Instruct or Evaluate: How People Choose to Teach Norms to Social Robots

Vivienne Bihe Chi  
Department of Cognitive, Linguistic, and Psychological Sciences  
Brown University  
Providence, RI, USA  
vivienne_ch@brown.edu

Bertram F. Malle  
Department of Cognitive, Linguistic, and Psychological Sciences  
Brown University  
Providence, RI, USA  
bfmalle@brown.edu

Abstract—Robots deployed in social settings must act appropriately—that is, in compliance with social and moral norms. However, efforts of teaching norms to robots have typically relied on single teaching methods (e.g., instruction, reward). By contrast, humans may naturally use more than one teaching method when training a novice. To test this claim in the domain of human-robot teaching, we present a novel paradigm in which participants interactively teach a simulated robot to behave appropriately in a healthcare setting, choosing to either instruct the robot or evaluate its proposed actions. We demonstrate that 89% of human teachers naturally adopt mixed teaching strategies. We further identify some of the factors that influence people’s choices. Results reveal that human teachers dynamically update their impression of the robot from early to late in the teaching session, and they choose their teaching strategy based on the robot’s specific actions and their accumulated perceptions of the robot’s learning progress.

Index Terms—social robots, human-robot teaching

I. INTRODUCTION

Social robots act in social contexts. Such contexts are defined and structured by norms—directives for how one should and should not behave. Robots increasingly take on roles of assistant and companion, so they must learn a large number of norms and the specific contexts in which those norms apply [1, 2, 3]. Due to the complex and continuous nature of natural environments, purely unsupervised learning of norms by observation or by trial and error appears unwise. Naturally, humans would make ideal teachers for robot learners [4, 5]. However, a more active, mixed-method interactive approach is needed that brings human teachers fully into the loop, saves time, and ensures safe training [6, 7].

Current training frameworks in HRI largely rely on a single teaching method [8] and place the burden on the human teacher to consistently produce high-quality feedback [9, 10, 11]. Yet, these practices may fail in real world settings. Most agents are not equipped to do one-shot learning, so the robot will repeatedly ask for human feedback on the same task. Despite some initial effort to adjust to inattentive or inaccurate human teachers [12] and to accommodate learning reward function from more than one type of information source [13, 14], most extant work takes the learner’s perspective without shedding light on what forms of teaching data are expected under which circumstances. In this study, we aim to fill this gap by studying human’s natural teaching behavior, which can then inform the design of an improved robot learning framework.

In the pedagogy literature, teaching strategies can be broadly divided into teacher-led (e.g., demonstration, instruction) and student-led (e.g., exploration with feedback) [15]. Teachers often choose between these teaching strategies deliberately as they monitor the learner’s learning progress and accordingly adjust the amount of direction they give [16, 17, 18]. Likewise, people who teach a robot may vary their teaching strategies. For social tasks that the robot is not familiar with, they are likely to give the robot specific action instructions [19]. However, providing instructions for an extended period of time demands significant effort from the teacher. Under some conditions, the teacher may let the learner attempt to perform the task first and follow up with evaluative feedback (e.g. “That’s good!”) [20]. The responsibility then falls on the learner to incorporate the feedback and improve their future performance.

In this study we examine an initial set of possible determinants of a teacher’s choice to either give direct instructions to a robot or let the robot propose an action and evaluate this proposal. We present an interactive teaching paradigm in which humans train a robot over the course of 18 trials, each time deciding whether to instruct or evaluate. We examine the impact of experimentally manipulated expectations about the robot, the robot’s specific actions, and their accumulated perceptions of the robot’s learning progress.

II. METHODS

A. Platform

We designed a platform that allowed participants to engage in an interactive training session with a robot agent on a cellphone device (Fig. 1). The top part of the screen mimics a messaging application (“chat app”) where robot and human teacher communicate via short messages. The area below the chat app is reserved for prompts from researchers (“Your response to the robot assistant?”) and response options for participants (e.g., teaching strategy choices, evaluation ratings).
**B. Participants**

We recruited 212 participants on Prolific (http://prolific.co), slightly oversampling beyond the pre-registered sample size of $n = 100$ for each between-subjects condition (targeting statistical power of $(1 - \beta) > .80$ at $\alpha < .05$ for an effect size of $d > .4$). Six participants were excluded following pre-registered criteria, leaving $n = 206$ for analysis. The pre-registration can be found at [https://osf.io/qwk86](https://osf.io/qwk86).

