Enhancing Goal Inference via Correction Timing

Anjiabei Wang
Yale University
New Haven, United States
anjiabei.wang@yale.edu

Shuangge Wang
Yale University
New Haven, United States
shuangge.wang@yale.edu

Tesca Fitzgerald
Yale University
New Haven, United States
tesca.fitzgerald@yale.edu

Abstract

Corrections offer a natural modality for people to provide feedback to a robot, by (i) intervening in the robot's behavior when they believe the robot is failing (or will fail) the task objectives and (ii) modifying the robot's behavior to successfully fulfill the task. Each correction offers information on what the robot should and should not do, where the corrected behavior is more aligned with task objectives than the original behavior. Most prior work on learning from corrections involves interpreting a correction as a new demonstration (consisting of the modified robot behavior), or a preference (for the modified trajectory compared to the robot's original behavior). However, this overlooks one essential element of the correction feedback, which is the human's decision to intervene in the robot's behavior in the first place. This decision can be influenced by multiple factors including the robot's task progress, alignment with human expectations, dynamics, motion legibility, and optimality. In this work, we investigate whether the timing of this decision can offer a useful signal for inferring these task-relevant influences. In particular, we investigate three potential applications for this learning signal: (1) identifying features of a robot's motion that may prompt people to correct it, (2) quickly inferring the final goal of a human's correction based on the timing and initial direction of their correction motion, and (3) learning more precise constraints for task objectives. Our results indicate that correction timing results in improved learning for the first two of these applications. Overall, our work provides new insights on the value of correction timing as a signal for robot learning.

Keywords

Interactive Robot Learning, Learning from Corrections

ACM Reference Format:
Anjiabei Wang, Shuangge Wang, and Tesca Fitzgerald. 2026. Enhancing Goal Inference via Correction Timing. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 13 pages.

1 Introduction

When a robot is performing a task and appears to be failing (or about to fail), a human can intervene and physically correct the robot's motion to achieve the intended task objective. Compared to traditional paradigms with designated teaching and deployment cycles, these intervention-based interactions offer a practical source of training data: 1) they are more data-efficient [16], as the robot is continuously deployed and corrected only when necessary; and 2) they enable more intuitive interactions [3, 34], since they draw on human domain expertise without requiring technical interfaces or specialized instruction for the humans. In practice, people provide corrections due to perceived violations of task constraints, such as task progress [2, 32], safety constraints [1, 41], or user preferences [49, 50], making them potentially highly informative toward learning these task constraints.

Prior work on learning from corrections [3, 20, 26, 27, 40] primarily formulates it as an Inverse Reinforcement Learning (IRL) problem, where the robot infers the reward function it should optimize during the task by treating corrections as evidence about the reward function's parameters. Within the IRL framework, prior work has focused on the spatial aspects of corrections — such as where and how the robot's motion is adjusted — while overlooking another critical dimension: the human's decision to intervene in the robot's behavior in the first place. Prior studies show that motion features such as efficiency, safety, legibility, and human-likeness shape how people interpret and respond to robots [1, 4, 12, 19, 35, 41]. Based on the interpretations on robot motions, humans may decide that the robot's current trajectory is inadequate and intervene to ensure task success. Thus, the decision to intervene reflects the moment when humans internally judge that the robot requires assistance — and provides insight into *why* the correction occurs [30, 46].

In this work, we focus on physical correction to a robot. We **hypothesize that the timing of human's intervention decision offers insight into the underlying task objectives**. We investigate this hypothesis in the context of three learning-related applications (and corresponding research questions):

RQ1: What features of the robot's motion prompt people to correct it? We aim to understand when and why corrections occur, enabling the future design of robot trajectories that either elicit informative corrections or avoid unnecessary ones.

RQ2: Can we utilize the information available at the onset of the correction to directly infer the task goal? This capability could enable future robots to respond to human corrections in real time.

RQ3: Can we learn more precise constraints about the task goal by using timing information? This would allow a robot to learn more precise task constraint information that may not be captured from spatial cues alone.

Our results indicate that timing information contributes meaningfully to the first two applications. Our main contributions include: **1)** We evaluate the contribution of individual trajectory-based features to correction timing prediction through a feature ablation study, providing insights into which aspects of robot motion influence humans' decisions to intervene. **2)** We demonstrate that incorporating timing information enhances early inference of human-intended goals from correction onset cues, but provides limited gains for refining fine-grained task constraint learning.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus.

2 Related Work

2.1 Learning from Corrections

Correction feedback occurs when a human intervenes during a robot's task execution to modify its ongoing motion and help it succeed. Prior work has leveraged such corrections as demonstrations. For instance, Bajcsy et al. [4], Losey et al. [34] interpret human corrections as intentional actions that reveal information about the task objective parameters, allowing the robot to update its model accordingly. Bobu et al. [6] argue that human corrections may fall outside the robot's hypothesis space and thus be misleading if incorporated directly; to address this, they introduce situational confidence, weighting each correction by how well the robot's current hypothesis can explain the human input. Zhang and Dragan [54] treat local corrections as partial demonstrations and extrapolate them to infer the human's intended full trajectory for objective learning.

Beyond demonstrations, other studies have modeled corrections as preferences, assuming that the corrected trajectory is more aligned with the human's desired objective than the original. For example, Jain et al. [20, 21], Wilde et al. [51] consider corrections as iterative, incremental trajectory refinements that gradually shift the robot's behavior toward the desired trajectory through comparison-based updates. Similarly, Mehta and Losey [36] treat corrections as evidence that the corrected trajectory segment is locally more desirable than nearby perturbations.

However, these works primarily focus on learning from the correction once it has occurred, rather than why or when the human decided to intervene. Korkmaz and Bıyık [30] introduced probabilistic models to predict both when an intervention occurs and how the human corrects, integrating these signals into policy learning and achieving high policy update performance. While their approach leverages intervention signals to improve robot behavior, it does not evaluate whether the model itself accurately reflects the underlying human decision-making process. In particular, their framework assumes Boltzmann-rational human actions [33], where human corrections are modeled probabilistically as being exponentially more likely for actions that yield higher expected rewards.

While the work above captures behavioral variability, it frames corrections as outcomes of optimization rather than as expressions of human intent shaped by task goals and interaction context. In practice, however, corrections can be indicative of humans' intent to adjust the robot's behavior toward a desired outcome: Jin et al. [25] infer instantaneous intent from directional feedback, while Schrum et al. [42] learn individualized mappings that translate human correction styles into intended objectives. In parallel, Bayesian frameworks for intent recognition in collaborative settings [18, 22] demonstrate how probabilistic inference can recover latent human goals from behavioral signals. Together, these insights motivate our approach: if corrections reflect a human's internal evaluation of the robot's performance, then their timing and spatial characteristics should provide implicit cues for goal inference.

2.2 Human Feedback Timing

In social psychology, Nosek [38] demonstrates that the timing of a person's response reveals underlying mental representations and cognitive processes, providing a window into how people internally evaluate and interpret stimuli. Similarly, research on conversational feedback shows that the timing of verbal backchannels reflects a listener's internal predictive model of dialogue, indicating moments of understanding or alignment with the speaker's intent [39]. These findings suggest that the timing of human feedback is not arbitrary but a behavioral manifestation of internal evaluation processes. Extending this idea to human–robot interaction, the timing of human feedback can convey implicit judgments about the robot's performance. For example, Wang et al. [48] examined how human correction feedback is affected by robot behavioral features — such as motion legibility and competency. They found that people tend to intervene earlier and sometimes provide unnecessary corrections when the robot appears more competent. Spencer et al. [44] show that the moment of human intervention signals when the robot's behavior becomes undesirable, effectively partitioning the action space into acceptable and unacceptable regions. Thus, feedback timing serves as a threshold on human tolerance for error and offers an implicit learning signal for the robot.

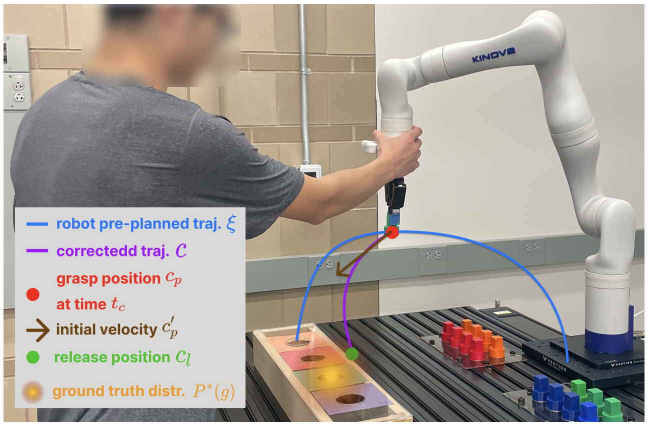


Figure 1: Definitions of correction-related terms when a human intervenes during a robot's pre-planned trajectory ξ, resulting in a corrected trajectory c. The timing information is represented by the correction time t_c, while the spatial information includes the grasp position c_p, the initial correction velocity c_p', and the release position c_l. The ground truth goal distribution is $\mathbb{P}^*(g)$.

While prior work demonstrates that intervention timing carries valuable information for learning, it does not examine which factors of robot motion drive humans' intervention decisions or how the timing of those decisions can be leveraged to infer task constraints. Our work aims to fill this gap by connecting humans' decisions to intervene with early goal inference and precise task constraint learning.

