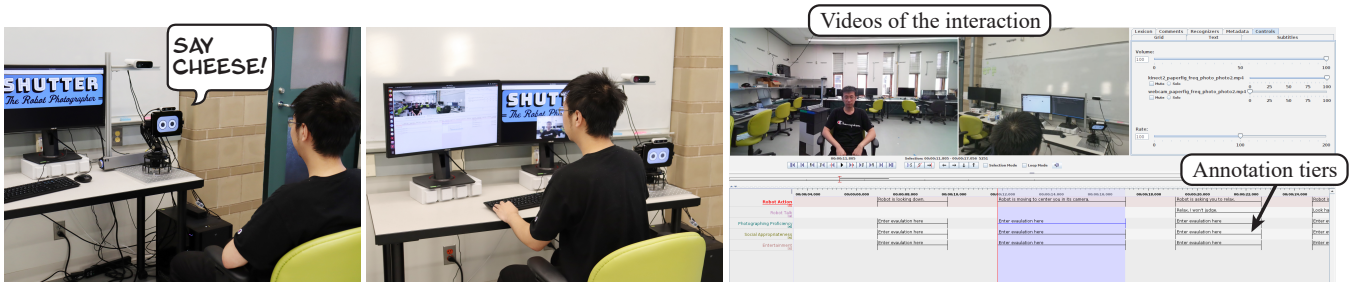
# Self-Annotation Methods for Aligning Implicit and Explicit Human Feedback in Human-Robot Interaction

Qiping Zhang
Yale University
qiping.zhang@yale.edu

Austin Narcomey
Yale University
austin.narcomey@yale.edu

Kate Candon
Yale University
kate.candon@yale.edu

Marynel Vázquez
Yale University
marynel.vazquez@yale.edu




(a) Human-Robot Interaction (b) Self-Annotation (c) ELAN Annotation Interface

**Figure 1: In our study, a participant interacted with a robot photographer (a) and then self-annotated the robot's performance (b) based on its *proficiency* in the photography task, the *social appropriateness* of its behavior, and its *entertainment* value. Each of these performance dimensions was annotated action-by-action in one tier of ELAN considering videos of the interaction (c).**

## ABSTRACT

Recent research in robot learning suggests that implicit human feedback is a low-cost approach to improving robot behavior without the typical teaching burden on users. Because implicit feedback can be difficult to interpret, though, we study different methods to collect fine-grained labels from users about robot performance across multiple dimensions, which can then serve to map implicit human feedback to performance values. In particular, we focused on understanding the effects of annotation order and frequency on human perceptions of the self-annotation process and the usefulness of the labels for creating data-driven models to reason about implicit feedback. Our results demonstrate that different annotation methods can influence perceived memory burden, annotation difficulty, and overall annotation time. Based on our findings, we conclude with recommendations to create future implicit feedback datasets in Human-Robot Interaction.

## CCS CONCEPTS

• **Human-centered computing → User studies**; • **Computing methodologies → Learning from implicit feedback**.

## KEYWORDS

Interactive robot learning; Implicit human feedback; Self-annotation

This work is licensed under a Creative Commons Attribution International 4.0 License.

*HRI '23, March 13–16, 2023, Stockholm, Sweden*
© 2023 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9964-7/23/03.
https://doi.org/10.1145/3568162.3576986

**ACM Reference Format:**
Qiping Zhang, Austin Narcomey, Kate Candon, and Marynel Vázquez. 2023. Self-Annotation Methods for Aligning Implicit and Explicit Human Feedback in Human-Robot Interaction. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23), March 13–16, 2023, Stockholm, Sweden.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3568162.3576986

## 1 INTRODUCTION

Most prior work in robot behavior adaptation and learning from human teachers relies on explicit feedback signals, such as demonstrations [4], preference [60], and evaluative feedback [39, 40]. Unfortunately, though, explicit human feedback requires deliberate effort, may diminish over time [11, 38, 45], and can interrupt the natural flow of human-robot interactions [52].

Recently, research has begun to investigate implicit human communicative signals in Human-Robot Interaction (HRI) as a complement or alternative signal to explicit feedback, e.g., [9, 21, 34, 44, 50]. Implicit signals are inevitably given off by people when they care about an agent's behavior. They can be reflected in human actions that change the physical state of the world [37, 61]. Also, implicit human feedback can be provided through nonverbal cues — such as body motion and gaze — that people naturally display when they interact socially. Nonverbal signals serve as a passive and burden-free information channel for people that sometimes persists even when they don't intend to communicate [36].

This work investigates annotation methods to align implicit human feedback with explicit feedback in HRI, as shown in Fig. 1. The goal is to understand how we can effectively gather data and use supervised learning to map nonverbal human behavior to explicit values of a robot's performance. While it is common for prior work in HRI to rely on third-party coders to gather labels that describe human nonverbal behavior [8, 13, 43, 53, 62, 73], we explore the

feasibility of gathering these labels from the interactants themselves, i.e., gathering *self-annotations*. This idea is inspired by work in Human-Computer Interaction [35] and Affective Computing [12]. Although gathering self-annotations may be hard in many cases (e.g., when users are children [12]), we deem it critical to make progress in making sense of implicit human feedback in HRI due to the challenges that interpreting this feedback brings along.

Implicit human feedback tends to be more difficult to interpret and less informative than explicit feedback [2, 21, 65]. The reason is threefold. First, certain signals, like facial expressions, can have different meanings based on the context in which they are produced [22, 28, 32, 58]. Second, natural human reactions to important events may have varied delays [3, 29, 39]. Lastly, while human feedback and rewards in robot learning are often represented with a single scalar value [16], people may care about multiple aspects of a robot's behavior. For example, a person could care about the proficiency of a robot in a task [52] but also its entertainment value during interactions [1, 75]. Thus, people's reactions to a robot's behavior may not just be due to a single performance factor, but their perception of the robot across multiple dimensions.

We conducted a user study to understand trade-offs in *when* to gather self-annotations about robot performance, and *how* to collect these annotations when performance is a multi-dimensional construct. In our study, participants naturally interacted with a robot photographer, as shown in Figure 1(a). The robot sequentially completed photo-taking tasks with the participant while we recorded the person's nonverbal behavior, which included implicit feedback about the robot's performance. Every so often, participants were asked to annotate their impression of the robot's performance using a well-established graphical user interface (GUI) for video annotation, as shown in Figures 1(b) and 1(c). The GUI allowed them to see a recording of their recent interaction with the robot and annotate the robot's current level of *proficiency* at the task, its *social appropriateness*, and its *entertainment* value. Also, after a photo-taking task was complete, we asked the participants to provide more sparse task-level ratings about their impression of the robot's performance through a survey. We considered the self-annotations and survey responses as explicit human feedback in the study.

We evaluated different ways in which the annotation process took place from a human and a computational perspective. Interestingly, the delay between when interactions happened and annotations were gathered affected the difficulty of the annotation process. Also, the delay and the order in which annotations were provided influenced the length of the annotation process. On the computational side, implicit feedback data helped predict per-action annotation labels that explicitly measure robot performance. We conclude this paper with a discussion of our findings and recommendations to create future implicit feedback datasets for HRI.

## 2 BACKGROUND

Our work focuses on fine-grained annotation of robot performance based on users' impressions of the robot's behavior. Thus, we first discuss related work on self-annotation. Then, we contextualize our effort with respect to a growing body of research on Interactive Robot Learning [16, 20, 21, 39, 46, 67], where a robot learns from human feedback through interaction rather than simply using a pre-coded environmental reward.

### 2.1 Self-Annotation

It is common for work in Human-Computer Interaction [42] and HRI [7] to gather participant self-reports of personal traits and experiences engaging in social encounters. For example, Celiktutan et al. [15] collected self-assessed personality and engagement towards a robot through survey instruments and then used the answers as ground truth labels to model these factors with machine learning. Furthermore, Tsoi et al. [70] collected impressions of a mobile robot at scale using simulated human-robot interactions. The opinions of those that interacted with the robot in simulation differed from those that passively observed the simulated interactions in video recordings. In contrast to using third-party observers or raters (e.g., as in [8, 13, 14, 43, 53, 62, 73]), self-annotation can be more reliable for gathering explicit labels that are reflective of the participant's own opinions during interactions [54]. As suggested by Jung [35], self-annotations based on video footage can be an effective mechanism for participants to recall and assess the emotional dynamics of their social interactions.

The importance of self-assessment measures in HRI motivated us to study the possibility of using self-annotations to align implicit human reactions to a robot's behavior with explicit feedback about its performance. In particular, we chose to study how to collect these annotations using ELAN [76], a well-established annotation tool for audio and video recordings. ELAN has been used widely for research in linguistics and multi-modal interaction [19, 23, 30, 33, 41, 59]. ELAN's interface was particularly well suited for our study because we wanted to gather labels for multiple robot performance factors, which could easily be annotated using ELAN's layered annotated scheme based on *tiers* (as shown in Figure 1(c)). A tier is a set of annotations that share common characteristics, e.g., they might all be about robot proficiency.

