admoni. Eye-Hand Behavior in Human-Robot Shared Manipulation. In *ACM/IEEE International Conference on Human-Robot Interaction*, volume Part F1350, 2018. ISBN 9781450349536. doi: 10.1145/3171221.3171287.

[6] Rowel Atienza and Alexander Zelinsky. Intuitive human-robot interaction through active 3D gaze tracking. *Springer Tracts in Advanced Robotics*, 15:172–181, 2005. ISSN 16107438. doi: 10.1007/11008941{_}19. URL http://link.springer.com/10.1007/11008941_19.

[7] Thomas Bader, Matthias Vogelgesang, and Edmund Klaus. Multimodal integration of natural gaze behavior for intention recognition during object manipulation. In *ICMI-MLMI'09 - Proceedings of the International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interfaces*, pages 199–206, 2009. ISBN 9781605587721. doi: 10.1145/1647314.1647350.

[8] Andreea Bobu, Dexter R.R. Scobee, Jaime F. Fisac, S. Shankar Sastry, and Anca D. Dragan. LESS is more: Rethinking probabilistic models of human behavior. In *ACM/IEEE International Conference on Human-Robot Interaction*, pages 429–437, New York, NY, USA, 3 2020. IEEE Computer Society. ISBN 9781450367462. doi: 10.1145/3319502.3374811. URL https://dl.acm.org/doi/10.1145/3319502.3374811.

[9] Yu Chen and D. H. Ballard. Learning to recognize human action sequences. In *Proceedings - 2nd International Conference on Development and Learning, ICDL 2002*, 2002. ISBN 0769514596. doi: 10.1109/DEVLRN.2002.1011726.

[10] Matthew Fontaine and Stefanos Nikolaidis. A Quality Diversity Approach to Automatically Generating Human-Robot Interaction Scenarios in Shared Autonomy. In *Robotics: Science and Systems*, 2021. doi: 10.15607/rss.2021.xvii.036. URL https://roboticsconference.org/program/papers/036/.

[11] Stefan Fuchs and Anna Belardinelli. Gaze-Based Intention Estimation for Shared Autonomy in Pick-and-Place Tasks. *Frontiers in Neurorobotics*, 15(647930), 4 2021. ISSN 16625218. doi: 10.3389/fnbot.2021.647930. URL https://dx.doi.org/10.3389%2Ffnbot.2021.647930.

[12] Deepak Gopinath, Siddarth Jain, and Brenna D. Argall. Human-in-the-loop optimization of shared autonomy in assistive robotics. *IEEE Robotics and Automation Letters*, 2(1):247–254, 1 2017. ISSN 23773766. doi: 10.1109/LRA.2016.2593928. URL http://ieeexplore.ieee.org/document/7518989/.

[13] Deepak E Gopinath and Brenna D Argall. Mode switch assistance to maximize human intent disambiguation. In *Robotics: Science and Systems*, volume 13, 2017. ISBN 9780992374730. doi: 10.15607/rss.2017.xiii.046. URL http://www.roboticsproceedings.org/rss13/p46.pdf.

[14] Mary Hayhoe. Vision using routines: A functional account of vision, 1 2000. ISSN 13506285. URL http://www.tandfonline.com/doi/abs/10.1080/135062800394676.

[15] Chien-Ming Huang and Bilge Mutlu. Anticipatory robot control for efficient human-robot collaboration. In *11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 83–90. IEEE, 3 2016. ISBN 978-1-4673-8370-7. doi: 10.1109/HRI.2016.7451737. URL http://ieeexplore.ieee.org/document/7451737/.

[16] Siddarth Jain and Brenna Argall. Probabilistic Human Intent Recognition for Shared Autonomy in Assistive Robotics. *ACM Transactions on Human-Robot Interaction*, 9(1):1–23, 12 2019. ISSN 2573-9522. doi: 10.1145/3359614. URL http://dl.acm.org/citation.cfm?doid=3375676.3359614.

[17] Shervin Javdani, Siddhartha S. Srinivasa, and J. Andrew Bagnell. Shared autonomy via hindsight optimization. In *Robotics: Science and Systems*, volume 11. MIT Press Journals, 2015. ISBN 9780992374716. doi: 10.15607/RSS.2015.XI.032.

[18] Shervin Javdani, Henny Admoni, Stefania Pellegrinelli, Siddhartha S. Srinivasa, and J. Andrew Bagnell. Shared autonomy via hindsight optimization for teleoperation and teaming. *The International Journal of Robotics Research*, 37(7):717–742, 6 2018. ISSN 0278-3649. doi: 10.1177/0278364918776060. URL http://journals.sagepub.com/doi/10.1177/0278364918776060.

[19] Shervin Javdani, Henny Admoni, Stefania Pellegrinelli, Siddhartha S Srinivasa, and J Andrew Bagnell. Shared autonomy via hindsight optimization for teleoperation and teaming. *The International Journal of Robotics Research*, 37(7):717–742, 2018. doi: 10.1177/0278364918776060. URL https://doi.org/10.1177/0278364918776060.