3 Problem Definition

Our goal is to model how the timing of a human's interventions can serve as a learning signal for the robot (toward inferring the task goal). We consider a scenario where the robot executes a pre-planned trajectory $\xi = \{x_t\}_{t=1}^{T}$, where x_t is the robot gripper position at each time step t. During this execution, the human may choose to intervene in the robot's motion at any time $t_c \in [1, T]$ to provide a correction. In doing so, they guide the robot through a trajectory $c = \{x_t\}_{t=t_c}^{T_c}$ (T_c being the time when the correction ends), from which we extract timing and spatial information. The timing information consists of t_c (the timestep at which the correction begins). The spatial information consists of $c_p = x_{t_c} \in \mathbb{R}^3$ (position where people grasp the robot gripper), $c'_p = \Delta_t c_p \in \mathbb{R}^3$ (the initial correction direction: velocity), and $c_l = x_{T_c} \in \mathbb{R}^3$ (where people release the gripper at the end of the correction). After the correction, the robot re-plans its trajectory based on its understanding of the task.

To address **RQ1**, we aim to identify which aspects of robot motion influences the human's decision to intervene. If certain motion features maximize a model's accuracy to predict correction timing, it suggests that those features are closely tied to the human intervention decision. Our goal is to learn the conditional distribution $\mathbb{P}(t_c \mid \xi, g)$ that models when corrections occur based on the robot original trajectory ξ and task goal $g \in \mathbb{R}^3$ (a gripper position), from which we can quantitatively evaluate which trajectory-based and task-related features explain what triggers intervention decision.

To address whether correction timing information improves direct task goal inference when using correction onset information (**RQ2**) and enhances precise task goal inference (**RQ3**), we consider the task goal g drawn from a task-dependent distribution $\mathbb{P}^*(g)$, which represents the likelihood of goal locations that result in successful task completion. This distribution is unknown to the robot and constitutes what it aims to learn. Our objective is to evaluate whether incorporating correction timing t_c can improve the inferred goal distribution $\mathbb{P}(g \mid \xi, t_c, s_c)$ — estimated from the robot's trajectory ξ, timing information t_c, and spatial correction cues s_c — to better approximate the true task constraint $\mathbb{P}^*(g)$, compared to inference using only spatial information $\mathbb{P}(g \mid \xi, s_c)$. When $s_c = (c_p, c'_p)$, representing the grasp position and initial correction direction, the resulting goal inference evaluates how well the model can infer the goal distribution directly from the onset of correction, addressing **RQ2**. When $s_c = c_l$, the release position, the model infers the precise goal distribution based on the end of the correction, corresponding to **RQ3**.

4 Approach

We propose a two-stage approach to model and utilize human correction timing for improved robot learning. In the first stage, we train a transformer-based timing prediction model to (i) take task- and motion-related trajectory features as input and (ii) output the probability of the human correcting the robot at each time step. This enables us to analyze which aspects of robot motion prompt human interventions. In the second stage, we train a goal inference model to integrate both the spatial (s_c) and temporal (t_c) characteristics of the correction to infer the goal distribution $\mathbb{P}(g \mid \xi, t_c, s_c)$. By comparing it against a spatial-only baseline $\mathbb{P}(g \mid \xi, s_c)$, we evaluate whether correction timing offers a useful signal for robot learning; i.e., whether it improves inference of human-intended correction goals and task constraints.

4.1 Predicting WHEN People Give Corrections

4.1.1 Feature Extraction

Prior studies have incorporated both velocity-related and task-related motion features into learning frameworks [4, 19], which are closely tied to what humans perceive and respond to during collaboration. Characteristics such as human-likeness [35], efficiency [19], and safety [1, 41] influence how easily humans can interpret and adapt to robot motion. In addition, legibility is known to play a key role in conveying a robot's intent [12]. The degree to which a robot's behavior aligns with what humans consider optimal strongly affects their likelihood of intervening [30, 46]. Taken together, these works suggest that the human decision to intervene are influenced by robot motions.

Table 1: Time-series features used for modeling robot behavior and human corrections. The design of these features is grounded in prior literature: each cited work motivates the inclusion of a specific aspect of robot motion shown to influence human perception and intervention behavior.

Feature	Equation
Expectation Alignment (Vel.) [13, 14]	$\frac{v_t \cdot v_t^{\text{opt}}}{\|v_t\|\|v_t^{\text{opt}}\|}$
Expectation Alignment (Pos.) [30, 46]	$\|x_t - x_t^{\text{opt}}\|$
Directness Alignment (Vel.) [13, 14]	$\frac{v_t \cdot v_t^g}{\|v_t\|\|v_t^g\|}$
Velocity Consistency [52, 53]	$\frac{v_t \cdot v_{t-1}}{\|v_t\|\|v_{t-1}\|}$
Legibility [12]	$\frac{\int \mathbb{P}(g \mid \xi_{S\to x_t})\gamma_t \, dt}{\int \gamma_t \, dt}$
Task Progress (Dist. to Goal) [13, 14]	$\|x_t - x_g\|$
Optimality [4, 6, 7, 34]	$\frac{\exp\left(-\mathcal{L}(\xi_{s\to x_t}) - \mathcal{L}^{\text{opt}}(\xi_{x_t\to g})\right)}{\exp\left(-\mathcal{L}(\xi_{s\to x_{t-1}}) - \mathcal{L}^{\text{opt}}(\xi_{x_{t-1}\to g})\right)}$

Human Expectation-related Features. Since humans have a preference for natural robot trajectories [37], we compute an optimal reference trajectory from each time step to the goal using a PID controller. This trajectory represents a smooth, dynamically feasible motion that a human might perceive as a natural movement toward the goal. From this trajectory, we obtain the *optimal next velocity* v_t^{opt} and *optimal next position* x_t^{opt} (Additional details about the PID controller and the construction of these reference trajectories are provided in Appendix A.1.1.). We then define two alignment-based features: (F1) the cosine similarity between the robot's current velocity v_t and the optimal velocity v_t^{opt} and (F2) the Euclidean distance between the current position x_t and the optimal next position x_t^{opt}.

In addition to this locally optimal reference, we also define a direct-to-goal expectation, where the robot is expected to move straight toward the goal without curvature. This yields: (F3) the cosine similarity between the current velocity v_t and a direct velocity vector v_t^g pointing from x_t to the goal g.

Dynamics-related Features. Robot dynamics and velocity patterns impact how human adapt to and interact with the robot [35, 52, 53]. For dynamics-related features, we included: (F4) motion consistency, measured by the cosine similarity between v_t and the previous velocity v_{t-1}.

Task-Performance-Related Features. To quantify how interpretable the robot's motion is to a human observer, we included (F5) a legibility score based on Dragan et al. [12], shown in Table 1 (see Appendix A.1.2 for details).

Additionally, we capture (F6) task progress as the Euclidean distance from the robot's current position x_t to the goal g, following prior work [13, 14].

And finally, (F7) Boltzmann rational optimality. Prior work on corrections [4, 6, 7, 34] models the likelihood of human intervention as a function of trajectory optimality. A common baseline is the Boltzmann rationality model [5, 55], which assumes that humans are exponentially more likely to provide a correction as the robot's behavior becomes increasingly sub-optimal. In our task, we focus on efficiency [12] and define optimality as the inverse of the total path length — specifically, the sum of the trajectory length up to the current time step and the length of the optimal path from that point to the goal. We encode optimality as a single time-varying feature: the ratio between the current and previous time step's optimality values (shown in Table 1).

4.1.2 Timing Prediction Model

Inputs. To analyze how features evolve over time and how they relate to human intervention, we convert each *original* robot trajectory into a temporal sequence. Each robot pre-planned trajectory $\xi = \{x_t\}_{t=1}^{T}$ is discretized into T time steps and converted into a featurized representation $\Phi(\xi, g) = \left[\phi_t^k(\xi, g)\right]_{t=1:T,\,k=1:7} \in \mathbb{R}^{T\times 7}$ with respect to some goal g.

Model Architecture. Given the temporal nature of the data, we adopted a transformer model [47] to preserve its sequential structure, which is essential for modeling correction decisions that are inherently non-Markovian [15]. For a given trajectory ξ, the model takes $\Phi(\xi, g)$ as input, applies masking to handle variable-length trajectories, and uses positional encoding to preserve temporal order. The encoder consists of two transformer layers, each comprising a multi-head self-attention mechanism (8 heads, embedding dimension 32) followed by a feed-forward network (64 hidden units) with residual connections, dropout, and layer normalization.

Outputs & Loss Function. The final output layer applies a sigmoid activation to produce a corresponding sequence of cumulative distribution function (CDF) probabilities $\mathbb{P}_{\text{CDF}}(t) = \mathbb{P}(t_c \le t \mid \xi, g)$ representing the likelihood of a correction occurring at or before that time step given the robot trajectory ξ and goal of the task g. These predictions are compared against ground truth labels $\hat{l}_t^i$, where $\hat{l}_t^i = 0$ if no correction has occurred or the time step precedes the correction, and $\hat{l}_t^i = 1$ for all time steps following the correction. The model is trained using a binary cross-entropy loss with an exponentially decayed learning rate schedule, and validation loss is monitored to select the best-performing checkpoint.

