Modeling Adaptive Expression of Robot Learning Engagement and Exploring its Effects on Human Teachers

SHUAI MA, Hong Kong University of Science and Technology,
MINGFEI SUN, University of Oxford,
XIAOJUAN MA, Hong Kong University of Science and Technology,

Robot Learning from Demonstration (RLfD) allows non-expert users to teach a robot new skills or tasks directly through demonstrations. Although modeled after human-human learning and teaching, existing RLfD methods make robots act as passive observers without the feedback of their learning statuses in the demonstration gathering stage. To facilitate a more transparent teaching process, we propose two mechanisms of *Learning Engagement*, Z2O-Mode and D2O-Mode, to dynamically adapt robots' attentional and behavioral engagement expressions to their actual learning status. Through an online user experiment with 48 participants, we find that, compared with two baselines, the two kinds of *Learning Engagement* can lead to users' more accurate mental models of the robot's learning progress, more positive perceptions of the robot, and better teaching experience. Finally, we provide implications for leveraging engagement expression to facilitate transparent human-AI (robot) communication based on our key findings.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; *Empirical studies in interaction design.*

Additional Key Words and Phrases: Human-robot interaction; Learning from demonstration; Transparent AI; Robot teaching; Robot engagement

1 INTRODUCTION

As a common form of AI systems, robots play an increasingly important role in assisting humans at work or in daily lives [20, 62, 72]. A traditional way to equip robots with the ability to perform tasks is to program the required skills into them in advance. However, in this way, robots can only execute fixed tasks and are not capable of learning new skills to adapt to users' diverse needs in dynamic and unstructured environments. Robot Learning from Demonstration (RLfD) [25], is thus proposed to address this challenge. RLfD provides a convenient, interactive method for non-expert users to teach a robot new skills by intuitively demonstrating examples [5, 6, 25], which has been successfully adopted in a variety of domains, such as daily task assistance [11, 120], manufacturing [36, 86], healthcare [11, 185], etc.

Generally, as shown in Figure 1 (a), a standard RLfD process usually includes two main stages: i) a demonstration gathering stage, where the human instructor shows demonstrations of a specific task, and ii) a policy deriving stage, where the robot student learns a policy from the human-demonstrated examples towards intended task outcomes [5]. These two stages in RLfD are carried out iteratively, with more demonstrations being added in the next round if the robot's learning outcome is not satisfactory. However, two main limitations exist in such a pipeline. First, it usually takes a long time for a robot to derive a rudimentary policy [5, 39]. Especially in the early stage of exploration-based learning (e.g., Reinforcement Learning-based policy deriving), it may be hard for the

Authors' addresses: Shuai Ma, Hong Kong University of Science and Technology,, shuai.ma@connect.ust.hk; Mingfei Sun, University of Oxford,, mingfei.sun@microsoft.com; Xiaojuan Ma, Hong Kong University of Science and Technology,, mxj@cse.ust.hk.

1073-0516/2022/11-ART $15.00
https://doi.org/10.1145/3571813

robot to show meaningful task outcomes, and thus the human teacher cannot get enough valuable information to assess the learning progress of the robot [94]. Second, robots usually remain static when human teachers give demonstrations [175], which makes it difficult for humans to reflect upon their teaching and make necessary adjustments during the interaction. Existing works have proposed methods to communicate what the robots have learned (or are yet to learn) to human teachers, including but not limited to demonstrating the robot's current learned policy by showing trajectories [31, 92], providing verbal/nonverbal cues about robots' learning status [21, 176], and visualizing where the robot succeeds and fails [159], etc. Nevertheless, these studies mainly present such robot learning feedback at the end of the policy deriving stage and overlook the demonstration gathering stage, where instructors receive little real-time response from the stationary robot student regarding what they are demonstrating [175]. It may impair teaching efficacy, decrease humans' interest in participating in RLfD [25, 94], and prevent human teachers from achieving their desired teaching objectives.