**C. Procedure**

Participants were invited to act in the role of a busy nurse responsible for training a robot assistant that is ready to learn and take over some tasks. Participants were told that the robot would consult with them in situations where the best action was unclear. As in real-life scenarios, the human teacher needed to provide sufficient direction for norm-compliant actions but also let the robot come up with action plans and evaluate them. Teachers had to choose their own preferred balance between direct instruction and evaluative feedback.

Participants were initially provided with two pieces of information about their robot assistant, which implemented the between-subjects experimental manipulation of Robot experience. In the experienced condition, the robot was described as having undergone some amount of training as a hospital assistant. Some initial observations indicated that it knew to close the door behind it and to comfort a patient in pain. In the inexperienced condition, the robot was described as having just started to undergo training as a hospital assistant. Some initial observations indicated that it knew to beeping, but in the second instance it proposed an appropriate action of moving slowly through the crowded room.

The main experiment (robot training session) consisted of 18 teaching trials (3 task blocks with 6 subtasks each). To design the teaching trials, we selected scenarios relevant to a nurse assistant’s daily responsibilities, such as measuring blood oxygen levels and delivering medicine to patients. Each scenario implicated one or more social-moral norms, and we prepared three possible actions per scenario: an inappropriate, acceptable, or appropriate behavior relative to the scenario’s norms. For example, for a scenario of seeing a patient’s family member smoke in the hospital room, the unacceptable action was not to intervene, the acceptable action was to make some noise, and the appropriate action was to announce a gentle reminder.

The robot’s learning progress over the course of the training session was fixed by having the robot propose an improving rate of appropriate behaviors over the three task blocks. Specifically, in block 1, the distribution of inappropriate-acceptable-appropriate behaviors was 2-4-0, in block 2 it was 1-3-2, and in block 3 it was 0-2-4.

In addition, some of the trials reintroduced a previous subtask so as to demonstrate even more clearly the robot’s learning progress. For example, in a scenario of moving across a crowded waiting room, in the first instance the robot proposed an acceptable action of moving quickly while beeping, but in the second instance it proposed an appropriate action of moving slowly through the crowded room.

Participants evaluated the robot’s action proposals on a five-point rating scale marked by the labels “That’s forbidden,” “… discouraged”, “… allowed”, “… appropriate,” and “… required.”

**D. Materials and Measures**

The experiment consisted of 18 teaching trials (3 task blocks with 6 subtasks each). To design the teaching trials, we selected scenarios relevant to a nurse assistant’s daily responsibilities, such as measuring blood oxygen levels and delivering medicine to patients. Each scenario implicated one or more social-moral norms, and we prepared three possible actions per scenario: an inappropriate, acceptable, or appropriate behavior relative to the scenario’s norms. For example, for a scenario of seeing a patient’s family member smoke in the hospital room, the unacceptable action was not to intervene, the acceptable action was to make some noise, and the appropriate action was to announce a gentle reminder.

The robot’s learning progress over the course of the training session was fixed by having the robot propose an improving rate of appropriate behaviors over the three task blocks. Specifically, in block 1, the distribution of inappropriate-acceptable-appropriate behaviors was 2-4-0, in block 2 it was 1-3-2, and in block 3 it was 0-2-4.

In addition, some of the trials reintroduced a previous subtask so as to demonstrate even more clearly the robot’s learning progress. For example, in a scenario of moving across a crowded waiting room, in the first instance the robot proposed an acceptable action of moving quickly while beeping, but in the second instance it proposed an appropriate action of moving slowly through the crowded room.

Participants evaluated the robot’s action proposals on a five-point rating scale marked by the labels “That’s forbidden,” “… discouraged”, “… allowed”, “… appropriate,” and “… required.”

**E. Design and Hypotheses**

We experimentally manipulated one potential determinant of teaching strategy choices: the initial expectation of the robot’s level of experience. We randomly assigned participants to either the experienced robot, which had prior training in the current domain, or the inexperienced robot, which had just started training (see II.C above for more detail). Both types of robots, however, were described to be “fully capable of handling the daily tasks at the hospital.”
We hypothesized that (1) the robot’s experience level would influence people’s teaching strategy choices, such that the less experienced robot would encourage a higher rate of instruct choices. We further hypothesized that (2) people will be sensitive to the robot’s displays of norm competence, such that its appropriate action proposals would be evaluated more positively, and its inappropriate proposals more negatively, than acceptable proposals.