Given the transformer-predicted CDFs for each trajectory, we can derive the probability density function (PDF) of a correction occurring at each time step t as:

$$\begin{aligned}\mathbb{P}(t \mid \xi, g) &= \mathbb{P}_{\text{CDF}}(t) - \mathbb{P}_{\text{CDF}}(t-1) \\ &= \mathbb{P}(t_c \le t \mid \xi, g) - \mathbb{P}(t_c \le t-1 \mid \xi, g).\end{aligned} \tag{1}$$

The value of the PDF at the actual observed correction time t_c is taken at $t = t_c$, averaged within a 1.2-second window, and any negative probabilities resulting from numerical artifacts are set to zero.

4.2 Enhancing Goal Inference

In this section, we introduce three models for inferring the goal distribution $\mathbb{P}(g \mid \cdot)$ that represents the robot's estimate of where the goal is. First, we define a **WHERE model** that relies solely on the spatial information of the correction to infer the goal distribution, serving as a baseline. Next, we present a **WHEN model** that relies on the timing of human intervention to perform goal inference, isolating the contribution of temporal information. Finally, we present a **COMBINED model** that integrates both spatial and timing information to infer the goal distribution. By comparing these models, we can evaluate whether the timing of human intervention provides additional informative signals beyond spatial cues, thus testing our hypotheses in **RQ2** and **RQ3**.

4.2.1 Inferring Goal using Timing Information

Using the timing prediction model before, the posterior distribution over goals given the correction timing (defined as **WHEN model** goal distribution) is:

$$\mathbb{P}(g \mid t_c, \xi) = \frac{\mathbb{P}(t_c \mid g, \xi)\,\mathbb{P}(g \mid \xi)}{\sum_{\hat{g}\in\mathcal{G}} \mathbb{P}(t_c \mid \hat{g}, \xi)\,\mathbb{P}(\hat{g} \mid \xi)} \propto \mathbb{P}(t_c \mid g, \xi), \tag{2}$$

Assuming our prior over candidate goals $\mathbb{P}(g \mid \xi) = \mathbb{P}(g)$ is uniform (i.e., the robot's pre-planned trajectories are generated independently of the sampled goal hypotheses).

4.2.2 RQ2: Goal Inference from *Start* of Correction

To address **RQ2**, we aim to infer the intended goal location using the spatial information available at the onset of the human correction; specifically, the position where the participant first grasps the robot gripper c_p and the velocity they apply at that moment c_p'. Because people do not always release the gripper precisely at the goal after completing the correction, we decompose the process into two stages. By first predicting where participants are likely to release the gripper and then inferring the goal from these predicted endpoints, the model captures the intermediate intent expressed through the correction motion, resulting in a more interpretable and realistic goal inference process.

In the first stage, we infer where people intend to move the robot. We implement a feedforward Multilayer Perceptron (MLP) to model $c_l = \text{MLP}(c_p, c_p')$; i.e., using the initial interaction (grasp position c_p and velocity c_p') to predict where the human releases the gripper c_l. The network consists of three fully connected layers with hidden dimension 64 and ReLU activations.

In the second stage, we use the predicted release position to infer the goal distribution. We fit a Gaussian Mixture Model (GMM) to model the distribution of release positions relative to the ground-truth goal position. This results in a model for $\mathbb{P}_{\text{GMM}}(c_l \mid g)$.

Combining both the MLP and GMM, we can approximate the **WHERE model** goal distribution as follows:

$$\mathbb{P}(g \mid c_p, c_p') \approx \mathbb{P}_{\text{GMM}}(\text{MLP}(c_p, c_p') \mid g), \tag{3}$$

assuming a uniform distribution over goals. (Detailed proof in Appendix A.2.1.)

4.2.3 Combining WHEN and WHERE

In order to infer the goal by leveraging both spatial and temporal information, we compute the posterior distribution over candidate goals conditioned on the robot trajectory ξ, the observed correction timing t_c, and the spatial information available at the onset of the correction, including the grasp position c_p and the initial correction velocity c'_p (the **COMBINED model**):

$$\mathbb{P}(g \mid t_c, c_p, c'_p, \xi) = \frac{\mathbb{P}(t_c \mid g, \xi)\, \mathbb{P}_{\text{GMM}}(\text{MLP}(c_p, c'_p) \mid g)}{\sum_{\hat{g} \in \mathcal{G}} \mathbb{P}(t_c \mid \hat{g}, \xi)\, \mathbb{P}_{\text{GMM}}(\text{MLP}(c_p, c'_p) \mid \hat{g})} \tag{4}$$

The full derivation of the COMBINED model is in Appendix A.2.2. Following Eq. 22, we also experiment with weighing $\mathbb{P}_w(t_c \mid g, \xi)$ and $\mathbb{P}(c \mid g)$ differently, so the **weighted COMBINED model** becomes:

$$\mathbb{P}_w(g \mid t_c, c_p, c'_p, \xi) = \frac{\mathbb{P}(t_c \mid g, \xi)^{\alpha} \cdot \mathbb{P}_{\text{GMM}}(\text{MLP}(c_p, c'_p) \mid g)^{1-\alpha}}{\sum_{\hat{g} \in \mathcal{G}} \mathbb{P}(t_c \mid \hat{g}, \xi)^{\alpha} \cdot \mathbb{P}_{\text{GMM}}(\text{MLP}(c_p, c'_p) \mid \hat{g})^{1-\alpha}} \tag{5}$$

where $\alpha \in [0, 1]$ is a geometric weight.

4.2.4 RQ3: Goal Inference from *End* of Correction

Additionally, we investigate whether timing information can enhance the precision of goal distribution inference. Instead of using the grasp position c_p and its corresponding velocity c'_p, we directly utilize previously fitted GMM $\mathbb{P}_{\text{GMM}}(c_l \mid g)$. The resulting posterior for the **WHERE model** is $\mathbb{P}(g \mid c_l) \propto \mathbb{P}_{\text{GMM}}(c_l \mid g)(c_l \mid g)$. The **weighted COMBINED model** becomes:

$$\mathbb{P}_w(g \mid t_c, c_l, \xi) = \frac{\mathbb{P}(t_c \mid g, \xi)^{\alpha} \cdot \mathbb{P}_{\text{GMM}}(c_l \mid g)^{1-\alpha}}{\sum_{\hat{g} \in \mathcal{G}} \mathbb{P}(t_c \mid \hat{g}, \xi)^{\alpha} \cdot \mathbb{P}_{\text{GMM}}(c_l \mid \hat{g})^{1-\alpha}}, \tag{6}$$

5 Evaluation

In prior work[1], we conducted a user study (approved by our Institutional Review Board) to collect correction data with $N = 120$ participants recruited from a university community, resulting in a total of 7,435 interaction episodes and 3,585 correction trajectories. We now use this data to train and test our models in an offline manner, enabling us to evaluate our research questions.

5.1 Data Collection

Each participant in our study was tasked with supervising a 7-DoF Kinova Gen3 robotic arm equipped with a Robotiq 2F-85 gripper, mounted on a horizontal linear actuator. During each participant's 1-hour session, the robot performed a series of 64 pick-and-place operations, where the goal was to insert various shapes into matching color-coded target holes (goals). The robot approached each sub-task by executing joint motion pre-planned with RRT* [45], which were smoothed before being executed through a velocity-based PID controller. Participants were free to intervene at any time during the robot's motion. Upon applying physical force to guide the robot, an admittance control scheme [24, 28] allowed the robot to respond compliantly. We used the recursive Newton-Euler algorithm for inverse dynamics and gravity compensation [8]. Corrections terminated when participants ceased applying force; the robot then replanned from the corrected state to the closest viable goal while preserving the end-effector's final orientation. Sensor readings from joint encoders, torque sensors, and control inputs were collected at 10Hz. Cartesian states are computed through forward dynamics [10].

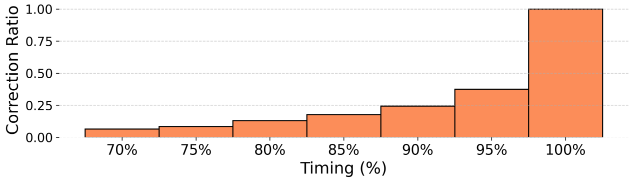


Figure 2: Corrections occurring at different percentage stages of the trajectory. The percentage denotes the portion of the robot's pre-planned trajectory that was executed before the first human correction occurred.

For the block placement task, we consider four distinct shapes — circle, square, triangle, and rectangle — each paired with four colored holes positioned at different locations on the board. Each unique shape–color pairing defines a distinct task (goal) constraint that the model is trained to learn.

Although participants could issue multiple interventions, we restrict our analysis to the *first correction* event in each trial to keep the problem tractable. We find that these first corrections occur throughout different phases of the robot's trajectory (Fig. 2), meaning that our study covers early and late corrections.

5.2 Model Training

For the timing prediction model, all trajectory inputs are derived from the robot's pre-planned trajectories. For non-corrected trials, we use the uninterrupted executed trajectory. For corrected trials, we combine the pre-correction executed segment with the intended (but unexecuted) remainder of the pre-planned trajectory.