Prior work has shown that annotation processes can result in cognitive load [74] and task switching costs [55], which may increase user response times and errors rates. We thought that these challenges would naturally translate to HRI, potentially posing a barrier to the effective collection of implicit feedback annotations. Thus, we sought to investigate:

***Research Question 1:*** *How do different self-annotation procedures influence human perception of the annotation process and the flow of human-robot interactions?*

### 2.2 Interactive Robot Learning

*2.2.1 Using Human Feedback as Reward.* For at least 20 years, researchers have investigated reward shaping by people in Reinforcement Learning [68]. For example, Isbell et al. [31] and Thomaz and Breazeal [69] trained a simulated robot by combining a human reward with a traditional environmental reward. More recently, Knox et al. [39, 40] proposed to explicitly model human feedback, such as a button press, and to learn entirely from human signals. Other work interpreted human-delivered training signals as alternative evaluative metrics (like optimality [25], an advantage function [49], or categorical feedback strategies [47]) or used the feedback for more direct policy learning [25, 49, 56].

Prior work in Interactive Robot Learning typically models sequential decision making as a Markov Decision Process, in which an agent aims to find a policy that maximizes the sum of rewards:

$$G = \sum_{t=0}^{\infty} \gamma^t r_t \tag{1}$$

where $r_t$ is the reward at time step $t$ and $\gamma$ is a discount factor. Our research is related to this line of work because we investigate methods to gather explicit feedback labels for implicit human feedback. This data could in turn be used to create reward functions from implicit feedback in HRI, as explained below.

*2.2.2 Implicit Human Feedback.* In addition to using explicit human feedback as a reward for robot learning, research has begun to explore using implicit feedback as well. For instance, Broekens [9] and Gordon et al. [24] explored interpreting human affective expressions as numeric reward signals using a predefined mapping from implicit feedback to social reinforcement. Veeriah et al. [72] and Zadok et al. [77] proposed to learn a value function or probabilistic model to ground the facial reactions of proficient observers (or demonstrators) and bias an agent's behavior. More recently, Li et al. [44] and Stiber et al. [65] demonstrated learning from human facial expressions only and Cui et al. [21] proposed to map facial reactions to task statistics, requiring no explicit human feedback. Unfortunately, these methods are data-hungry due to the high-dimensional nature of implicit signals, such as facial expressions. This motivated us to explore methods for collecting high-quality human feedback datasets from which we can build reliable models for interpreting implicit feedback:

***Research Question 2:*** *How do different self-annotation procedures for robot performance influence the usefulness of the data for mapping implicit signals to explicit measures of performance?*

One novel aspect of our work is that we consider robot performance a multi-dimensional construct. This consideration opens doors for future work in multi-objective learning in HRI [6, 18] using implicit human feedback in the form of nonverbal behavior.

## 3 METHOD

This section describes our study for evaluating self-annotation methods for aligning implicit and explicit human feedback in HRI. The study protocol was approved by our local Institutional Review Board and refined through pilot studies.

In our study, a participant interacted with a robot photographer twice. An interaction consisted of the robot completing 3 photo-taking tasks with the participant. Before taking a photo of the person, the robot adjusted the framing direction of its camera and/or chatted with the users, as in Fig. 1(a). The robot employed a sub-optimal policy while interacting with users, as it did not always choose in the most task-effective action given the context of the interaction. Thus, as the policy was executing, the behavior of the robot naturally prompted people to react to it, providing implicit feedback via their nonverbal behavior.

After every photo-taking task, participants provided explicit feedback about the performance of the robot through a survey. In general, we considered performance as a 3-dimensional construct that included: (1) the robot's *proficiency* at an interactive task with the user, (2) the *social appropriateness* of the robot's behavior, and (3) the robot's *entertainment* value during the interaction. We further refer to these factors as the *dimensions* of the robot's performance.

### 3.1 Annotation Procedures

We focused our study on evaluating the effect of two independent variables on the annotation experience and the data collected through self-annotation:

**Annotation Order:** The participants provided their opinion of the robot's performance on an action-by-action level using ELAN. They either annotated all three dimensions of robot performance together for a single robot action at a time (***By-Action*** order), or annotated all actions for a single performance dimension at a time (***By-Dimension*** order). These two annotation orders are explained in Fig. 2 in relation to ELAN's interface. Intuitively, these orders can be thought as completing the annotations by "column" (per action) or by "row" (per dimension or tier) of the ELAN timeline.

**Annotation Frequency:** Participants self-annotated the robot's performance after each photo-taking task was completed (***Per-Task*** frequency), or annotated the three tasks after the whole interaction had completed (***Per-Interaction*** frequency). These annotation frequencies are depicted in Fig. 3. Because an interaction was composed of several photo-taking tasks, the Per-Task frequency meant that participants provided self-annotations through ELAN more frequently than when providing annotations Per-Interaction.

Our independent variables led to 4 annotation procedures, each corresponding to a study condition: 1) By-Action & Per-Task (BA-PT), 2) By-Action & Per-Interaction (BA-PI), 3) By-Dimension & Per-Task (BD-PT), and 4) By-Dimension & Per-Interaction (BD-PI).

**Figure 2: In our study, participants provided self-annotations in one of two orders: By-Action, annotating one dimension at a time for every action (a); and By-Dimension, annotating one action at a time for every dimension (b). "Dim 1", "Dim 2" and "Dim 3" corresponded to the proficiency, social appropriateness, and entertainment dimensions of the robot's performance, respectively. The ❶ and ❷ symbols denote the order in which the annotations were provided.**

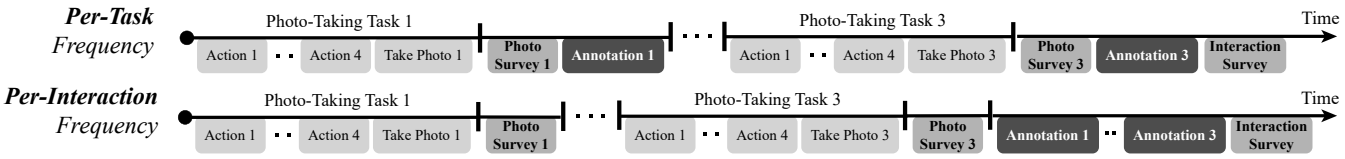



**Figure 3: In our study, participants provided self-annotations after a photo was taken (Per-Task frequency, top) and after a whole interaction (composed of 3 photo-taking tasks) was complete (Per-Interaction frequency, bottom). The order of these two procedures was counterbalanced across participants to account for any potential ordering effects.**

## 3.2 Study Design and Hypotheses

Our study had a mixed design: Annotation Order was run between-subjects, and Annotation Frequency was run within-subjects. We counterbalanced Per-Task and Per-Interaction frequencies to account for potential ordering effects. Thus, participants experienced 2 interactions with the robot, each with 3 photo-taking tasks.

We expected the self-annotation procedures to affect (1) people's perception of their human-robot interactions and the annotation process, and (2) the effectiveness of grounding implicit human feedback into explicit performance values. More specifically, in relation to our Research Question 1 (in Sec. 2.1), we hypothesized:

**H1.** Providing self-annotations with the Per-Task frequency will disrupt the interaction more than the Per-Interaction frequency.
**H2.** The Per-Task frequency will lead to less perceived workload, memory burden, and annotation difficulty than Per-Interaction.
**H3.** The annotation process with the By-Dimension order will be more time-consuming than with the By-Action order.

The first hypothesis, H1, is a direct result of the Per-Task frequency happening more often, in the middle of interactions, than the Per-Interaction frequency (as illustrated by the dark gray "Annotation" regions of Fig. 3). H2 is motivated by Per-Task frequency resulting in more dispersed periods of time in which participants provide self-annotations in comparison to Per-Interaction frequency, when self-annotations are all collected at the end of an interaction. Also, when participants provide self-annotations of the performance of the robot in ELAN, they need to remember how they felt when the interaction took place. We suspected that this would be easier with the Per-Task frequency because the interactions would have taken place closer in time to the annotation event than with the Per-Interaction frequency. Finally, H3 is motivated by the expectation that participants will need to re-watch the videos of the human-robot interaction multiple times with the By-Dimension order. With the By-Action order, they could instead watch a portion of the video of their interaction around the time of a given action and then label the performance of the robot across the three dimensions one after the other, without having to necessarily re-watch the recording.

In relation to our Research Question 2 (in Sec. 2.2), we expected:

**H4.** A predictive model trained with our study data will infer the explicit feedback labels from implicit nonverbal reactions better than: a) random guessing, and b) informed guessing based on the distribution of explicit feedback labels.
**H5.** Annotating By-Action & Per-Task (BA-PT) will result in a more accurate model for grounding implicit reactions to explicit feedback labels than the other conditions.