III. RESULTS

A. Predicting People’s Teaching Strategy

Participants were free to use either one of the two teaching strategies (instruct vs. evaluate) in each of 18 trials. Of the 206 participants, 89% adopted a mixed teaching strategy, 10% chose to always give action instructions, and 1% chose to always evaluate the robot’s proposals. Given that most people used a mixed teaching strategy, we aimed to predict their strategy choice at each step of the training process.

To begin, a person’s strategy choice for the very first subtask should be influenced by the robot’s level of experience, as manipulated in the instruction phase of the experiment. An initial generalized linear model showed that, by itself, this manipulation was weak and nonsignificant (\( p = .21 \)). We then introduced each individual’s overall teaching strategy (number of instruct choices averaged over the remaining 17 subtasks) as another factor, along with the interaction term. In this model, overall teaching strategy strongly predicted first-task choice, \( z = 3.54, p < .001 \); further, participants introduced to an experienced robot had a lower probability of giving action instructions in the first subtask (\( M = 0.45 \)) than those introduced to an inexperienced robot (\( M = 0.56 \)), \( z = 2.65, p = .008 \); finally, the two predictors’ interaction (\( z = 2.32, p = .02 \)) indicated that those who were training an experienced robot made an initial teaching choice that was very similar to their general strategy, whereas those training an inexperienced robot made an initial teaching choice that was largely unrelated to their teaching strategy in the rest of the task (Fig. 2).

We then reversed the analysis to predict people’s overall teaching strategy from the first-subtask teaching choice and robot experience level. Experience level did not predict a person’s overall teaching strategy, \( z = 0.7, p = 0.48 \), whereas first-subtask choice strongly predicted overall strategy, \( t = 4.0, p < 0.001 \). Thus, even though the first teaching choice is guided by people’s initial expectation of the robot’s experience (see Fig. 2), as people continue to engage with the robot they adopt a personally preferred teaching strategy, and the robot’s initial level of experience no longer affects their choices. In fact, at the end of the experiment, only 24.7% of participants accurately recalled whether the robot had initially been introduced as “experienced” or “inexperienced.”

Next, we examined which factors predict people’s strategy choices to either instruct or evaluate in each of 18 subtasks. We found, first, an overall increase in the tendency to offer instruction. Relative to first task block (\( M = 0.53 \)), probability of instruction increased in the second block (\( M = 0.71 \)), \( z = 6.4, p < .001 \), and also from the second block to the third block (\( M = 0.80 \)), \( z = 8.72, p < .001 \). This overall main effect is ambiguous—it might indicate a loss of trust in the robot over the course of training, or a desire to perfect the improving robot.

More meaningful is the impact of seeing a previous subtask reoccur for a second or third time. Here, probability of instruction decreases from the initial instance (\( M = 0.53 \)) to the reoccurring instance (\( M = 0.43 \)), \( z = -3.8, p = 0.001 \). This decline of instruction suggests that people expected the robot to have learned the behavior the first time around and therefore may not need instruction when the same task reoccurs.

The major predictor of strategy choice, however, was the robot’s learning progress—the increasing norm appropriateness of its action proposals over the course of the training. As a measure of this accumulating evidence of learning, we computed, at each trial starting with the second, the sum of previous action proposals the robot made, coding inappropriate proposals as -1, acceptable ones as 0, and appropriate ones as +1. This sum score weights both the quality and number of proposals the participant had received. (Note that the participant received proposals only in trials in which they chose to evaluate.) We found that, from trial to trial, a higher accumulated learning score predicted a lower probability of instruction, \( z = -8.4, p < .001 \). As the robot’s learning progress became more palpable, people moved away from instructing the robot and letting it instead act somewhat more autonomously (while still evaluating its progress).