We first create sets corresponding to correction timing percentages (≤70%, ≤80%, ≤90% and ≤100%), measured as the proportion of the trajectory executed prior to the first correction. For each percentage set, we split the trajectories involving human corrections into a training set (60%), a validation set (10%), and a test set (30%). For correction timing prediction, we trained the transformer model using features extracted from the training and validation sets, and included an equal number of uncorrected trajectories during training to enable the model to predict both *if* and *when* corrections occur. For goal distribution inference, we trained the MLP and fitted the GMM using the same training and validation sets with only the trajectories involving corrections.

5.3 Evaluation Metrics

For **RQ1**, we evaluate the accuracy of correction timing prediction and analyze the contribution of each trajectory-based feature through ablation studies. For **RQ2** and **RQ3**, we assess the WHEN model's ability to enhance the precision of goal inference.

5.3.1 Transformer Accuracy

We evaluate the model's correction timing prediction accuracy to assess whether the trajectory-based features effectively capture the motion factors that drive human intervention. Strong prediction performance indicates that these features are informative and

[1] Citation redacted for anonymity.

relevant for answering **RQ1** — understanding what aspects of robot motion lead to human decisions to intervene. We compare performance of our multi-feature model against a single-feature Boltzmann baseline, where the only input to the transformer is the Boltzmann optimality feature $\Phi_t = [\phi_t^7]$, representing the commonly used model for correction behavior in prior work. All evaluations are conducted over 200 random training, validation, and test splits of the dataset to ensure that the observed performance is robust and statistically reliable. All sets include equal numbers of correction-inducing and uncorrected trajectories.

(1) **F1 Score**: We report the mean F1 score across test set to measure how well the model is able to predict that a correction has happened and has not on a timestep level. We set a threshold of 0.5 to convert the predicted probabilities p_t into binary labels: $l_t = H(p_t - 0.5)$ and compare to the ground truth labels, where $H(\cdot)$ is the unit step function.

(2) **Correction MAE**: To evaluate the accuracy of correction timing prediction, we define the predicted correction time as the first time step at which the predicted label l_t transitions from 0 to 1 and remains 1 thereafter. For each trajectory, we compute the absolute time difference between the predicted correction time t_c and the ground truth correction time $\hat{t_c}$. The overall mean absolute error (MAE) is then reported as MAE $= \frac{1}{N}\sum_{i=1}^{N}\left|\hat{t_c}^{(i)} - t_c^{(i)}\right|$, averaged across all test trajectories.

(3) **Predicted Correction Ratio**: We compare the number of trajectories predicted to involve corrections with the number of trajectories that actually involve corrections in the test set, assessing how accurately the model identifies trajectories that involve human intervention.

5.3.2 Feature Importance

We further analyze the contribution of each feature to address **RQ1**. We conduct a feature ablation study, excluding one feature at a time during training and evaluating the resulting drop in F1 score. This analysis reveals which aspects of the robot's motion most strongly influence human correction timing decisions at different stages of the trajectory. The evaluation is conducted over 200 random training, validation, and test splits of the dataset.

5.3.3 Goal Inference Accuracy

We investigate whether timing information improves goal inference using information from the start of corrections (**RQ2**) and end of corrections (**RQ3**) For all models, we uniformly sample candidate goal positions $\hat{g}$ on the (x, y) plane at $z = 0$, assuming that the z-coordinate has minimal influence on participants' perception of the goal. The sampled region spans 40 cm × 60 cm around the board area where the actual goals are located, with a sampling resolution of $\Delta x = \Delta y = 1$ cm.

We compare the mean Kullback–Leibler Divergence (KLD) [31] between the inferred and ground truth goal distributions for the **WHEN**, **WHERE**, and weighted **COMBINED** models across all trajectories in the test set to evaluate their goal inference accuracy. For the weighted COMBINED model, we set $\alpha = 0.8$ to balance the contributions of timing and spatial information (details about how we picked α in Appendix A.5). The evaluation is conducted over 50 random training, validation, and test splits of the dataset.

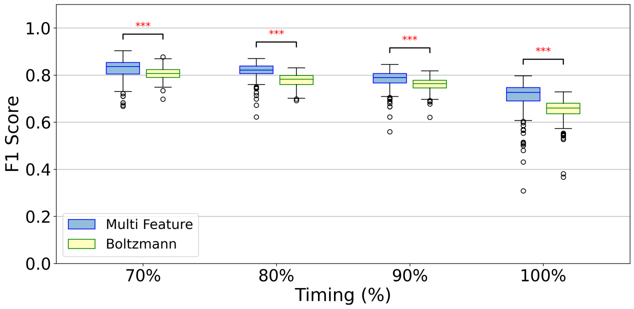


Figure 3: F1 score comparison between the multi-feature model and the Boltzmann baseline at different correction timing percentages. Each percentage represents the portion of the trajectory completed before the first correction occurred.

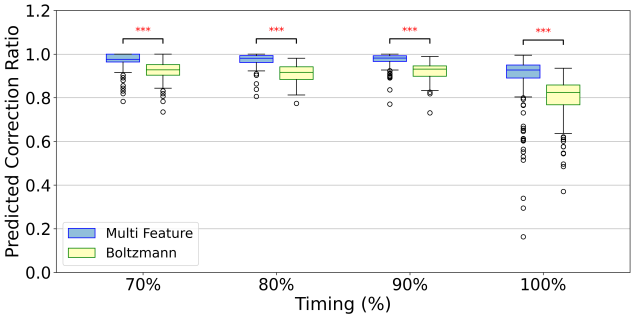


Figure 4: Ratio of predicted to actual corrections for the multi-feature model and the Boltzmann baseline across different correction timing percentages.

Ground Truth. To estimate the ground truth goal distribution for each shape, we employ a sim-to-real approach [11]. We first conduct block-dropping simulations [9] across a range of potential target positions, recording whether each drop successfully landed in the actual target region (1 for success, 0 for failure). Using eigenentropy [17], we then select a set of 100 poses that were maximally informative. These sampled poses were executed in the real world, and the corresponding outcomes were recorded. By comparing the simulation and real-world results for these poses, we identified the most accurate simulator for each shape.

With the selected simulators, we performed 10^5 additional block drops at randomly sampled poses. We recorded the poses that resulted in successful placements and aggregated these outcomes to construct the ground-truth goal distribution. This distribution was modeled using a GMM centered at the true target positions and evaluated at $z = 0.08$, since small variations in the z-coordinate do not affect whether the block successfully lands in the target. For each colored target, we applied an offset to shift the GMM center to align with the actual absolute target position corresponding to that specific color.

6 Results

(1) **Correction Timing Accuracy**: Across 200 runs, the multi-feature model consistently outperformed the Boltzmann baseline in F1 score, particularly when later corrections were included (Fig. 3).

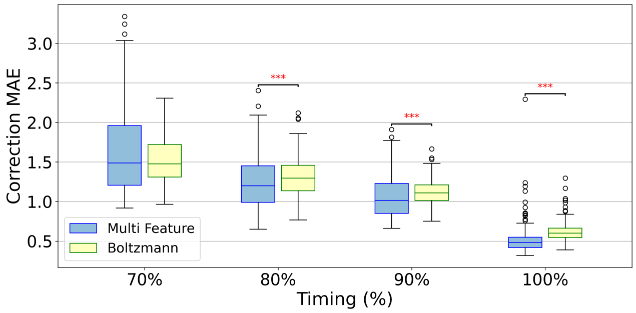


Figure 5: Mean difference between predicted and actual correction timing (in seconds) for the multi-feature model and the Boltzmann baseline across different correction timing percentages. Lower MAE means better accuracy.

It also achieved better correction timing accuracy at 80%, 90% and 100% of the trajectory (Fig. 5), and was significantly more accurate in predicting the number of actual corrections across all correction timing percentages (Fig. 4). At the same time, both models showed declining F1 scores and predicted/actual correction ratios as later corrections were considered, while the correction timing becomes more accurate.

(2) **Feature Importance**: For feature importance in Fig. 6, we observed that different trajectory features affected the model's F1 score at different stages of the trajectory, mostly resulting in a decrease when the feature was removed. Some feature removals led to minimal or insignificant changes. Boltzmann optimality removal led to performance drops at 80%, 90%, and 100%. Optimal velocity alignment removal decreased performance at 80%, direct velocity alignment at 80% and 90%, and velocity consistency when corrections occurring at the end of the trajectory (100%) were included. However, distance to goal removal led to a noticeable increase in F1 score at 70%, 80%, and 90% of the trajectory.

(3) **Goal Inference from *Start* of Correction**: As shown in Fig. 7a, when using the grasping location c_p and initial velocity c'_p to infer the goal distribution, the COMBINED model achieves significantly lower KLD than either the WHEN or WHERE models alone for each actual target before the latest corrections are included (70%, 80%, 90%). However, when later corrections are included (100%), the COMBINED model performs worse than the WHERE model.

(4) **Goal Inference from *End* of Correction**: As shown in Fig. 7b, when using the leaving location c_l to infer the goal distribution, the COMBINED model achieves lower KLD than the WHEN model but does not outperform the WHERE model across all correction timing percentages.