In comparison, human-human teaching and learning is a more reciprocal process between the instructor and learner. Human instructors would often like to get instant feedback from learners, which will help them perceive learners' learning status and adjust the teaching content accordingly [108, 192]. It has been shown that human learners usually convey their learning status by natural and intuitive communicative acts, such as gaze, motions, gestures, facial expressions, etc. [84, 106]. Based on these acts, instructors can appropriately update their beliefs about learners' knowledge and skills [13, 21]. Since RLfD is inherently a human-robot interaction process, we can adopt a similar idea to human learning into robot learning, where showing robots' learning status during the demonstration gathering stage in RLfD may help human instructors assess the effects of their demonstrations on robots' learning progress promptly. One naive solution to this problem is directly presenting robots' learning outcomes when the human teacher performs the demonstration. However, there exist two problems. First, as mentioned earlier, in the early iterations of the demonstration gathering stage [101], the learning outcome is usually unshaped. Second, when displaying learning outcomes (based on past training rounds), robots' behaviors are independent of what the teacher is currently demonstrating. If performed simultaneously, robots' actions may seem inconsistent with and/or irrelevant to the instructor's teaching dynamics in the ongoing round, leading to misunderstanding and confusion [99].

To address these issues, we propose a novel concept, *Learning Engagement*, for robot students to communicate their learning status to human teachers by showing engagement with teachers' live example(s) in the demonstration gathering stage. Engagement can be defined as the participation and involvement in an interaction, including emotional engagement, behavioral engagement, and cognitive engagement [166]. Student engagement is a typical signal for instructors to perceive the learner's inner status [192]. We thus propose to have robots convey their engagement in the demonstration process as a kind of feedback on their learning status, similar to human teaching and learning. The design of proper robot learning engagement expression should satisfy the following design requirements. **R1**) It needs to be closely coupled with human instructors' real-time demonstration behaviors and adaptive to the dynamics of humans' movements to enhance a real-time interactive experience because, in human teaching and learning, students are expected to engage in and adapt to the teaching dynamic [51]. **R2**) It should be able to reflect the actual learning status of the robot since human teachers tend to deem students' – humans' [192] or robots' [106] – engagement as an effective lens to inspect and update their mental model of learners' learning progress. **R3**) It should be represented in a human-understandable manner instead of simply displaying the updates of algorithmic parameters in robots' policy space to instructors as the RLfD users are mainly non-experts [5].

As shown in Figure 1(b), our proposed robot *Learning Engagement* expression is controlled by two components: humans' ongoing demonstrations so that the robot's engagement behaviors can correspond to the humans' teaching dynamic (**R1**), and robots' current learning status to help human teachers perceive the robot's learning progress accurately (**R2**). To achieve **R3**, we design two kinds of engagement cues. One is attentional engagement named *Gaze Following*, and the other is behavioral engagement called *Rhythm Synchrony*. To connect with human

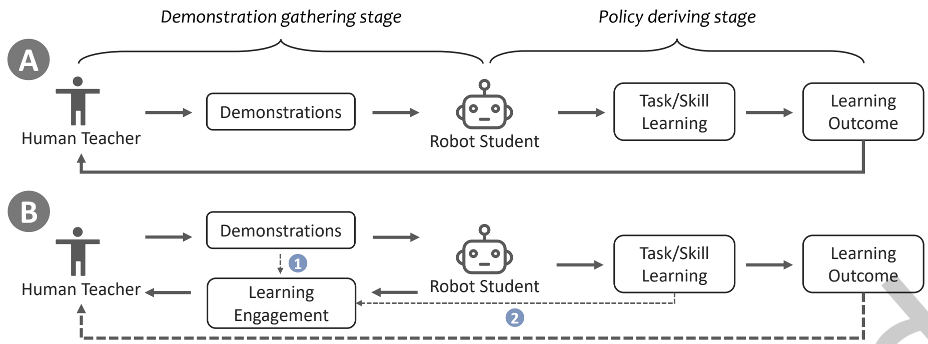


Fig. 1. The difference between a traditional RLfD pipeline and our improved pipeline. a) In the traditional RLfD process, a human teacher first gives a demonstration, then the robot learns to derive a policy, and finally executes the skill/task to showcase its learning outcomes to the human. b) With our proposed *Learning Engagement*, robots can engage in the demonstration gathering stage. Specifically, robots' *Learning Engagement* is driven by two parts: (1) humans' demonstrations to adapt the engagement expression with the dynamic of humans' teaching; (2) robot's inner algorithmic status to reflect its actual learning status.