The last two hypotheses, H4 and H5, are defined in relation to each robot performance dimension. In particular, H4 is motivated by existing research on reasoning about human implicit feedback [21, 44, 72], which suggests that implicit human feedback can be mapped to task statistics (from which rewards can be inferred) or mapped directly to a reward value. Worth noting, to the best of our knowledge, our work is the first to explore disentangling implicit feedback into multiple robot performance dimensions. That is, H4 posits that we can create three models that map implicit human feedback to explicit ratings for robot proficiency, social appropriateness, and entertainment value, respectively. Each of these predictive models would be more accurate than random or informed guessing for the respective dimension. Finally, H5 is motivated by H2 and H3. We thought that By-Action order and Per-Task frequency would be faster and easier for the participants, resulting in the BA-PT condition providing labels with higher quality (more closely matching the implicit reactions) than the other conditions. In turn, higher quality labels would lead to a more accurate model for grounding implicit reactions to explicit feedback.

## 3.3 Setup

The study took place in a laboratory on a university campus. As shown in Fig. 4, the room contained a chair where the participant could sit, a desk with a desktop computer, a keyboard and mouse, two monitors, and a small table-top robot. The setup also included three sensors for recording the study and recognizing changes in the state of the interaction: (1) a Kinect Azure sensor was positioned above and slightly behind the robot to visually capture human reactions to the robot's behavior and perform face tracking of users during the interaction; (2) a RealSense RGB-D camera was mounted on the robot's head and used to take photos of participants; and (3) a Logitech webcam was placed behind the user to record the robot's behaviors during the interaction. The participants used the video from the Kinect and the webcam as a reference for self-annotations, as exemplified in Fig. 1(c) and Fig. 2.

The robot, called Shutter, had a screen face (with a RealSense camera on top) mounted on a 4 degrees-of-freedom arm. The arm allowed the face and camera to pan, tilt, and move up and down as well as forward and backwards. Inspired by Adamson et al. [1], we made the robot behave like a portrait photographer. The robot could use humor to elicit spontaneous reactions during photography events, change its pose to adjust its camera framing direction, and take photos of participants. While the robot was moving, its eyes continuously followed the participant via face tracking to demonstrate the robot's presence and engagement. The robot was

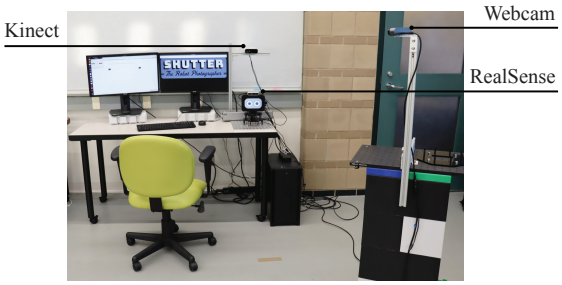



**Figure 4: Study setup. Participants sat at the desk as in Fig. 1.**

controlled with the Robot Operating System (ROS) [57] and its photos were displayed on a monitor placed next to it during the study. The other monitor on the desk was used to gather self-annotations.

## 3.4 Procedure

First, the participants completed a demographics survey. The experimenter then introduced the robot photographer, explained the photo-taking task, and demonstrated how to use the annotation tool to label the robot's performance. The participants were instructed to consistently follow a certain annotation order (By-Action or By-Dimension) throughout the study.

Afterwards, the participants experienced two interactions with the robot, each with a given annotation frequency (Per-Task or Per-Interaction) as exemplified in Fig. 3. In each interaction, the robot completed three photo-taking tasks and, after each task, the participant completed a survey to rate the robot's overall task performance based on our three dimensions (proficiency, social appropriateness, and entertainment value). This survey is depicted as "Photo Survey" in Fig. 3. Once an interaction and the corresponding annotations were complete, the participants answered a few additional survey questions regarding their perception of the annotation process with the last annotation frequency that they experienced. This survey corresponds to "Interaction Survey" in Fig. 3.

The study typically took 45 min to 1 hour to complete. Participants were compensated with USD $20 for their time.

*3.4.1 Photo-Taking Tasks.* Before taking a photo in a given photo-taking task, the robot sequentially executed four pre-defined actions that involved changes in the robot's pose and/or spoken dialog. Specifically, robot actions were of the following types:

*1. Center face.* The robot followed the participant to center their face on the RealSense image used to capture photos.

*2. Avoiding.* The robot intentionally oriented its head away from the participant's face.

*3. Fixed Poses.* The robot moved to a pre-defined pose (e.g., resting).

*4. Joke.* The robot told a joke verbally.

*5. Smile.* The robot told the participant to smile (e.g., as in Fig. 1(a)).

*6. Relax.* The robot told the participant to relax.

Unbeknownst to the participants, we used weighted random sampling to choose varied actions for the robot from the 6 action types in order to maintain a consistent rate of sub-optimal behavior. We prevented the same action type from consecutively occurring more than twice to avoid user boredom or confusion.

*3.4.2 Self-Annotation.* The participants reviewed their past interaction and performed self-annotation of the robot's performance using ELAN, as shown in Fig. 1. They could freely choose how they would replay the time-aligned videos of their interaction with the robot in order to recall what happened in the photo-taking tasks, but we required that they consistently followed the assigned annotation order for the entire study.

As shown in Fig. 2, there were 5 tiers under the timeline of the videos in ELAN: The first tier contained pre-generated text fields that described the actions taken by the robot; the second tier showed the robot's dialog if the robot talked during an action; and the remaining three tiers were used by the participants to indicate whether they thought that the robot performed well (in regard to its photography proficiency, social appropriateness, and entertainment value). Participants provided ratings for these three dimensions on a 7-point Likert responding format (with 1 being "strongly disagree" that the robot performed well and 7 being "strongly agree").

## 3.5 Measures

We collected a variety of measures based on the survey responses, the self-annotation data, and the videos recorded during the study.

**Human's Perception of the Annotation.** The interaction survey, which was administered after 3 photos and corresponding annotations (as in Fig. 3), asked the participants to indicate their perceived workload for the self-annotation task in ELAN. Inspired by the NASA Task Load Index method [26], we measured the workload through participant's agreement to five prompts in relation to the annotations: 1) "It was mentally demanding"; 2) "It was physically demanding"; 3) "I felt heavy time pressure"; 4) "I was successful"; 5) "I felt stressed". Additionally, the survey asked for participants' agreement with the statements "It was difficult for me to annotate the interaction", "It was hard to remember what happened during the interaction when annotating", and "The annotations were disruptive to the interaction". All these prompts were contextualized for the Per-Interaction frequency by making reference to "annotate a session as a whole" (e.g., "It was mentally demanding to annotate a session as a whole") and for the Per-Task frequency by making reference to "annotate a session separately for each photo". Ratings were provided on a 7-point Likert responding format. These measures served to evaluate H1, H2, and H3.

**Explicit Human Feedback.** Participants provided two types of explicit human feedback about the behavior of the robot in our study: *per-action explicit feedback* through the self-annotation process in ELAN, and *per-photo explicit feedback* through the photo survey after each task (as in Fig. 3). More specifically, the photo survey asked the participants to indicate if they agreed that the robot performed well while taking the last photo. For each of the performance dimensions, we gathered impressions of the robot's performance in these surveys with two statements, as shown in Table 1. Agreement was provided on a 7-point Likert responding format from "Strongly Disagree" (1) to "Strongly Agree" (7). The responses per dimension had high internal consistency with Cronbach's $\alpha$ above the nominal 0.7 threshold. Thus, we combined survey responses into a single performance value per dimension and per photo. The explicit feedback data was used to evaluate H4 and H5.

**Table 1: Items for measuring the robot's performance.**

| | |
|---|---|
| **Photography Proficiency** | (Cronbach's $\alpha$ = 0.9685) |
| - The robot is a competent photographer. | |
| - The robot is capable of taking my photo proficiently. | |
| **Social Appropriateness** | (Cronbach's $\alpha$ = 0.9303) |
| - The robot behaves in a socially acceptable manner. | |
| - The robot follows social norms. | |
| **Entertainment Value** | (Cronbach's $\alpha$ = 0.9259) |
| - The robot is entertaining. | |
| - The robot is amusing to interact with. | |

**Table 2: Participant demographics by Annotation Order.**

| Ann. Order | N | #Female | #Male | Avg. Age (SE) |
|---|---|---|---|---|
| By-Action | 20 | 10 | 10 | 24.20 (1.07) |
| By-Dimension | 20 | 8 | 12 | 25.20 (1.52) |
| All | 40 | 18 | 22 | 24.70 (0.92) |

**Implicit Human Feedback.** We used OpenFace 2.0 [5] to extract features from raw videos of human reactions captured by the Kinect sensor in our study setup (Fig. 4). The videos consisted of 30 image frames per second. For each image frame, OpenFace extracted head poses, gazes, and activation of facial action units in a [0,5] interval (0 representing no activation and 5 being maximum intensity). For our analysis with machine learning, we aggregated the predicted features using max and min pooling in each feature dimension extracted from consecutive image frames to allow the series of input features to cover a large enough temporal window of reactions. These implicit feedback features served as inputs to predictive models created for H4 and H5.

**Screen Recordings.** We captured screen recordings while participants were annotating their videos in ELAN. This data served to check that participants provided annotations in the right order (according to the experimental condition) and to compute the amount of time that it took participants to complete the annotations.