7 Discussion

Can we predict intervention timing using robot motion features? We demonstrate that the model successfully predicts both *whether* and *when* a correction will occur, achieving strong F1 scores (Fig. 3), high correction capture ratios (Fig. 4), and low mean absolute timing errors (Fig. 5). These results indicate that human intervention decisions are non-Markovian — they depend not only on the current state but also on the preceding trajectory history. Moreover, our multi-feature model consistently outperforms the Boltzmann baseline, particularly when later corrections are included, suggesting that human intervention decisions are influenced by multiple aspects of robot motion rather than a single optimality-based factor.

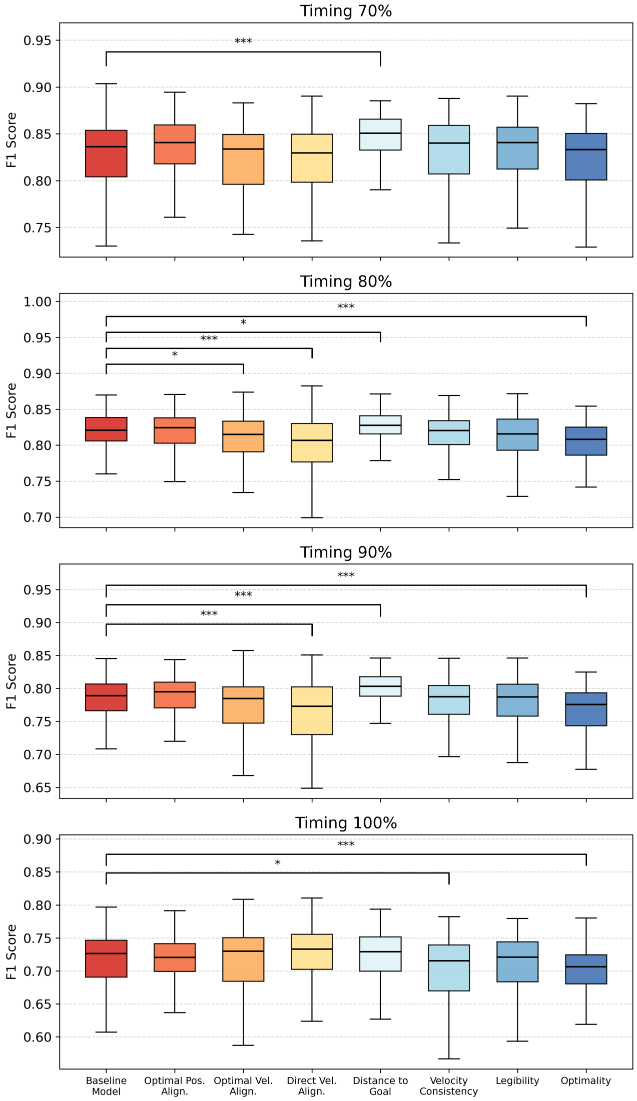
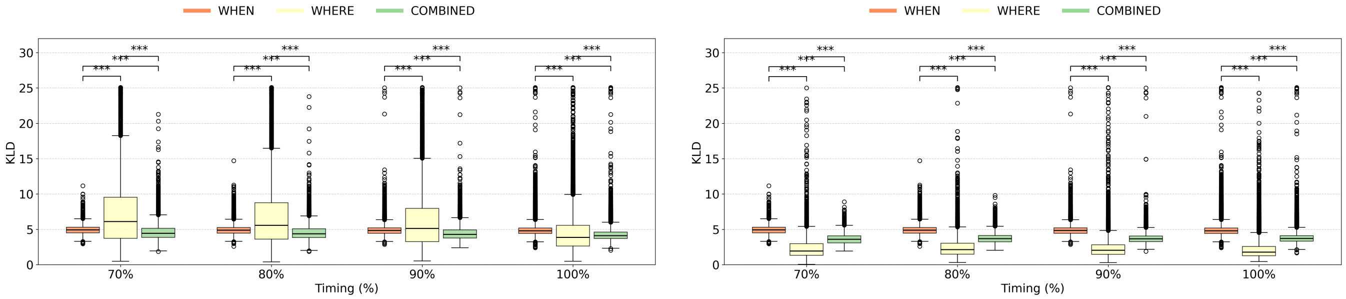


Figure 6: F1 scores for the baseline model and ablation models excluding individual features, evaluated across different correction timing percentages. Statistical significance indicates that the F1 score differs significantly from the baseline, suggesting that the removed feature has a meaningful influence on prediction performance. A decrease in F1 score implies that the excluded feature contributed positively to the model.

Which features of robot motion influence intervention decisions? The feature ablation results (Fig. 6) show that multiple motion features jointly influence correction prediction, with no single feature causing a drastic performance drop. Removing optimality consistently reduces performance at 80%, 90%, and 100%, indicating its stable relevance for later-stage corrections. Direct velocity alignment has the strongest impact at 80% and 90%, suggesting it shapes mid-trajectory intervention decisions. Velocity consistency affects performance only at 100%, aligning with the



(a) Results when using *start*-of-correction data to infer goal distribution.

(b) Results when using *end*-of-correction data to infer goal distribution.

Figure 7: KLD between the ground-truth goal distributions and the inferred goal after each correction in the test set, evaluated over 50 random dataset splits. Results are shown for each goal inference model (WHEN, WHERE, and COMBINED) across different correction timing percentages. The results are aggregated across all targets and shapes.

robot's natural deceleration near task completion, while optimal velocity alignment matters mainly at 80%. Interestingly, removing distance-to-goal improves performance, suggesting that humans attend less to absolute goal distance and more to motion cues — such as alignment and efficiency — that signal whether the robot behaves as expected.

Can timing information improve inference of human intended goals and task objectives? We find that when using information available at the onset of a correction, the COMBINED model achieves higher goal inference accuracy than both the WHEN and WHERE models alone for earlier corrections (70%, 80%, and 90%) (Fig. 7a). This indicates that timing information enhances early goal inference from correction onset before the correction completes.

In contrast, when goal inference is based on the release position of the gripper, the COMBINED model does not outperform the WHERE model (Fig. 7b). In our dataset, participants typically released the gripper very close to the intended goal, making the release position itself an almost complete indicator of the goal distribution. Consequently, timing information adds little value for refining task objectives in this setting.

In what settings it useful to incorporate timing for learning from corrections? Correction timing enhances prediction of participants' intended correction goals when combined with initial contact information, enabling earlier inference of intent without observing the full correction. Building on shared autonomy frameworks like Javdani et al. [23], such early intent prediction could help robots use feedback more efficiently and provide proactive assistance as soon as an intervention begins.

Although timing does not improve task constraint inference from correction endpoints, this likely reflects our task's simplicity; in more complex settings with less direct correction–goal mappings, timing may remain a valuable complementary signal.

When are human corrections most informative? The timing prediction model performs best for mid-trajectory corrections, showing higher F1 scores (Fig. 3) and correction detection rates (Fig. 4). In our dataset, many interventions happen very late (Fig. 2), with participants often waiting until the end of the task to intervene. These late-stage corrections are harder to predict due to the limited number of positive samples (less time steps before which correction occurs) and provide less informative feedback as much of the task has already been executed. The smaller mean timing error at late stages (Fig. 5) likely results from the greater number of late corrections.

Consistent with this, early goal inference performs best when corrections occur earlier, whereas the benefit of timing information diminishes for late corrections (100%) (Fig. 7a). Since participants often grasp the robot near the goal, the WHERE model already predicts the intended endpoint accurately, leaving little room for timing cues to help. Additionally, with minimal trajectory remaining, late corrections provide limited opportunity for timing-based inference.

Limitations and Future Work. We found no improvement in goal inference from post-correction positions, likely because our task was simple and direct — the end position already revealed the goal. Future work will explore more complex settings where corrections are less explicit, and timing may better reveal underlying task constraints. While our model predicts whether and when corrections occur, it has yet to be integrated into real-time planning. Embedding this predictive ability into online control could help robots anticipate human goals and adapt dynamically. Finally, our current feature set, though effective, may not capture all factors driving human corrections; richer, task-specific features could enhance future models.

8 Conclusion

Our results show that correction timing provides meaningful insight into human intervention behavior and offers valuable information for early goal inference. We successfully model both whether and when humans issue corrections, showing that these decisions depend on the trajectory history rather than the current state alone. Feature ablation analysis indicates that multiple aspects of robot motion jointly influence intervention timing. Based on this, integrating timing with spatial cues in the COMBINED model improves goal inference accuracy, enabling earlier prediction of human intended correction goal and supporting more proactive robot assistance. However, timing adds little benefit when goal inference is based on the gripper's final release position, as this already provides a near-complete signal of the task goal. Overall, correction timing is most valuable for early intent inference when spatial information is incomplete, highlighting its role as a complementary signal for learning from human feedback.

Acknowledgments

If you wish to include any acknowledgments in your paper (e.g., to people or funding agencies), please do so using the 'acks' environment. Note that the text of your acknowledgments will be omitted if you compile your document with the 'anonymous' option.

References

[1] Abdullah Cihan Ak, Eren Erdal Aksoy, and Sanem Sariel. 2023. Learning failure prevention skills for safe robot manipulation. *IEEE Robotics and Automation Letters* 8, 12 (2023), 7994–8001.