demonstration (**R1**), in *Gaze Following*, robots track humans' key movements during a demonstration with their eye gaze to suggest their focus of attention. In *Rhythm Synchrony*, robots' limb movements follow the rhythm of humans' demonstrated actions to show a sense of involvement. We verify the communication effectiveness of these two engagement cues through two pilot studies. To control robots' expressed level of engagement according to actual learning status (**R2**), we incorporate the *Learning Status Indicator* from the backend learning algorithm as a regulator and design two types of *Learning Engagement* dynamics, i.e., From-Zero-to-One Engagement mode (Z2O-Mode) and From-Disorderly-to-Orderly Engagement mode (D2O-Mode), based on the common phenomena and educational theories in human learning [25, 147, 186]. Specifically, in Z2O-Mode, with the progress of robot learning status, robot engagement will gradually change from zero intensity to full intensity. By contrast, in D2O-Mode, as the robot learning status improves, robot engagement will change from a chaotic state (full intensity filled with random noise) to an orderly state (full intensity with no noise). With these two designs, we explore the following research questions: **RQ1**. How will the proposed *Learning Engagement* affect users' mental model of robots' learning progress? and **RQ2**. How will the proposed *Learning Engagement* affect users' overall perceptions of the robot and the RLfD process?

To answer the two research questions, we first developed an online simulated RLfD platform on which users can teach skills via webcam to virtual Pepper robots [122]. Robots were equipped with Z2O-Mode and D2O-Mode and two baseline modes for comparison, namely 1) Full-Mode, where robots always showed the same highly active engagement, regardless of their real learning statuses, 2) None-Mode, where robots always remained motionless in humans' teaching process. We then conducted a controlled, within-subjects online user study with 48 participants to investigate users' perceptions and experiences when teaching robots in different engagement modes. In the study, participants were asked to teach four given daily skills to four robots. Each robot was deployed in one engagement mode, and each skill-teaching task contained five teaching rounds. We collected participants' perceptions via in-task and post-task questionnaires and interviewed them after the study. As the user study was conducted online due to the pandemic and some experimental factors, we took several measures to enhance participants' sense of participation and reality.

For **RQ1**, results suggested that compared with the two baselines, the proposed *Z2O-Mode* and *D2O-Mode* had significantly different effects on users' perceptions of robot engagement, robot learning status, and the expectations of robot learning outcomes. Also, we verified that participants' perceived robot engagement had a significant positive correlation with the perceived robot learning status. Moreover, we found that *Learning Engagement* helped users perceive the actual learning status of the robot significantly more accurately compared with the two baselines. For **RQ2**, we found that with the proposed *Learning Engagement*, users had significantly better perceptions of the acceptability of robot behavior, the robot's intelligence, and the robot's potential for further improvement. And the analysis of participants' button-clicking logs revealed that participants' adjustment of demonstration and re-watching behaviors were significantly different when teaching robots in different engagement modes. Last but not least, participants reported being significantly less tired when teaching robots and significantly more willing to continue to teach the robot in the Z2O-Mode and D2O-Mode than in the baseline modes. We derive possible explanations of these findings from a qualitative analysis of participants' verbal feedback during the teaching process and post-study semi-structured interviews.

To the best of our knowledge, this work is a first step to enhancing the transparency of robot learning progress by incorporating robots' learning status into learning engagement expression. The key contributions of this work include: (1) We proposed *Learning Engagement* to facilitate the transparency of robots' learning status to non-expert users. Specifically, we proposed two *Leaning Engagement* modes, *Z2O-Mode* and *D2O-Mode* by incorporating robots' actual learning status into the proposed two kinds of engagement design, *Gaze Gollowing* and *Rhythm Synchrony*; (2) We conducted a controlled user study with 48 participants to investigate the effects of different engagement expressions on users' perceptions from multiple perspectives. Based on our key findings, we provided practical implications for the engagement design of robots and broader AI systems.