## 3.6 Participants

We recruited 46 participants using flyers and word of mouth. Participants' were required to be at least 18 years of age, fluent in English, and have normal or corrected-to-normal vision. The study had a final sample size of 40 participants due to technical problems with the robot and accidental major deviations from the experimenter's script in 6 sessions. A total of 20 participants were assigned to each Annotation Order, as shown in Table 2. Most participants were university students, and ages ranged from 19 to 48 years old. All participants reported using computers daily yet were somewhat unfamiliar with robots (M = 3.20, SE = 0.36) on a 7-point Likert responding format (1 being lowest). Only one participant in the By-Dimension order had interacted with the Shutter robot before, and another participant in BD had prior experience with ELAN.

## 3.7 Data Validation

Two members of our team manually inspected the screen recordings to verify that participants correctly followed the instructed Annotation Order. All the participants provided annotations as instructed and there were no missing annotations in the data.

We evaluated the consistency of explicit human feedback data between per-action annotations and per-photo survey ratings for each performance dimension. In this analysis, we considered the per-action annotations as an instantaneous reward $r$. Following eq. (1), we computed the sum of undiscounted rewards $G = \sum_{t=0}^{T} r_t$ and checked whether these values correlated with the per-photo explicit feedback from survey ratings. We found significant correlations in general ($p < 0.0001$). Pearson's correlation was 0.63, 0.67 and 0.68 for proficiency, social appropriateness, and entertainment value, respectively. These moderate-to-strong correlations suggest that the more fine-grained, per-action annotations are generally consistent with overall robot performance after a photo-taking task.

# 4 RESULTS

This section first presents our results for our first three hypotheses about human perception of the annotation process (RQ1 in Sec. 2.1). Then, it describes the results for our last two hypotheses about mapping implicit signals to explicit measures of performance (RQ2).

## 4.1 Perceptions of the Annotation Process

We used linear mixed models estimated with REstricted Maximum Likelihood (REML) [27, 66] to evaluate perceptions of the annotation process. Unless otherwise noted, the analyses considered Participant ID as a random effect, and Annotation Order (By-Action, By-Dimension), Annotation Frequency (Per-Task, Per-Interaction), and Interaction Number (1st, 2nd) as main effects. Post-hoc Student's t-tests and Tukey HSD tests were used when appropriate.

**Interaction Disruption.** There was low agreement with the annotations disrupting human-robot interactions (M = 2.6, SE = 0.16). REML analyses resulted in no significant differences; however, there were two trends. The first trend was for Interaction Number ($p = 0.053$). The 1st interaction had an average disruption rating of 2.88 (SE = 0.23) while the 2nd interaction had 2.33 (SE = 0.23). The second trend was for Annotation Frequency having a significant effect on disruption ($p = 0.053$), in line with H1. The average disruption for the Per-Task frequency was 2.88 (SE = 0.26) while for the Per-Interaction frequency was 2.33 (SE = 0.20). We suspect that the lack of significance and the low disruption ratings were because we primed the participants at the beginning of the study about providing self-annotations for robot performance, setting expectations for stopping and resuming the interaction with the robot early on.

**Workload, Memory Burden and Difficulty.** The ratings for mental demand (M = 3.16, SE = 0.19), physical demand (M = 2.28, SE = 0.13), time pressure (M = 1.91, SE = 0.10) and frustration (M = 2.30, SE = 0.13) were generally low, and there was a perception of high level of success (M = 6.04, SE = 0.08) in the self-annotation process. REML analyses showed significant differences only for the interaction between Annotation Order and Frequency on mental demand ($p = 0.03$) and on perceived time pressure ($p = 0.01$). However, post-hoc tests resulted in no significant differences.

As shown in Fig. 5, most participants did not agree that it was hard to remember what happened during the interaction when providing self-annotations (M = 2.61, SE = 0.17) nor with the annotations being difficult (M = 2.39, SE = 0.17). Interestingly, Annotation Frequency led to significant differences on both perceptions. For

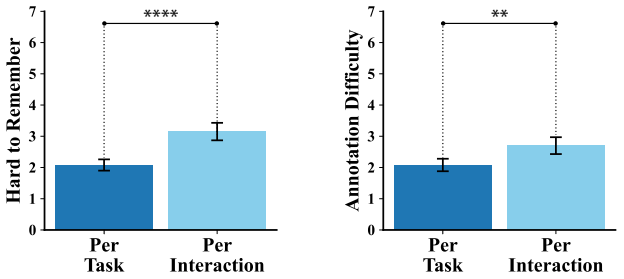



**Figure 5: Perception of how hard it was to remember what happened during the interaction (left) and annotation difficulty (right). (****) and (**) denote $p < 0.0001$ and $p < 0.01$.**

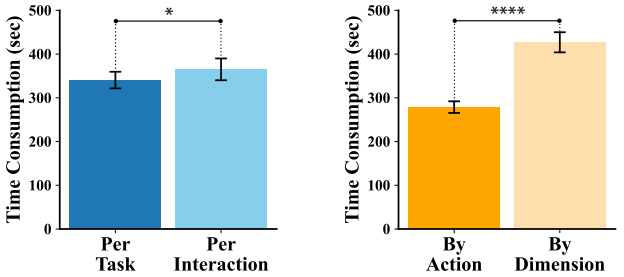



**Figure 6: Amount of time spent on self-annotations by Annotation Frequency (left) and Annotation Order (right). (****) and (*) denote $p < 0.0001$ and $p < 0.05$, respectively.**

how hard it was to remember (p < 0.0001), Per-Task frequency (M = 2.08, SE = 0.18) made it significantly less challenging than Per-Interaction frequency (M = 3.15, SE = 0.28). Similarly, for annotation difficulty (p = 0.0038), Per-Task frequency (M = 2.08, SE = 0.20) led to significantly lower ratings than Per-Interaction frequency (M = 2.70, SE = 0.27). These results provided support for H2.

**Annotation Time.** An REML analysis indicated that Interaction Number (F[1,37] = 71.99, p < 0.0001), Annotation Frequency (F[1,37] = 4.59, p = 0.04), and Annotation Order (F[1,37] = 21.04, p < 0.0001) had a significant effect on how long it took participants to complete the annotations. First, participants spent significantly more time annotating the 1st interaction (M = 401.33, SE = 22.32) than the 2nd interaction (M = 304.35, SE = 19.25). This was expected due to the unfamiliarity of the annotation process earlier in the study. Second, annotating with the Per-Interaction frequency (M = 365.08, SE = 24.91) was significantly more time-consuming than with the Per-Task frequency (M = 340.6, SE = 19.00), possibly because of the extra memory burden and difficulty induced by the Per-Interaction frequency (as in Fig. 5). Lastly, the By-Dimension order (M = 427.00, SE = 23.02) took significantly more time in comparison to the By-Action order (M = 278.68, SE = 13.31). This last finding supports H3. Fig. 6 shows results by Annotation Frequency and Order.

## 4.2 Making Sense of Implicit Feedback

To investigate the usefulness of implicit feedback data for predicting per-action explicit measures of robot performance, we implemented a two-layered, bidirectional Gated Recurrent Unit (GRU) network [17, 63] in PyTorch. The neural network took as input a temporal

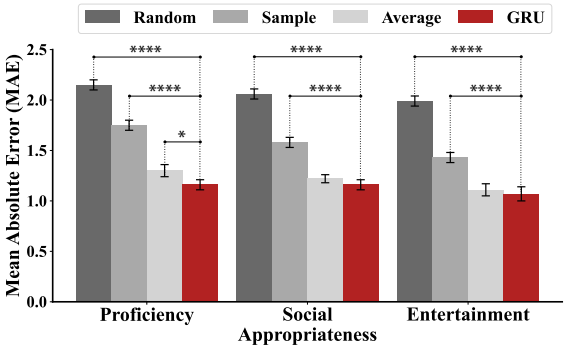



**Figure 7: MAE for predicting per-action robot performance. Baseline models (Random, Sample and Average) led to significantly different results from each other, but we only show significant differences in the plot with our GRU model due to limited space. (****) and (*) denote $p < 0.0001$ and $p < 0.05$.**

series of implicit feedback features (as explained in Sec. 3.5) concatenated with a one-hot encoding of robot action, which served to contextualize the implicit feedback. The model output a per-action performance value in $[1, 7]$ (for a single robot performance dimension). The model was trained in a supervised manner on the Mean Absolute Error (MAE) with respect to per-action annotations provided by the participants during the study. In general, we used a batch size of 256, a learning rate of 1e−5 and the AdamW optimizer [48] to train the model. We employed leave-one-out cross-validation to evaluate accuracy with MAE, where each participant served as the test set once. See the supplementary material for more details.

**Comparison with Guessing.** To validate that implicit nonverbal reactions contain useful information that helps infer explicit feedback labels, we compared our GRU model with 3 baselines: *Random*, which predicted a label of 1-7 uniformly at random; *Sample*, which predicted a value in 1-7 by sampling from the ground truth label distribution of the training set; and *Average*, which always predicted the average of all the ground truth labels in the training set. Fig. 7 shows the MAE results averaged across all 40 leave-one-out folds.