[2] Christopher G Atkeson and Stefan Schaal. 1997. Robot learning from demonstration. In *ICML*, Vol. 97. 12–20.

[3] Andrea Bajcsy, Dylan P Losey, Marcia K O'Malley, and Anca D Dragan. 2018. Learning from physical human corrections, one feature at a time. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 141–149.

[4] Andrea Bajcsy, Dylan P Losey, Marcia K O'malley, and Anca D Dragan. 2017. Learning robot objectives from physical human interaction. In *Conference on Robot Learning*. PMLR, 217–226.

[5] Chris L Baker, Joshua B Tenenbaum, and Rebecca R Saxe. 2007. Goal inference as inverse planning. In *Proceedings of the annual meeting of the cognitive science society*, Vol. 29.

[6] Andreea Bobu, Andrea Bajcsy, Jaime F Fisac, Sampada Deglurkar, and Anca D Dragan. 2020. Quantifying hypothesis space misspecification in learning from human–robot demonstrations and physical corrections. *IEEE Transactions on Robotics* 36, 3 (2020), 835–854.

[7] Andreea Bobu, Andrea Bajcsy, Jaime F Fisac, and Anca D Dragan. 2018. Learning under misspecified objective spaces. In *Conference on robot learning*. PMLR, 796–805.

[8] Justin Carpentier and Nicolas Mansard. 2018. Analytical derivatives of rigid body dynamics algorithms. In *Robotics: Science and Systems (RSS)*.

[9] Erwin Coumans and Yunfei Bai. 2016. Pybullet, a python module for physics simulation for games, robotics and machine learning.

[10] Erwin Coumans and Yunfei Bai. 2016–2021. PyBullet, a Python module for physics simulation for games, robotics and machine learning. http://pybullet.org.

[11] Longchao Da, Justin Turnau, Thirulogasankar Pravan Kutralingam, Alvaro Velasquez, Paulo Shakarian, and Hua Wei. 2025. A survey of sim-to-real methods in rl: Progress, prospects and challenges with foundation models. *arXiv preprint arXiv:2502.13187* (2025).

[12] Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. 2013. Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 301–308.

[13] Deepak E Gopinath and Brenna D Argall. 2020. Active intent disambiguation for shared control robots. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28, 6 (2020), 1497–1506.

[14] Deepak E Gopinath, Andrew Thompson, and Brenna D Argall. 2022. Information Theoretic Intent Disambiguation via Contextual Nudges for Assistive Shared Control. In *International Workshop on the Algorithmic Foundations of Robotics*. Springer, 239–255.

[15] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[16] Ryan Hoque, Ajay Mandlekar, Caelan Reed Garrett, Ken Goldberg, and Dieter Fox. 2023. Interventional data generation for robust and data-efficient robot imitation learning. In *First Workshop on Out-of-Distribution Generalization in Robotics at CoRL 2023*.

[17] Jiajing Huang, Hyunsoo Yoon, Teresa Wu, Kasim Selcuk Candan, Ojas Pradhan, Jin Wen, and Zheng O'Neill. 2023. Eigen-Entropy: A metric for multivariate sampling decisions. *Information sciences* 619 (2023), 84–97.

[18] Santiago Iregui, Joris De Schutter, and Erwin Aertbelien. 2021. Reconfigurable constraint-based reactive framework for assistive robotics with adaptable levels of autonomy. *IEEE Robotics and Automation Letters* 6, 4 (2021), 7397–7405.

[19] Abhinav Jain, Daphne Chen, Dhruva Bansal, Sam Scheele, Mayank Kishore, Hritik Sapra, David Kent, Harish Ravichandar, and Sonia Chernova. 2020. Anticipatory human-robot collaboration via multi-objective trajectory optimization. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 11052–11057.

[20] Ashesh Jain, Shikhar Sharma, Thorsten Joachims, and Ashutosh Saxena. 2015. Learning preferences for manipulation tasks from online coactive feedback. *The International Journal of Robotics Research* 34, 10 (2015), 1296–1313.

[21] Ashesh Jain, Brian Wojcik, Thorsten Joachims, and Ashutosh Saxena. 2013. Learning trajectory preferences for manipulators via iterative improvement. *Advances in neural information processing systems* 26 (2013).

[22] Siddarth Jain and Brenna Argall. 2019. Probabilistic human intent recognition for shared autonomy in assistive robotics. *ACM Transactions on Human-Robot Interaction (THRI)* 9, 1 (2019), 1–23.

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

[24] Rajat Kumar Jenamani, Daniel Stabile, Ziang Liu, Abrar Anwar, Katherine Dimitropoulou, and Tapomayukh Bhattacharjee. 2024. Feel the Bite: Robot-Assisted Inside-Mouth Bite Transfer using Robust Mouth Perception and Physical Interaction-Aware Control. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 313–322.

[25] Wanxin Jin, Todd D Murphey, Zehui Lu, and Shaoshuai Mou. 2022. Learning from human directional corrections. *IEEE Transactions on Robotics* 39, 1 (2022), 625–644.

[26] Mrinal Kalakrishnan, Peter Pastor, Ludovic Righetti, and Stefan Schaal. 2013. Learning objective functions for manipulation. In *2013 IEEE International Conference on Robotics and Automation*. IEEE, 1331–1336.

[27] Martin Karlsson, Anders Robertsson, and Rolf Johansson. 2017. Autonomous interpretation of demonstrations for modification of dynamical movement primitives. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 316–321.

[28] Min Jun Kim, Fabian Beck, Christian Ott, and Alin Albu-Schäffer. 2019. Model-free friction observers for flexible joint robots with torque measurements. *IEEE Transactions on Robotics* 35, 6 (2019), 1508–1515.

[29] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[30] Yigit Korkmaz and Erdem Bıyık. 2025. Mile: Model-based intervention learning. *arXiv preprint arXiv:2502.13519* (2025).

[31] Solomon Kullback and Richard A Leibler. 1951. On information and sufficiency. *The annals of mathematical statistics* 22, 1 (1951), 79–86.

[32] Sulabh Kumra, Shirin Joshi, and Ferat Sahin. 2021. Learning robotic manipulation tasks via task progress based gaussian reward and loss adjusted exploration. *IEEE Robotics and Automation Letters* 7, 1 (2021), 534–541.

[33] Cassidy Laidlaw and Anca Dragan. 2022. The boltzmann policy distribution: Accounting for systematic suboptimality in human models. *arXiv preprint arXiv:2204.10759* (2022).

[34] Dylan P Losey, Andrea Bajcsy, Marcia K O'Malley, and Anca D Dragan. 2022. Physical interaction as communication: Learning robot objectives online from human corrections. *The International Journal of Robotics Research* 41, 1 (2022), 20–44.

[35] Pauline Maurice, Meghan E Huber, Neville Hogan, and Dagmar Sternad. 2017. Velocity-curvature patterns limit human–robot physical interaction. *IEEE robotics and automation letters* 3, 1 (2017), 249–256.

[36] Shaunak A. Mehta and Dylan P. Losey. 2024. Unified Learning from Demonstrations, Corrections, and Preferences during Physical Human–Robot Interaction. *J. Hum.-Robot Interact.* 13, 3, Article 39 (Aug. 2024), 25 pages. https://doi.org/10.1145/3623384

[37] Katherine J Mimnaugh, Markku Suomalainen, Israel Becerra, Eliezer Lozano, Rafael Murrieta-Cid, and Steven M LaValle. 2021. Defining preferred and natural robot motions in immersive telepresence from a first-person perspective. *arXiv preprint arXiv:2102.12719* (2021).

[38] Brian A Nosek. 1999. Response latency in social psychological research. (1999).

[39] Michael Paierl, Anneliese Kelterer, and Barbara Schuppler. 2025. Distribution and Timing of Verbal Backchannels in Conversational Speech: A Quantitative Study. *Languages* 10, 8 (2025), 194.

[40] Nathan D Ratliff, J Andrew Bagnell, and Martin A Zinkevich. 2006. Maximum margin planning. In *Proceedings of the 23rd international conference on Machine learning*. 729–736.

[41] William Saunders, Girish Sastry, Andreas Stuhlmueller, and Owain Evans. 2017. Trial without error: Towards safe reinforcement learning via human intervention. *arXiv preprint arXiv:1707.05173* (2017).

[42] Mariah L Schrum, Erin Hedlund-Botti, Nina Moorman, and Matthew C Gombolay. 2022. Mind meld: Personalized meta-learning for robot-centric imitation learning. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 157–165.

[43] Pavel Senin. 2008. Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA* 855, 1-23 (2008), 40.

[44] Jonathan Spencer, Sanjiban Choudhury, Matthew Barnes, Matthew Schmittle, Mung Chiang, Peter Ramadge, and Sidd Srinivasa. 2022. Expert intervention learning: An online framework for robot learning from explicit and implicit human feedback. *Autonomous Robots* (2022), 1–15.

[45] Ioan A. Şucan, Mark Moll, and Lydia E. Kavraki. 2012. The Open Motion Planning Library. *IEEE Robotics & Automation Magazine* 19, 4 (December 2012), 72–82.

https://doi.org/10.1109/MRA.2012.2205651 https://ompl.kavrakilab.org.