[20] Hong Jun Jeon, Smitha Milli, and Anca Dragan. Reward-rational (implicit) choice: A unifying formalism for reward learning. In *Advances in Neural Information Processing Systems*, volume 2020-Decem, 2020.

[21] Roland S Johansson, Göran Westling, Anders Bäckström, and J Randall Flanagan. Eye–Hand Coordination in Object Manipulation. *Journal of Neuroscience*, 21(17):6917–6932, 9 2001. ISSN 1529-2401. URL http://www.ncbi.nlm.nih.gov/pubmed/11517279.

[22] Hong Jun Jeon, Dylan Losey, and Dorsa Sadigh. Shared Autonomy with Learned Latent Actions. 2020. doi: 10.15607/rss.2020.xvi.011. URL https://doi.org/10.15607/rss.2020.xvi.011.

[23] Thomas C. Kubler, Dennis R. Bukenberger, Judith Ungewiss, Alexandra Worner, Colleen Rothe, Ulrich Schiefer, Wolfgang Rosenstiel, and Enkelejda Kasneci. Towards automated comparison of eye-tracking recordings in dynamic scenes. In *EUVIP 2014 - 5th European Workshop on Visual Information Processing*. Institute of Electrical and Electronics Engineers Inc., 1 2015. ISBN 9781479945726. doi: 10.1109/EUVIP.2014.7018371.

[24] Minae Kwon, Erdem Biyik, Aditi Talati, Karan Bhasin, Dylan P Losey, and Dorsa Sadigh. When humans aren't optimal: Robots that collaborate with risk-aware humans. In *ACM/IEEE International Conference on Human-Robot Interaction*, pages 43–52, 2020. ISBN 9781450367462. doi: 10.1145/3319502.3374832. URL https://doi.org/10.1145/3319502.3374832.

[25] Michael F. Land and Mary Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25):3559–3565, 2001. ISSN 00426989. doi: 10.1016/S0042-6989(01)00102-X.

[26] Yann Seing Law-Kam Cio, Maxime Raison, Cedric Leblond Menard, and Sofiane Achiche. Proof of Concept of an Assistive Robotic Arm Control Using Artificial Stereovision and Eye-Tracking. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(12):2344–2352, 12 2019. ISSN 15580210. doi: 10.1109/TNSRE.2019.2950619.

[27] Songpo Li, Xiaoli Zhang, and Jeremy D. Webb. 3-D-Gaze-Based Robotic Grasping Through Mimicking Human Visuomotor Function for People with Motion Impairments. *IEEE Transactions on Biomedical Engineering*, 64(12):2824–2835, 12 2017. ISSN 15582531. doi: 10.1109/TBME.2017.2677902. URL https://ieeexplore.ieee.org/document/7870669/.

[28] Dylan P. Losey, Craig G. McDonald, Edoardo Battaglia, and Marcia K. O'Malley. A review of intent detection, arbitration, and communication aspects of shared control for physical human–robot interaction, 1 2018. ISSN 00036900. URL http://asmedigitalcollection.asme.org/appliedmechanicsreviews/article-pdf/70/1/010804/5964415/amr_070_01_010804.pdf.

[29] Neil Mennie, Mary Hayhoe, and Brian Sullivan. Look-ahead fixations: Anticipatory eye movements in natural tasks. *Experimental Brain Research*, 179(3):427–442, 5 2007. ISSN 00144819. doi: 10.1007/s00221-006-0804-0. URL http://link.springer.com/10.1007/s00221-006-0804-0.

[30] Katharina Muelling, Arun Venkatraman, Jean Sebastien Valois, John E Downey, Jeffrey Weiss, Shervin Javdani, Martial Hebert, Andrew B Schwartz, Jennifer L Collinger, and J. Andrew Bagnell. Autonomy infused teleoperation with application to brain computer interface controlled manipulation. *Autonomous Robots*, 41(6):1401–1422, 2017. ISSN 15737527. doi: 10.1007/s10514-017-9622-4.

[31] Benjamin A. Newman, Abhijat Biswas, Sarthak Ahuja, Siddharth Girdhar, Kris K. Kitani, and Henny Admoni. Examining the Effects of Anticipatory Robot Assistance on Human Decision Making. In *International Conference on Social Robotics (ICSR)*, volume 12483 LNAI, pages 590–603. Springer, Cham., 11 2020. ISBN 9783030620554. doi: 10.1007/978-3-030-62056-1{_}49. URL https://doi.org/10.1007/978-3-030-62056-1_49.

[32] Benjamin A. Newman, Reuben M. Aronson, Siddartha S. Srinivasa, Kris Kitani, and Henny Admoni. HARMONIC: A Multimodal Dataset of Assistive Human-Robot Collaboration. *The International Journal of Robotics Research*, 41(1):3–11, 7 2022. ISSN 17413176. doi: 10.1177/02783649211050677. URL http://journals.

sagepub.com/doi/10.1177/02783649211050677.