We analyzed the MAE results per performance dimension using REML analyses that considered Model (Random, Sample, Average, GRU) as a main effect and Participant ID as a random effect. For proficiency, the REML analysis indicated that Model had a significant effect on the MAE (F[3,117] = 161.71, p < 0.0001). A Tukey HSD post-hoc test then showed that the GRU model (M = 1.16, SE = 0.05) had a significantly lower error than all the baselines. For the other dimensions, the REML analyses also indicated a significant effect of Model (p < 0.0001). The post-hoc tests revealed that our GRU model was significantly better (lower MAE) than the Random and Sample baselines. Although our GRU model did not result in significantly lower MAE than the Average baseline for the social appropriateness and entertainment value dimensions, our results are in-line with H4, especially when considering photography proficiency.

**Comparison of Annotation Procedures.** To investigate if the annotation procedures led to differences in model accuracy, we built separate datasets for BA-PT, BA-PI, BD-PT, and BD-PI, considering the implicit feedback, one-hot action encodings and per-action annotations for each condition. Then, for each performance dimension,

**Table 3: Average MAE (with Std. Err.) on per-action annotations for our GRU model. Data were split by study condition.**

| Dimension | BA-PT | BA-PI | BD-PT | BD-PI |
|---|---|---|---|---|
| Proficiency | 1.26 (0.10) | 1.14 (0.07) | 1.27 (0.10) | 1.35 (0.10) |
| Social App. | 1.19 (0.09) | 1.15 (0.08) | 1.21 (0.08) | 1.27 (0.08) |
| Entertainment | 1.28 (0.15) | 1.14 (0.11) | 1.02 (0.14) | 0.91 (0.07) |

we trained our GRU model using leave-one-out cross-validation on each subset of the data. Table 3 shows the MAE results for the GRU models per performance dimension. While all four GRU models achieved low errors, an REML analysis revealed no significant differences for these results. Thus, we found no support for H5.

## 5 DISCUSSION

**Findings in relation to RQ1.** Our first research question was about whether self-annotation procedures influence human perception of the flow of interactions and the annotation process. First, we found a trend toward more disruptions to the interaction with the robot when annotating with the Per-Task frequency than with the Per-Interaction frequency. This trend was in line with our first hypothesis (H1). Second, we found that the Per-Task frequency led to significantly lower memory burden and annotation difficulty than the Per-Interaction frequency, partially supporting H2. We suspect that the latter findings were in turn reflected in significantly less annotation time with Per-Task than with Per-Interaction frequency. Also, providing the self-annotations with By-Action order consumed less time than with the By-Dimension order, as expected for H3. Taken together, these findings suggest that care must be taken when designing self-annotation methods because procedural details can influence how hard it is for robot users to provide data and the time spent on annotations. Potentially, the self-annotation method can also result in disruptions to interactions, which could influence human engagement and perceptions of the robot.

**Findings in relation to RQ2.** Our second research question was about the usefulness of the self-annotation data for mapping implicit feedback in the form of nonverbal signals to explicit measures of robot performance. As hypothesized in H4, we found value in predicting explicit feedback labels by interpreting implicit feedback with a machine learning model, especially for the proficiency dimension. While the GRU model achieved significantly lower MAE than the Average model for predicting photography proficiency, the broader takeaway from our results in Fig. 7 is that our annotation methods enabled the strong performance of not only our GRU model but also the Average baseline. Each of these models relied upon the dataset of per-action labels produced by our annotation methods: the Average baseline capitalized on the consistency of these labels across participants and interactions while the GRU further capitalized on the alignment of explicit per-action labels with implicit feedback in order to interpret nonverbal reactions.

Although we found no support for BA-PT leading to better data for implicit feedback models (H5), all the models that were trained separately on data from our four annotation procedures led to small mean absolute errors with respect to ground truth explicit feedback. This suggests that all four methods for gathering self-annotations could aid in building better implicit feedback models in the future.

**Limitations & Future Work.** Our work primarily focused on evaluating methods for annotating robot performance in human-robot interactions, so the scope of implicit data modalities and model architectures is limited and can be expanded in future work. Worth noting, we used the boundaries of high-level robot actions to align implicit reaction features with explicit human labels in a photography setting. However, other robotics tasks like social navigation might be harder to discretize, for which continuous annotation [51, 71] could be valuable to enable fine-grained mapping of human reactions to explicit feedback about robot performance.

Our work explored the relationship between implicit reactions grounded in explicit labels and rewards used in Reinforcement Learning. In particular, our GRU model mapped implicit feedback to explicit feedback, which could be considered instantaneous rewards (as in Sec. 3.7). In the future, we want to investigate how per-action rewards can be leveraged for robot behavior adaptation in HRI. We suspect that this may require incorporating human biases in how we reason about the rewards [64] as well as considering peculiarities of the interaction scenario. For instance, we observed cases in our study where later robot actions (closer to capturing a photo) influenced the robot's performance for the photo-taking task more than earlier actions.

Finally, by simultaneously collecting self-annotations for three dimensions of robot performance, our work sets the stage for future work in multi-objective robot learning. Although we chose performance objectives specifically for robot photography, other robotic tasks may require measuring performance with different or more dimensions. Also, future work could investigate whether personalizing models for mapping nonverbal implicit feedback to explicit feedback improves prediction accuracy. For our robot photographer, we suspect that conditioning our GRU model based on individual user characteristics could lead to better results in the future because perceptions of robot dimensions are inherently subjective [10].

**Recommendations for Implicit Feedback Datasets.** We hope that future work leverages our findings to create high-quality implicit feedback datasets in HRI. To this end, we first recommend increasing the frequency of annotations during human-robot interactions whenever possible to reduce memory burden and the annotation difficulty perceived by users. However, such improvements may lead to more disruptions to interactions, even when annotations are gathered after well-structured robot tasks (as in our study). Therefore, our second recommendation is to conduct annotation pilots to find a balance between how challenging the annotation process is to users and how disruptive annotating is to the interaction. Third, we recommend labeling all robot performance dimensions for small slices of time (e.g., per robot action) and familiarizing users with the annotation process early on during data collection. This can save time, potentially reducing the cost of creating the dataset.

## ACKNOWLEDGEMENTS

Thanks to Ryan Kim for helping implement and test code for this project. This work was partially funded by the National Science Foundation (NSF) under Grants No. IIS-1924802 and IIS-2106690. Any opinions, findings, and conclusions in this paper are those of the authors and do not necessarily reflect the views of the NSF.

# REFERENCES


[1] Timothy Adamson, C. Burton Lyng-Olsen, Kendrick Umstattd, and Marynel Vázquez. 2020. Designing Social Interactions with a Humorous Robot Photographer. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) *(HRI '20)*. Association for Computing Machinery, New York, NY, USA, 233–241. https://doi.org/10.1145/3319502.3374809

[2] Neziha Akalin and Amy Loutfi. 2021. Reinforcement Learning Approaches in Social Robotics. *Sensors* 21, 4 (2021). https://doi.org/10.3390/s21041292

[3] Riku Arakawa, Sosuke Kobayashi, Yuya Unno, Yuta Tsuboi, and Shin-ichi Maeda. 2018. DQN-TAMER: Human-in-the-Loop Reinforcement Learning with Intractable Feedback. https://doi.org/10.48550/ARXIV.1810.11748

[4] Brenna D. Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. 2009. A survey of robot learning from demonstration. *Robotics and Autonomous Systems* 57, 5 (2009), 469–483. https://doi.org/10.1016/j.robot.2008.10.024

[5] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. OpenFace 2.0: Facial Behavior Analysis Toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. 59–66. https://doi.org/10.1109/FG.2018.00019

[6] Santosh Balajee Banisetty, Scott Forer, Logan Yliniemi, Monica Nicolescu, and David Feil-Seifer. 2021. Socially Aware Navigation: A Non-Linear Multi-Objective Optimization Approach. *ACM Trans. Interact. Intell. Syst.* 11, 2, Article 15 (jul 2021), 26 pages. https://doi.org/10.1145/3453445

[7] Christoph Bartneck, Tony Belpaeme, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Šabanović. 2020. *Human-Robot Interaction: An Introduction*. Cambridge University Press. https://doi.org/10.1017/9781108676649

[8] Atef Ben-Youssef, Chloé Clavel, Slim Essid, Miriam Bilac, Marine Chamoux, and Angelica Lim. 2017. UE-HRI: A New Dataset for the Study of User Engagement in Spontaneous Human-Robot Interactions. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction* (Glasgow, UK) *(ICMI '17)*. Association for Computing Machinery, New York, NY, USA, 464–472. https://doi.org/10.1145/3136755.3136814

[9] Joost Broekens. 2007. Emotion and Reinforcement: Affective Facial Expressions Facilitate Robot Learning. In *Artifical Intelligence for Human Computing*, Thomas S. Huang, Anton Nijholt, Maja Pantic, and Alex Pentland (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 113–132.

[10] Kate Candon, Zoe Hsu, Yoony Kim, Jesse Chen, Nathan Tsoi, and Marynel Vázquez. 2023. Nonverbal Human Signals Can Help Autonomous Agents Infer Human Preferences for Their Behavior. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS '23)*. IFAAMAS.