[46] Ran Tian, Chenfeng Xu, Masayoshi Tomizuka, Jitendra Malik, and Andrea Bajcsy. 2023. What matters to you? towards visual representation alignment for robot learning. *arXiv preprint arXiv:2310.07932* (2023).

[47] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[48] Shuangge Wang, Anjiabei Wang, Sofiya Goncharova, Brian Scassellati, and Tesca Fitzgerald. 2025. Effects of Robot Competency and Motion Legibility on Human Correction Feedback. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 789–799. https://doi.org/10.1109/HRI61500.2025.10974241

[49] Nils Wilde, Alexandru Blidaru, Stephen L Smith, and Dana Kulić. 2020. Improving user specifications for robot behavior through active preference learning: Framework and evaluation. *The International Journal of Robotics Research* 39, 6 (2020), 651–667.

[50] Nils Wilde, Dana Kulić, and Stephen L Smith. 2018. Learning user preferences in robot motion planning through interaction. In *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 619–626.

[51] Nils Wilde, Dana Kulić, and Stephen L Smith. 2020. Learning user preferences from corrections on state lattices. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 4913–4919.

[52] Jiexin Xie, Zhenzhou Shao, Yue Li, Yong Guan, and Jindong Tan. 2019. Deep reinforcement learning with optimized reward functions for robotic trajectory planning. *IEEE Access* 7 (2019), 105669–105679.

[53] Qingfeng Yao, Jilong Wang, Shuyu Yang, Cong Wang, Hongyin Zhang, Qifeng Zhang, and Donglin Wang. 2022. Imitation and adaptation based on consistency: A quadruped robot imitates animals from videos using deep reinforcement learning. In *2022 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 1414–1419.

[54] Jason Y Zhang and Anca D Dragan. 2019. Learning from extrapolated corrections. In *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 7034–7040.

[55] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, Anind K Dey, et al. 2008. Maximum entropy inverse reinforcement learning.. In *Aaai*, Vol. 8. Chicago, IL, USA, 1433–1438.

A Appendix

A.1 Feature Details

A.1.1 Optimal Trajectories

We generate a reference trajectory using a PID controller following the original pre-planned trajectory to obtain the optimal velocities v_t^{opt} and optimal positions x_t^{opt} used to compute the alignment-related features. At each time step t along an original trajectory, the controller produced a "natural" trajectory from the robot's current state—specifically its position x_t and velocity v_t—toward the goal G^*. We consider this reference trajectory to be "optimal" in the sense that it reflects what a human might intuitively expect the robot to do from that point onward, based on the smoothness and predictability of PID-generated motion [35]. We then compute the position and velocity by interpolating the dynamics 0.25 seconds into the future along this reference trajectory to represent x_t^{opt} and v_t^{opt}, respectively.

A.1.2 Legibility

To represent legibility, we adopt the formulation from Dragan et al. [12] and compute a score l_t at each time step t as:

$$l_t = \frac{\int \mathbb{P}(G^* \mid \xi_{S \to x_t})\, \gamma_t \, dt}{\int \gamma_t \, dt} \tag{7}$$

where $\gamma_t = T - t$ is a discount factor that gives more weight to earlier parts of the trajectory and T is the total trajectory duration. The term $\mathbb{P}(G^* \mid \xi_{S \to x_t})$ is the probability that the robot is going to goal G^*, given the observed trajectory prefix $\xi_{S \to x_t}$, and is computed as:

$$\mathbb{P}(G^* \mid \xi_{S \to x_t}) \propto \frac{\exp\left(-C(\xi_{S \to x_t}) - C(\xi^{\text{opt}}_{x_t \to G^*})\right)}{\exp\left(-C(\xi^{\text{opt}}_{S \to G^*})\right)} \mathbb{P}(G^*) \tag{8}$$

This probability measures how efficient it is to reach the goal G^* via the current observed trajectory snippet $\xi_{S \to x_t}$, followed by an optimal completion $\xi^{\text{opt}}_{x_t \to G^*}$, relative to the cost of the optimal trajectory $\xi^{\text{opt}}_{S \to G^*}$. In our implementation, we define the cost function $C(\cdot)$ as the total trajectory length. Higher legibility indicates that the robot's motion strongly suggests its intention to reach the goal G^* early on.

A.2 Derivation of the Goal Inference Formulation

A.2.1 Inferring from Where People Grasp the Gripper

First we derive the posterior over goals given the release position is

$$\mathbb{P}(g \mid c_l) \propto \mathbb{P}(c_l \mid g) \cdot \mathbb{P}(g) \propto \mathbb{P}(c_l \mid g), \tag{9}$$

assuming a uniform prior over goals.

During testing, for each trajectory, given the grasp position c_p and velocity c'_p, the MLP predicts the release position:

$$\hat{c}_l = f_{\text{MLP}}(c_p, c'_p). \tag{10}$$

We assume c_l and $\hat{c}_l$ follow the same distribution. Through the MLP and using the Baye's Rule from Eq. 9, the posterior over goals conditioned on the grasp position and velocity becomes (defined as **WHERE model** goal distribution):

$$\mathbb{P}(g \mid c_p, c'_p) = \mathbb{P}(g \mid \hat{c}_l) \propto \mathbb{P}(\hat{c}_l \mid g), \tag{11}$$

where $\mathbb{P}(\hat{c}_l \mid g)$ is calculated using the GMM fitted for $\mathbb{P}(c_l \mid g)$.

A.2.2 Combining When and Where

To derive the posterior over goals using both spatial and timing information:

$$\mathbb{P}(g \mid t_c, c_p, c'_p, \xi) = \frac{\mathbb{P}(t_c, c_p, c'_p, \xi \mid g)\mathbb{P}(g)}{\mathbb{P}(t_c, c_p, c'_p, \xi)} \tag{12}$$

We expand the joint probability in the numerator using the chain rule:

$$\mathbb{P}(t_c, c_p, c'_p, \xi \mid g) = \mathbb{P}(t_c, c_p, c'_p \mid g, \xi)\, \mathbb{P}(\xi \mid g). \tag{13}$$

In our setup, the robot's pre-planned trajectories are generated independently of the sampled goal hypotheses. Therefore, we assume that the trajectory ξ and the goal g are independent a priori, such that $\mathbb{P}(\xi \mid g) = \mathbb{P}(\xi)$. Substituting this into Eq. 12 gives:

$$\mathbb{P}(g \mid t_c, c_p, c'_p, \xi) = \frac{\mathbb{P}(t_c, c_p, c'_p \mid g, \xi)\, \mathbb{P}(g)}{\mathbb{P}(t_c, c_p, c'_p \mid \xi)}. \tag{14}$$

We assume conditional independence between the correction timing t_c and the correction initiation parameters (c_p, c'_p) given the goal g and the trajectory ξ, as they capture distinct aspects of the human intervention process: *when* the human decides to intervene, whereas (c_p, c'_p) describe *how* the correction is initiated spatially and dynamically.:

$$\mathbb{P}(t_c, c_p, c'_p \mid g, \xi) = \mathbb{P}(t_c \mid g, \xi)\, \mathbb{P}(c_p, c'_p \mid g, \xi). \tag{15}$$

This assumption is supported empirically in our correlation analysis (Appendix A.3.1).

Substituting back:

$$\mathbb{P}(g \mid t_c, c_p, c'_p, \xi) = \frac{\mathbb{P}(t_c \mid g, \xi)\mathbb{P}(c_p, c'_p \mid g, \xi)\mathbb{P}(g)}{\mathbb{P}(t_c, c_p, c'_p \mid \xi)} \tag{16}$$

The denominator is a normalization constant computed as the sum over all possible goals $\mathcal{G}$.

$$\mathbb{P}(t_c, c_p, c'_p \mid \xi) = \sum_{\hat{g} \in \mathcal{G}} \mathbb{P}(t_c \mid \hat{g}, \xi)\mathbb{P}(c_p, c'_p \mid \hat{g}, \xi)\mathbb{P}(\hat{g}) \tag{17}$$

Thus, we have

$$\mathbb{P}(g \mid t_c, c_p, c'_p, \xi) = \frac{\mathbb{P}(t_c \mid g, \xi)\mathbb{P}(c_p, c'_p, \mid g, \xi)\mathbb{P}(g)}{\sum_{\hat{g} \in \mathcal{G}} \mathbb{P}(t_c \mid \hat{g}, \xi)\mathbb{P}(c_p, c'_p \mid \hat{g}, \xi)\mathbb{P}(\hat{g})} \tag{18}$$

To calculate $\mathbb{P}(c_p, c'_p \mid \hat{g}, \xi)$, we assume that c_p, c'_p is independent of ξ since the correction initiation should reflect human's intent based on the task goal rather than the shape of the robot's pre-planned trajectory, so it becomes $\mathbb{P}(c_p, c'_p \mid \hat{g})$. This assumption is partially supported by our correlation analysis in Appendix A.3.2, where the grasp position c_p exhibits negligible correlation with trajectory similarity across different ξ's. c'_p is dependent on where the correction begins, so the weak dependence of c_p suggests that the influence of ξ on correction initiation can be reasonably ignored for our modeling purposes.