B. Predicting People’s Evaluations

When people chose to evaluate, they evaluated the robot’s proposed action on a 1-5 rating scale. The overall mean of evaluations was toward the positive side (\( M = 2.93 \)), reflecting the experimental design in which we ensured that the robot’s action proposals tended toward appropriateness (the 18 pre-programmed proposals contained 3 inappropriate, 9 acceptable, and 6 appropriate actions). Importantly, we verified
that people’s evaluation scores for the robot’s specific proposals were responsive to the appropriateness of the proposals. Evaluation ratings for inappropriate actions ($M = 2.55$) were indeed lower than ratings for acceptable actions ($M = 2.96$), $t = -4.83, p < 0.001$, which were in turn lower than ratings for appropriate actions ($M = 4.09$), $t = 18.36, p < .001$.

Finally, returning to the prediction of teaching strategy choice, now at a trial-by-trial level, we wondered how the robot’s action proposal on a given trial affected people’s teaching choice on the subsequent trial. We conducted a generalized mixed effects model on the subset of trials immediately following a trial in which people had requested a proposal and evaluated it. Subject was a random effect, and last observed proposal (inappropriate, acceptable, appropriate), task block, task re-occurrence, and accumulating evidence of appropriateness (see above) were fixed effects, along with the interaction between last observation and accumulating evidence. Consistent with the previous pattern in people’s trial-by-trial decisions, we found an increase in instruction from first to second to third block, a decrease of instruction when a task reoccurred, and a decrease in instruction as people’s evidence of the robot’s appropriateness increased. Additionally, we found that an appropriate action proposal in the last trial decreased people’s likelihood to instruct the robot in the next trial, $z = -3.85, p < .001$, whereas an inappropriate proposal tended to increase that likelihood, $z = 1.85, p = .06$. Most interestingly, the impact of those previous-trial proposals varied as a function of the person’s accumulated evidence of robot competence. The stronger this accumulated evidence was, the more people were alarmed by an inappropriate proposal and the more likely they offered to instruct the robot in the next trial, $z = 5.5, p < .001$ (Fig. 3).

**IV. DISCUSSION**

We had hypothesized that teachers presented with an experienced robot would set high expectations and let the robot propose actions that they can evaluate. The experimental manipulation of robot experience did influence people’s first choice of teaching (they were more likely to *evaluate* than *instruct* an experienced robot), but this effect did not last over the teaching session. Instead, what increasingly guided teachers’ choices were their habitual preferences for one or the other teaching strategy and the immediate and accumulated evidence they gathered about the robot’s learning progress. People’s choices were dynamically responsive both to the appropriateness of the robot’s most recent action proposal and to the gradually stabilizing evidence of the robot’s learning progress. As a result, once people gathered more concrete evidence for themselves about the robot’s competence level, they were able to take it into account in their teaching strategy choices and evaluation scores. This observation illustrates that human teachers hardly hold a fixed impression of a novel robot; instead, they are sensitive to and respond to the robot’s developing competence in the teaching-guided interactions.

The current experimental design has several limitations. First, teachers gained direct information about the robot’s learning progress only on trials when they chose the *evaluate* strategy and thereby received the robot’s action proposal. In a new experiment, we plan to consistently give human teachers information about the robot’s current learning progress regardless of their teaching strategy choice. Further, we have implicitly assumed but not directly measured that people’s choice to instruct the robot reflects lower trust in the robot’s capacities and a choice of evaluation reflects higher trust. We plan to directly test this assumption in a new experiment that measures the relationship, across a full teaching session, between teaching strategy choices and subjective trust in the robot’s action plans.

**V. CONCLUSION**

This paper presents a novel experimental paradigm to examine which strategies human teachers choose when interactively training a social robot. We focused on two naturally occurring teaching strategies—*instruction* vs. *evaluation*—and found that most people used both strategies. More important, they used them in selective ways. For example, teaching by instruction was favored for the first encounter with a less experienced robot and for surprising inappropriate action proposals when the robot was otherwise progressing very well. Teaching by evaluation was favored as the robot’s learning progress became more evident, and for reoccurring tasks. The results suggest that human teachers dynamically update their perceptions and their teaching of a robot trainee. They do so both trial by trial, in response to the robot’s most recent performance, and over the course of the entire session, replacing initial expectations with accumulating evidence of the robot’s learning progress. Given that human teachers use multiple strategies and do so in systematic and meaningful ways, learning algorithms in social robots must be flexible and sensitive to the complex human teaching they are likely to receive.
REFERENCES