[11] Kate Candon, Helen Zhou, Sarah Gillet, and Marynel Vázquez. 2023. Verbally Soliciting Human Feedback in Continuous Human-Robot Collaboration: Effects of the Framing and Timing of Reminders. In *Proceedings of the 18th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

[12] Ginevra Castellano, Hatice Gunes, Christopher Peters, and Björn W. Schuller. 2015. Multimodal Affect Recognition for Naturalistic Human-Computer and Human-Robot Interactions. In *The Oxford Handbook of Affective Computing*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199942237.013.026 arXiv:https://academic.oup.com/book/0/chapter/212016009/chapter-ag-pdf/44596643/book_28057_section_212016009.ag.pdf

[13] Ginevra Castellano, Iolanda Leite, André Pereira, Carlos Martinho, Ana Paiva, and Peter W Mcowan. 2013. Multimodal Affect Modeling and Recognition for Empathic Robot Companions. *International Journal of Humanoid Robotics* 10, 01 (2013), 1350010. https://doi.org/10.1142/S0219843613500102 arXiv:https://doi.org/10.1142/S0219843613500102

[14] Oya Celiktutan, Evangelos Sariyanidi, and Hatice Gunes. 2018. Computational Analysis of Affect, Personality, and Engagement in Human-Robot Interactions. In *Computer Vision for Assistive Healthcare*, Marco Leo and Giovanni Maria Farinella (Eds.). Academic Press, 283–318. https://doi.org/10.1016/B978-0-12-813445-0.00010-1

[15] Oya Celiktutan, Efstratios Skordos, and Hatice Gunes. 2019. Multimodal Human-Human-Robot Interactions (MHHRI) Dataset for Studying Personality and Engagement. *IEEE Transactions on Affective Computing* 10, 4 (2019), 484–497. https://doi.org/10.1109/TAFFC.2017.2737019

[16] Sonia Chernova and Andrea L. Thomaz. 2014. *Robot Learning from Human Teachers*. Morgan & Claypool Publishers.

[17] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*. Association for Computational Linguistics, Doha, Qatar, 103–111. https://doi.org/10.3115/v1/W14-4012

[18] Sammy Christen, Stefan Stevšić, and Otmar Hilliges. 2019. Demonstration-Guided Deep Reinforcement Learning of Control Policies for Dexterous Human-Robot Interaction. In *2019 International Conference on Robotics and Automation (ICRA)*. 2161–2167. https://doi.org/10.1109/ICRA.2019.8794065

[19] Joe Connolly, Viola Mocz, Nicole Salomons, Joseph Valdez, Nathan Tsoi, Brian Scassellati, and Marynel Vázquez. 2020. Prompting Prosocial Human Interventions in Response to Robot Mistreatment. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) *(HRI '20)*. Association for Computing Machinery, New York, NY, USA, 211–220. https://doi.org/10.1145/3319502.3374781

[20] Y Cui, P Koppol, H Admoni, S Niekum, R Simmons, A Steinfeld, and T Fitzgerald. 2021. Understanding the Relationship between Interactions and Outcomes in Human-in-the-Loop Machine Learning. In *International Joint Conference on Artificial Intelligence*.

[21] Yuchen Cui, Qiping Zhang, Brad Knox, Alessandro Allievi, Peter Stone, and Scott Niekum. 2021. The EMPATHIC Framework for Task Learning from Implicit Human Feedback. In *Proceedings of the 2020 Conference on Robot Learning (Proceedings of Machine Learning Research, Vol. 155)*, Jens Kober, Fabio Ramos, and Claire Tomlin (Eds.). PMLR, 604–626. https://proceedings.mlr.press/v155/cui21a.html

[22] Matthew Dailey, Carrie Joyce, Michael Lyons, Miyuki Kamachi, Hanae Ishi, Jiro Gyoba, and Garrison Cottrell. 2010. Evidence and a Computational Explanation of Cultural Differences in Facial Expression Recognition. *Emotion (Washington, D.C.)* 10 (12 2010), 874–93. https://doi.org/10.1037/a0020019

[23] Mary Ellen Foster, Andre Gaschler, and Manuel Giuliani. 2017. Automatically classifying user engagement for dynamic multi-party human–robot interaction. *International Journal of Social Robotics* 9, 5 (2017), 659–674.

[24] Goren Gordon, Samuel Spaulding, Jacqueline Kory Westlund, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal. 2016. Affective Personalization of a Social Robot Tutor for Children's Second Language Skills. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* (Phoenix, Arizona) *(AAAI'16)*. AAAI Press, 3951–3957.

[25] Shane Griffith, Kaushik Subramanian, Jonathan Scholz, Charles L Isbell, and Andrea L Thomaz. 2013. Policy Shaping: Integrating Human Feedback with Reinforcement Learning. In *Advances in Neural Information Processing Systems*, C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger (Eds.), Vol. 26. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2013/file/e034fb6b66aacc1d48f445ddfb08da98-Paper.pdf

[26] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[27] David A. Harville. 1977. Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *J. Amer. Statist. Assoc.* 72, 358 (1977), 320–338. http://www.jstor.org/stable/2286796

[28] Ran R Hassin, Hillel Aviezer, and Shlomo Bentin. 2013. Inherently ambiguous: Facial expressions of emotions, in context. *Emotion Review* 5, 1 (2013), 60–65.

[29] William E Hockley. 1984. Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10, 4 (1984), 598.

[30] Nancy Ide and James Pustejovsky. 2017. *Handbook of Linguistic Annotation* (1st ed.). Springer Publishing Company, Incorporated.

[31] Charles Isbell, Christian R. Shelton, Michael Kearns, Satinder Singh, and Peter Stone. 2001. A Social Reinforcement Learning Agent. In *Proceedings of the Fifth International Conference on Autonomous Agents* (Montreal, Quebec, Canada) *(AGENTS '01)*. Association for Computing Machinery, New York, NY, USA, 377–384. https://doi.org/10.1145/375735.376334

[32] Rachael E Jack and Philippe G Schyns. 2015. The human face as a dynamic tool for social communication. *Current Biology* 25, 14 (2015), R621–R634.

[33] Dinesh Babu Jayagopi, Samira Sheikhi, David Klotz, Johannes Wienke, Jean-Marc Odobez, Sebastian Wrede, Vasil Khalidov, Laurent Nguyen, Britta Wrede, and Daniel Gatica-Perez. 2012. The Vernissage Corpus: A Multimodal Human-Robot-Interaction Dataset. (2012), 8. http://infoscience.epfl.ch/record/182715

[34] Hong Jun Jeon, Smitha Milli, and Anca Dragan. 2020. Reward-rational (implicit) choice: A unifying formalism for reward learning. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 4415–4426. https://proceedings.neurips.cc/paper/2020/file/2f10c1578a0706e06b6d7db6f0b4a6af-Paper.pdf

[35] Malte F Jung. 2016. Coupling interactions and performance: Predicting team performance from thin slices of conflict. *ACM Transactions on Computer-Human Interaction (TOCHI)* 23, 3 (2016), 1–32.

[36] Adam Kendon. 1988. Goffman's approach to face-to-face interaction. *Erving Goffman: Exploring the interaction order* (1988).

[37] Ross A. Knepper, Christoforos I. Mavrogiannis, Julia Proft, and Claire Liang. 2017. Implicit Communication in a Joint Action. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna, Austria) *(HRI '17)*. Association for Computing Machinery, New York, NY, USA, 283–292. https://doi.org/10.1145/2909824.3020226

[38] W Bradley Knox, Brian D Glass, Bradley C Love, W Todd Maddox, and Peter Stone. 2012. How humans teach agents. *International Journal of Social Robotics* 4, 4 (2012), 409–421.

[39] W. Bradley Knox and Peter Stone. 2009. Interactively Shaping Agents via Human Reinforcement: The TAMER Framework. In *Proceedings of the Fifth International Conference on Knowledge Capture* (Redondo Beach, California, USA) *(K-CAP '09)*. Association for Computing Machinery, New York, NY, USA, 9–16. https:

//doi.org/10.1145/1597735.1597738
[40] W Bradley Knox, Peter Stone, and Cynthia Breazeal. 2013. Training a robot via human feedback: A case study. In *International Conference on Social Robotics*. Springer, 460–470.
[41] Kazunori Komatani and Shogo Okada. 2021. Multimodal human-agent dialogue corpus with annotations at utterance and dialogue levels. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–8.
[42] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research methods in human-computer interaction*. Morgan Kaufmann.
[43] Iolanda Leite, Marissa McCoy, Daniel Ullman, Nicole Salomons, and Brian Scassellati. 2015. Comparing Models of Disengagement in Individual and Group Interactions. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, Oregon, USA) *(HRI '15)*. Association for Computing Machinery, New York, NY, USA, 99–105. https://doi.org/10.1145/2696454.2696466
[44] Guangliang Li, Hamdi Dibeklioğlu, Shimon Whiteson, and Hayley Hung. 2020. Facial feedback for reinforcement learning: a case study and offline analysis using the TAMER framework. *Autonomous Agents and Multi-Agent Systems* 34, 1 (2020), 1–29.
[45] Guangliang Li, Hayley Hung, Shimon Whiteson, and W. Bradley Knox. 2013. Using Informative Behavior to Increase Engagement in the Tamer Framework. In *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems* (St. Paul, MN, USA) *(AAMAS '13)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 909–916.
[46] Jinying Lin, Zhen Ma, Randy Gomez, Keisuke Nakamura, Bo He, and Guangliang Li. 2020. A Review on Interactive Reinforcement Learning From Human Social Feedback. *IEEE Access* 8 (2020), 120757–120765.
[47] Robert Loftin, Bei Peng, James Macglashan, Michael L. Littman, Matthew E. Taylor, Jeff Huang, and David L. Roberts. 2016. Learning Behaviors via Human-Delivered Discrete Feedback: Modeling Implicit Feedback Strategies to Speed up Learning. *Autonomous Agents and Multi-Agent Systems* 30, 1 (jan 2016), 30–59. https://doi.org/10.1007/s10458-015-9283-7
[48] Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*. https://openreview.net/forum?id=Bkg6RiCqY7
[49] James MacGlashan, Mark K Ho, Robert Loftin, Bei Peng, Guan Wang, David L Roberts, Matthew E Taylor, and Michael L Littman. 2017. Interactive learning from policy-dependent human feedback. In *International Conference on Machine Learning*. PMLR, 2285–2294.
[50] Emily McQuillin, Nikhil Churamani, and Hatice Gunes. 2022. Learning Socially Appropriate Robo-waiter Behaviours through Real-time User Feedback. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 541–550. https://doi.org/10.1109/HRI53351.2022.9889395
[51] David Melhart, Antonios Liapis, and Georgios N Yannakakis. 2019. PAGAN: Video affect annotation made easy. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 130–136.
[52] Adam Norton, Henny Admoni, Jacob Crandall, Tesca Fitzgerald, Alvika Gautam, Michael Goodrich, Amy Saretsky, Matthias Scheutz, Reid Simmons, Aaron Steinfeld, et al. 2022. Metrics for Robot Proficiency Self-assessment and Communication of Proficiency in Human-robot Teams. *ACM Transactions on Human-Robot Interaction (THRI)* 11, 3 (2022), 1–38.
[53] Catharine Oertel and Giampiero Salvi. 2013. A Gaze-Based Method for Relating Group Involvement to Individual Engagement in Multimodal Multiparty Dialogue. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (Sydney, Australia) *(ICMI '13)*. Association for Computing Machinery, New York, NY, USA, 99–106. https://doi.org/10.1145/2522848.2522865
[54] Damilola Omitaomu, Shabnam Tafreshi, Tingting Liu, Sven Buechel, Chris Callison-Burch, Johannes Eichstaedt, Lyle Ungar, and João Sedoc. 2022. Empathic Conversations: A Multi-level Dataset of Contextualized Conversations. *arXiv preprint arXiv:2205.12698* (2022).
[55] Dim P. Papadopoulos, Jasper R. R. Uijlings, Frank Keller, and Vittorio Ferrari. 2017. Extreme Clicking for Efficient Object Annotation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.
[56] Patrick M. Pilarski, Michael R. Dawson, Thomas Degris, Farbod Fahimi, Jason P. Carey, and Richard S. Sutton. 2011. Online human training of a myoelectric prosthesis controller via actor-critic reinforcement learning. In *2011 IEEE International Conference on Rehabilitation Robotics*. 1–7. https://doi.org/10.1109/ICORR.2011.5975338
[57] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Ng. 2009. ROS: an open-source Robot Operating System. *ICRA Workshop on Open Source Software* 3.
[58] Anne Richards, Christopher C French, Andrew J Calder, Ben Webb, Rachel Fox, and Andrew W Young. 2002. Anxiety-related bias in the classification of emotionally ambiguous facial expressions. *Emotion* 2, 3 (2002), 273.
[59] Katharina Rohlfing, Daniel Loehr, Susan Duncan, Amanda Brown, Amy Franklin, Irene Kimbara, Jan-Torsten Milde, Fey Parrill, Travis Rose, Thomas Schmidt, et al. 2006. Comparison of multimodal annotation tools-workshop report. *Gesprächforschung-Online-Zeitschrift zur Verbalen Interaktion* 7 (2006), 99–123.
[60] Dorsa Sadigh, Anca D. Dragan, S. Shankar Sastry, and Sanjit A. Seshia. 2017. Active Preference-Based Learning of Reward Functions. In *Robotics: Science and Systems*.
[61] Dorsa Sadigh, Shankar Sastry, Sanjit A. Seshia, and Anca D. Dragan. 2016. Planning for Autonomous Cars that Leverage Effects on Human Actions. In *Proceedings of Robotics: Science and Systems*. AnnArbor, Michigan. https://doi.org/10.15607/RSS.2016.XII.029
[62] Hanan Salam, Oya Celiktutan, Isabelle Hupont, Hatice Gunes, and Mohamed Chetouani. 2016. Fully automatic analysis of engagement and its relationship to personality in human-robot interactions. *IEEE Access* 5 (2016), 705–721.
[63] M. Schuster and K.K. Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing* 45, 11 (1997), 2673–2681. https://doi.org/10.1109/78.650093
[64] Rohin Shah, Noah Gundotra, Pieter Abbeel, and Anca Dragan. 2019. On the Feasibility of Learning, Rather than Assuming, Human Biases for Reward Inference. In *Proceedings of the 36th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 97)*, Kamalika Chaudhuri and Ruslan Salakhutdinov (Eds.). PMLR, 5670–5679. https://proceedings.mlr.press/v97/shah19a.html
[65] Maia Stiber, Russell Taylor, and Chien-Ming Huang. 2022. Modeling Human Response to Robot Errors for Timely Error Detection. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 676–683. https://doi.org/10.1109/IROS47612.2022.9981726
[66] Walter W Stroup. 2012. *Generalized linear mixed models: modern concepts, methods and applications*. CRC press.
[67] Halit Bener Suay and Sonia Chernova. 2011. Effect of human guidance and state space size on Interactive Reinforcement Learning. In *2011 RO-MAN*. 1–6. https://doi.org/10.1109/ROMAN.2011.6005223
[68] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.
[69] Andrea L. Thomaz and Cynthia Breazeal. 2008. Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artificial Intelligence* 172, 6 (2008), 716–737. https://doi.org/10.1016/j.artint.2007.09.009
[70] Nathan Tsoi, Mohamed Hussein, Olivia Fugikawa, J. D. Zhao, and Marynel Vázquez. 2021. An Approach to Deploy Interactive Robotic Simulators on the Web for HRI Experiments: Results in Social Robot Navigation. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 7528–7535. https://doi.org/10.1109/IROS51168.2021.9636319
[71] Jose Vargas Quiros, Stephanie Tan, Chirag Raman, Laura Cabrera-Quiros, and Hayley Hung. 2022. Covfee: an extensible web framework for continuous-time annotation of human behavior. In *Understanding Social Behavior in Dyadic and Small Group Interactions (Proceedings of Machine Learning Research, Vol. 173)*, Cristina Palmero, Julio C. S. Jacques Junior, Albert Clapés, Isabelle Guyon, Wei-Wei Tu, Thomas B. Moeslund, and Sergio Escalera (Eds.). PMLR, 265–293. https://proceedings.mlr.press/v173/vargas-quiros22a.html
[72] Vivek Veeriah, Patrick M Pilarski, and Richard S Sutton. 2016. Face valuing: Training user interfaces with facial expressions and reinforcement learning. *arXiv preprint arXiv:1606.02807* (2016).
[73] Marynel Vázquez, Aaron Steinfeld, Scott E. Hudson, and Jodi Forlizzi. 2014. Spatial and Other Social Engagement Cues in a Child-Robot Interaction: Effects of a Sidekick. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 391–398.
[74] Erik Wallen, Jan L Plass, and Roland Brünken. 2005. The function of annotations in the comprehension of scientific texts: Cognitive load effects and the impact of verbal ability. *Educational Technology Research and Development* 53, 3 (2005), 59–71.
[75] Klaus Weber, Hannes Ritschel, Ilhan Aslan, Florian Lingenfelser, and Elisabeth André. 2018. How to Shape the Humor of a Robot - Social Behavior Adaptation Based on Reinforcement Learning. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction* (Boulder, CO, USA) *(ICMI '18)*. Association for Computing Machinery, New York, NY, USA, 154–162. https://doi.org/10.1145/3242969.3242976
[76] Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: a Professional Framework for Multimodality Research. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. European Language Resources Association (ELRA), Genoa, Italy. http://www.lrec-conf.org/proceedings/lrec2006/pdf/153_pdf.pdf
[77] Dean Zadok, Daniel McDuff, and Ashish Kapoor. 2021. Modeling Affect-based Intrinsic Rewards for Exploration and Learning. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. 13078–13084. https://doi.org/10.1109/ICRA48506.2021.9562098

# Supplementary Material

This supplementary material provides more details about our study and the models created for mapping implicit human feedback to specific robot performance measures.