Eventually, the equation becomes:

$$\mathbb{P}(g \mid t_c, c_p, c_p', \xi) = \frac{\mathbb{P}(t_c \mid g, \xi)\mathbb{P}(c_p, c_p', \mid g)\mathbb{P}(g)}{\sum_{\hat{g} \in \mathcal{G}} \mathbb{P}(t_c \mid \hat{g}, \xi)\mathbb{P}(c_p, c_p' \mid \hat{g})\mathbb{P}(\hat{g})} \tag{19}$$

Potential goals are uniformly sampled, thus $P(\hat{g})$ is uniform, so the posterior (**COMBINED model** goal distribution posterior) becomes:

$$\mathbb{P}(g \mid t_c, c_p, c_p', \xi) = \frac{\mathbb{P}(t_c \mid g, \xi)\mathbb{P}(c_p, c_p' \mid g)}{\sum_{\hat{g} \in \mathcal{G}} \mathbb{P}(t_c \mid \hat{g}, \xi)\mathbb{P}(c_p, c_p' \mid \hat{g})} \tag{20}$$

where $\mathbb{P}(t_c|g, \xi)$ is derived from Eq.1, and through the MLP:

$$\mathbb{P}(c_p, c_p' \mid g) \propto \mathbb{P}(\hat{c}_l \mid g) = \mathbb{P}_{\text{GMM}}(\text{MLP}(c_p, c_p') \mid g). \tag{21}$$

Thus,

$$\mathbb{P}(g \mid t_c, c_p, c_p', \xi) = \frac{\mathbb{P}(t_c \mid g, \xi)\, \mathbb{P}_{\text{GMM}}(\text{MLP}(c_p, c_p') \mid g)}{\sum_{\hat{g} \in \mathcal{G}} \mathbb{P}(t_c \mid \hat{g}, \xi)\, \mathbb{P}_{\text{GMM}}(\text{MLP}(c_p, c_p') \mid \hat{g})} \tag{22}$$

A.3 Correlations

A.3.1 Correlation Between Correction Timing t_c and Spatial Information (c_p, c_p')

To verify the assumption that correction timing is largely independent of spatial displacement, we computed the Pearson correlation between the normalized correction timing (in seconds) and the grasp positions relative to the actual goal and the initial correction velocity along each spatial dimension. Across all 3,589 filtered correction samples, we observe the following correlations for the grasp position c_p and the initial correction velocity c_p':

Table 2: Pearson correlation between normalized correction timing t_c and correction spatial features.

	Corr. w/ x	Corr. w/ y	Corr. w/ z
Grasp position c_p	0.092	0.039	-0.600
Initial correction velocity c_p'	0.077	-0.010	0.264

The results show weak correlations between t_c and both x and y components of c_p and c_p', suggesting that the timing of intervention is largely independent of where and how participants initiate the correction in the horizontal plane. The moderate correlations along the z-axis likely reflect the vertical component of the motion, which is peripheral to goal inference that primarily occurs in the x–y plane. Therefore, we approximate t_c and (c_p, c_p') as conditionally independent given g and ξ.

A.3.2 Correlation Between Grasp Position c_p and Trajectory ξ Similarities

To verify that the spatial similarity between grasp locations is independent from the overall trajectory similarity, we computed both dynamic time warping (DTW) [43] distances between full correction trajectories and Euclidean distances between their corresponding correction endpoints. From 3,589 correction trajectories, we randomly sampled 353,500 trajectory pairs and calculated the Pearson and Spearman correlations between the two distance measures.

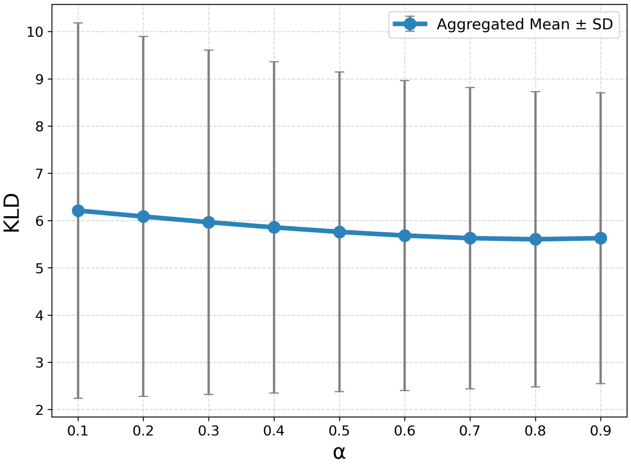


Figure 8: KLD vs. different weight α

The Pearson correlation coefficient was $r = 0.02$ ($p < 0.001$) and the Spearman rank correlation was $\rho = 0.04$ ($p < 0.001$), both indicating negligible relationships between trajectory and endpoint similarities. These results confirm that trajectories that are temporally or kinematically similar do not necessarily result in spatially similar correction endpoints. Therefore, it is reasonable to treat the temporal and spatial components of human corrections as independent in our modeling framework.

A.4 Model Details

A.4.1 Transformer

We optimize binary cross-entropy over valid time steps only, using the mask M as sample weights. We use Adam [29] with gradient clipping ($\|g\|_2$ clipnorm = 1.0) and an exponentially decaying learning rate (η_0=10^{-3}, decay_steps = 1000, decay_rate = 0.9). Model selection is done via a held-out validation set by monitoring validation loss and saving the best checkpoint. Random seeds are fixed for reproducibility. Input features are rescaled with the same preprocessing on train/validation splits.

A.4.2 MLP

The model is trained using mean squared error (MSE) loss and optimized with the Adam [29] optimizer with an exponentially decayed learning rate. Validation loss is monitored at each epoch, and the best-performing model is saved based on the lowest validation loss. All inputs and outputs are normalized using statistics computed from the training set to ensure stable convergence.

A.5 Geometric Weight α

To determine the balance between timing and spatial information in the COMBINED model, we varied $\alpha \in [0.1, 0.9]$ in the weighted posterior:

$$\mathbb{P}_w(g \mid t_c, c_p, c_p', \xi) = \frac{\mathbb{P}(t_c \mid g, \xi)^{\alpha} \cdot \mathbb{P}_{\text{GMM}}(\text{MLP}(c_p, c_p') \mid g)^{1-\alpha}}{\sum_{\hat{g} \in \mathcal{G}} \mathbb{P}(t_c \mid \hat{g}, \xi)^{\alpha} \cdot \mathbb{P}_{\text{GMM}}(\text{MLP}(c_p, c_p') \mid \hat{g})^{1-\alpha}} \tag{23}$$

For testing, we evaluated how the KLD between the inferred and ground-truth goal distributions (using correction onset information c_p, c'_p) varies with α at 80% correction timing percentages, aggregated across all targets and shapes over 5 runs. The results in Fig. 8 show that $\alpha = 0.8$ yields the lowest mean KLD. Therefore, we set $\alpha = 0.8$ as the weighting parameter for all COMBINED models.

A.6 KLD Plot for Different Targets Using Correction Onset Information

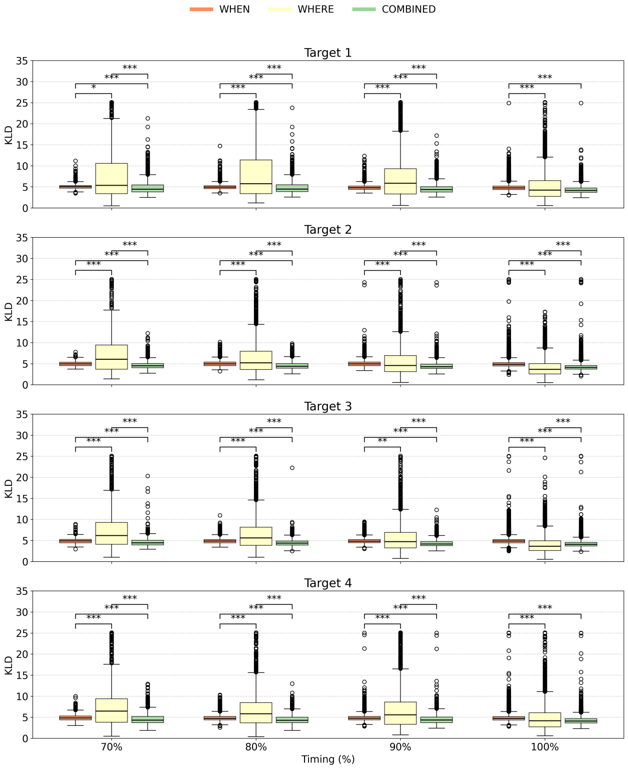


Figure 9: KLD between the ground-truth goal distributions and the inferred goal using the start of the correction information after each correction in the test set, evaluated over 50 random dataset splits. Results are shown for each goal inference model (WHEN, WHERE, and COMBINED) across different correction timing percentages. The results are aggregated across shapes.

We evaluate goal inference performance at each target using the grasp location c_p and initial velocity c'_p as inputs. As shown in Fig. 9, the COMBINED model yields significantly lower KLD than either the WHEN or WHERE models for all targets when earlier corrections are considered (70%, 80%, 90%). When late corrections are included (100%), the COMBINED model performs comparably to WHERE for side targets (targets 1 and 4; no significant difference) but slightly worse for center targets (targets 2 and 3). These results suggest that timing information is most beneficial for early corrections and becomes less informative when corrections occur late and target distinctions are less pronounced.