[33] Yosef Razin and Karen Feigh. Learning to predict intent from gaze during robotic hand-eye coordination. In *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, volume 31, pages 4596–4602, 2 2017. URL www.aaai.org.

[34] Siddharth Reddy, Anca D Dragan, and Sergey Levine. Shared Autonomy via Deep Reinforcement Learning. *Robotics: Science and Systems*, 2018.

[35] Charles Schaff and Matthew Walter. Residual Policy Learning for Shared Autonomy. In *Robotics: Science and Systems*, 2020. doi: 10.15607/rss.2020.xvi.072. URL https://ttic.uchicago.edu/.

[36] Lei Shi, Cosmin Copot, and Steve Vanlanduit. Application of Visual Servoing and Eye Tracking Glass in Human Robot Interaction: A case study. In *2019 23rd International Conference on System Theory, Control and Computing (ICSTCC)*, pages 515–520. IEEE, 10 2019. ISBN 978-1-7281-0699-1. doi: 10.1109/ICSTCC.2019.8886064. URL https://ieeexplore.ieee.org/document/8886064/.

[37] Arjun Sripathy, Andreea Bobu, Daniel S. Brown, and Anca D. Dragan. Dynamically Switching Human Prediction Models for Efficient Planning. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3495–3501. Institute of Electrical and Electronics Engineers (IEEE), 10 2021. doi: 10.1109/icra48506.2021.9561430.

[38] Janis Stolzenwald and Walterio W Mayol-Cuevas. Rebellion and Obedience: The Effects of Intention Prediction in Cooperative Handheld Robots. In *IEEE International Conference on Intelligent Robots and Systems*, pages 3012–3019, 2019. ISBN 9781728140049. doi: 10.1109/IROS40897.2019.8967927. URL https://arxiv.org/abs/1903.08158.

[39] Enkelejda Tafaj, Gjergji Kasneci, Wolfgang Rosenstiel, and Martin Bogdan. Bayesian online clustering of eye movement data. *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12*, page 285, 2012. doi: 10.1145/2168556.2168617. URL http://dl.acm.org/citation.cfm?id=2168556.2168617.

[40] Katherine M. Tsui, Aman Behal, David Kontak, and Holly A. Yanco. I want that: Human-in-the-loop control of a wheelchair-mounted robotic arm. *Applied Bionics and Biomechanics*, 8(1):127–147, 2011. ISSN 17542103. doi: 10.3233/ABB-2011-0004.

[41] Ming Yao Wang, Alexandros A. Kogkas, Ara Darzi, and George P. Mylonas. Free-View, 3D Gaze-Guided, Assistive Robotic System for Activities of Daily Living. In *IEEE International Conference on Intelligent Robots and Systems*, pages 2355–2361. Institute of Electrical and Electronics Engineers Inc., 12 2018. ISBN 9781538680940. doi: 10.1109/IROS.2018.8594045.

[42] Xiaoyu Wang, Alireza Haji Fathaliyan, and Veronica J. Santos. Toward Shared Autonomy Control Schemes for Human-Robot Systems: Action Primitive Recognition Using Eye Gaze Features. *Frontiers in Neurorobotics*, 14:66, 10 2020. ISSN 16625218. doi: 10.3389/fnbot.2020.567571. URL https://www.frontiersin.org/article/10.3389/fnbot.2020.567571/full.

[43] Weilie Yi and Dana Ballard. Recognizing behavior in hand-eye coordination patterns. *International Journal of Humanoid Robotics*, 6(3):337–359, 2009. doi: 10.1142/S0219843609001863. URL http://www.ncbi.nlm.nih.gov/pubmed/20862267.

[44] Brian D Ziebart, Andrew Maas, J Andrew Bagnell, and Anind K Dey. Maximum Entropy Inverse Reinforcement Learning. In *Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 3*, AAAI'08, page 1433–1438. AAAI Press, 2008. ISBN 9781577353683.

[45] Brian D Ziebart, Andrew Maas, J Andrew Bagnell, and Anind K Dey. Human behavior modeling with maximum entropy inverse optimal control. In *AAAI Spring Symposium - Technical Report*, volume SS-09-04, pages 92–97, 2009. ISBN 9781577354116. URL http://www.cs.cmu.edu/~bziebart/publications/human-behavior-bziebart.pdf.

[46] 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 *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3931–3936. IEEE, 10 2009. ISBN 978-1-4244-3803-7. doi: 10.1109/IROS.2009.5354147. URL http://ieeexplore.ieee.org/document/5354147/.

[47] Matthew Zurek, Andreea Bobu, Daniel S Brown, and Anca D Dragan. Situational Confidence Assistance for Lifelong Shared Autonomy. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2783–2789, 2021. doi: 10.1109/icra48506.2021.9561839. URL http://arxiv.org/abs/2104.06556.