## A ADDITIONAL DETAILS ABOUT OUR STUDY

### A.1 Motivation for Study Conditions

For annotation order, the By-Dimension (BD) and By-Action (BA) conditions were motivated by the multi-dimensionality of robot performance. We expected the BD condition to take longer time than BA (H3). However, it was also possible for BA to take longer due to task switching costs from considering different performance dimensions, and these switches happened more often with BA.

For annotation frequency, the Per-Task and Per-Interaction conditions were motivated by the belief that the time delay between the interaction and the annotations would influence the annotation process and resulting data. Less delay would potentially reduce memory burden and annotation difficulty (H2) and facilitate mapping nonverbal behavior to robot performance (H5), but cause more disruption to the interaction (H1).

### A.2 Demonstration of ELAN

Before the participants started their interactions with the robot photographer, they were shown how to use the annotation tool to label the robot's performance. The experimenter opened ELAN on a sample video, explained the interface and annotation format, and demonstrated annotating a few robot actions. Participants were then asked to operate ELAN by themselves to make sure they fully understood the procedures and could annotate smoothly. Once they had no further questions, the annotation sessions officially started for the participant.

### A.3 Annotation Distributions

The distributions of per-action annotations in every dimension of robot performance widely spanned the [1,7] range. In particular, the distributions of annotations per action were as follows: proficiency had M = 4.50 and SD = 1.58; social appropriateness resulted in M = 4.79 and SD = 1.44; and entertainment value led to M = 4.60 and SD = 1.38. As discussed in Sec. 3.7, we computed the undiscounted sum of per-action annotations $G = \sum_{t=0}^{T} r_t$ and checked their correlation with the per-photo explicit feedback from survey ratings. Fig. 8 shows the distributions of per-photo annotations (based on survey ratings) and the sum of per-action annotations.

### A.4 Qualitative Responses & Other Findings

We observed several responses related to our hypotheses in the open-ended survey given at the end of user study, despite their sparseness. In particular, 3 out of 40 participants directly mentioned disruption in open-ended feedback, and all three agreed that the Per-Task frequency were more disruptive than Per-Interaction. Also, 2 out of 40 participants explicitly commented about memory, and both agreed that the Per-Interaction frequency made it harder to remember the robot's performance than Per-Task. Lastly, 5 out of 40 participants explicitly mentioned their experience with the assigned annotation order, and all of them felt that it would be more natural to annotate By-Action than By-Dimension.

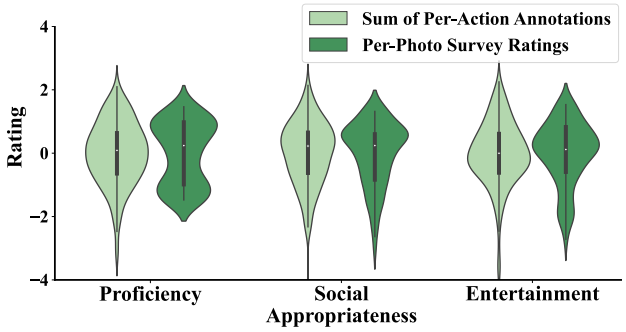



**Figure 8: Distribution of undiscounted sum of rewards $G$ (based on per-action self-annotations) and per-photo survey ratings. Each distribution was whitened to have $\mu = 0, \sigma = 1$.**

While using ELAN, participants seemed more focused on the robot than their own video, sometimes laughing again at the robot's jokes and mistakes. Therefore, analyzing participants' reactions when reviewing their own interaction with robots can be an interesting direction for future investigation.

## B MODEL DESIGN & TRAINING

### B.1 Data Partition for Leave-One-Out Cross-Validation

We used leave-one-out cross-validation to evaluate model generalization to new users. Each data fold was partitioned into training-validation-test sets. We used early stopping (based on the lowest MAE on the validation) to choose model hyperparameters for the final evaluation across the test tests.

To compare our GRU model with the three baselines described in Sec. 4.2 (H4), we built 40 data folds from the entire dataset. Each fold included the data for 35 participants in the training set, the data for 4 participants in the validation set, and the data for 1 participant in the test set. The data for one participant corresponded to all the implicit nonverbal reactions and corresponding per-action explicit performance labels that were collected in the two interactions that the person experienced with the robot (under the Per-Task and Per-Interaction annotation frequencies).

To compare the GRU models across the four annotation procedures, we first divided the entire dataset into 4 subsets according to the annotation procedure the data was collected with. Then, we built 20 data folds from each subset. Each fold included 17 participants in the training set, 2 participants in the validation set, and 1 participant in the test set.

### B.2 Feature Extraction with OpenFace

The outputs we selected from OpenFace 2.0 [5] to model implicit human feedback are: [AU01_r, AU02_r, AU04_r, AU05_r, AU06_r, AU07_r, AU09_r, AU10_r, AU12_r, AU14_r, AU15_r, AU17_r, AU20_r, AU23_r, AU25_r, AU26_r, AU45_r, gaze_angle_x, gaze_angle_y, pose_Tx, pose_Ty, pose_Tz, pose_Rx, pose_Ry, pose_Rz].

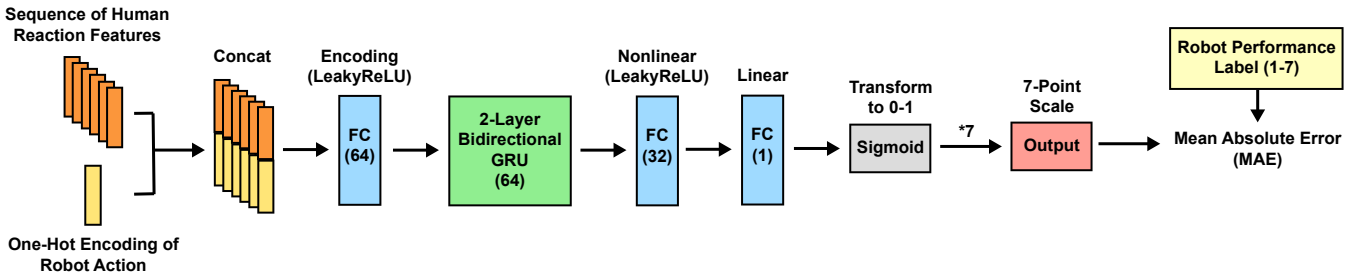



**Figure 9: The architecture of our GRU model for learning from implicit human feedback. After concatenating the series of implicit reaction features with the one-hot encoding of robot action, the input sequence was encoded by a dense non-linear layer activated with LeakyReLU for every time-step. The encoded features were then passed into a 2-layer bidirectional GRU network, from which the last hidden state was fed into another nonlinear dense layer with LeakyReLU activation and finally passed into a linear layer outputting a single value. This value was transformed by a sigmoid function and then multiplied by 7 to compute the Mean Absolute Error (MAE) against the ground-truth robot performance label.**

The AUx_r signals are from regression model to capture the intensity (in 0–5) of activation of facial action units. The gaze_angle signals are eye gaze direction averaged for both eyes. The pose_T and pose_R signals are detected head translation and rotation with respect to the camera. These nonverbal features were lower level and more perceptual than affective features because we did not want to constrain our robot performance predictions to a specific set of emotion categories or valence. Rather, we considered a wider range of features that describe human nonverbal reactions to better alleviate the ambiguity of implicit human feedback.

## B.3 Model Architecture & Training Details

Fig. 9 illustrates our model architecture. For training our GRU model, each example corresponded to a series of human reaction features and high-level robot action label (inputs) along with a corresponding performance label in 1–7 (output) from the self-annotations. We captured the raw human reaction videos during the human-robot interactions, and for each image frame we extracted the implicit features including facial action units (FAUs), head poses, and gazes, as discussed in Sec. 3.5 of the paper. We then aggregated the predicted features using max and min pooling in each feature dimension within mini-time windows of 5 consecutive image frames (corresponding to 1/6 of a second of reaction data). This helped reduce the length of the series of features input to the model, facilitating learning in practice. The temporal sequence of implicit feedback features was then concatenated with the one-hot encoding of robot action to contextualize the implicit feedback.

Depending on how long the individual robot actions were during the interaction, our input feature sequences ended up having a length ranging from 5 to 64 time-steps. For each time-step, our GRU model first encoded the input features with a dense non-linear layer with 64 hidden units and LeakyReLU activation function. The encoded features were then passed into a 2-layer bidirectional GRU network with a hidden vector of size 64. The last hidden state from the GRU layers was concatenated, fed into another nonlinear dense layer with LeakyReLU activation and, finally, passed into a linear layer that output a single value. This value was transformed by a sigmoid function and then multiplied by 7 to constraint the output of the network to a positive value less than or equal to 7.

The model was trained using the mean absolute error (MAE) against the ground-truth labels as a loss. To update model parameters, we used the AdamW optimizer, with a learning rate of 1e-5 and a batch size of 256. We used a dropout rate of 0.2 to the GRU, and L2 regularization with a weight decay of 0.1. All the results were obtained with an Intel Core i7 10700K 8-Core 3.6GHz desktop computer that had an NVIDIA 24GB GeForce RTX 3090 GPU.