1.1 Contribution to HCI community

Despite the increasing adoption of AI techniques in everyday applications [41, 197], it is challenging for users to build an accurate mental model of the often black-box AI systems [41, 104, 140]. HCI researchers have been putting efforts into designing more transparent communication approaches between AI and humans, such as making model behavior more intelligible through explanations (a.k.a. explainable AI [40, 54, 104, 189]) and informing users of the inner working status of the AI system (e.g., capabilities, confidence, uncertainties, trustworthiness, etc. [9, 10, 48, 129]) in easily understandable ways [48, 151]. Our work carries on the HCI community's exploration of using human-centric design methods to help users interpret the opaque AI system [9, 40]. Specifically, we investigate the possibility of translating the inner status of an AI agent (the robot's learning status in this case) into its engagement expression based on HCI design theories, such as the "mental model" of users [128], "gulf of evaluation" [126], and "computers are social actors (CASA)" [124]. Our results confirm that lay users can directly understand the conveyed information by applying their established communication strategies from human-human interaction [163].

In brief, in this paper, we take a modest step forward in making the inner status of AI systems transparent for non-expert users via engagement expression. Our work has showcased that through the computational design of adaptive engagement expression (e.g., the proposed *Learning Engagement*), users can well grasp the internal status of an AI system (progress of the learning algorithm of a robot in this work) through engagement communication. By taking teaching robots as an example, we hope our proposed methods will provide insights into designing more transparent and interpretable AI systems for the broader HCI community.

2 RELATED WORK

2.1 Robot Learning from Demonstration (RLfD)

Robot Learning from Demonstration (RLfD), also called "imitation learning" or "programming by demonstration" [155], is a paradigm for non-expert users to teach a robot new tasks or skills by simply providing (live) examples [5]. RLfD research community has proliferated over the past decades, with a wide range of approaches developed for gathering samples from humans, modeling the tasks, and deriving the policies [25]. RLfD methods have been widely and successfully adopted in various real-world applications, such as manufacturing [36, 86], assistive and healthcare robotics [11, 96], etc. There are three kinds of interfaces for humans to provide demonstrations to robots in RLfD, including kinesthetic, teleoperation, and passive observation [143]. Similar to human-human teaching face-to-face, passive robot observation enables users to naturally teach skills using their own body, which is particularly easy for users to perform without operator training and has been successfully applied to various tasks [76]. In this paper, we also take the commonly used passive observation approach as the demonstration method [28, 138].

Existing research on RLfD mainly focused on two aspects: technical development and interactive experience. The former concerns algorithms and techniques for deriving robot policy to fulfill different kinds of tasks, such as using supervised learning to map states to actions [26, 153], utilizing demonstrations for reward shaping [125, 191] in Reinforcement-Learning-based RLfD [15, 69], and recovering rewards from collected demonstrations via Inverse Reinforcement Learning [1, 31], to name a few. As for the interactive experience, prior works primarily look into how to make a robot communicate with humans to collect demonstrations more efficiently, including but not limited to designing robots' feedback to convey information about what they want from humans by verbal and/or nonverbal cues [21, 176], letting robots ask questions during demonstration [17], showing robots' informative feedback to increase humans' engagement [100], visualizing where a robot learner succeeds or fails [159] to human instructors. Taking the two aspects together, Senft et al. proposed SPARC [160, 161, 193], a framework allowing humans to teach robots to interact socially in the real world by which the robot can progressively learn appropriate autonomous behavior from in-situ human demonstrations and guidance.

However, these works mainly study the pre- or post-demonstration gathering stages in an RLfD process and largely overlook how the robot's behavior could be leveraged to improve RLfD effectiveness and user experience in the demonstration stage. More specifically, existing studies mostly situate in scenarios where i) the human has finished showing demonstrations and the robot is running its learning algorithm with the captured demonstration as input, or ii) the robot performs the task based on a trained policy while communicating some information such as its capability or uncertainty to elicit the next round of human instruction [92, 100]. Nevertheless, in the demonstration gathering stage, the robot simply remains stationary, waiting for the completion of the demonstration. We argue that similar to reciprocity in human learning [177, 187], the demonstration stage can be turned into a two-way interaction with real-time feedback from robot learners to human teachers. Hence, in this paper, we aim to design appropriate mechanisms to make robot students engage in the teaching process. To the best of our knowledge, the most relevant work to ours is [175], where the authors explore how different kinds of robot engagement cues may affect users' perceptions. However, in their work, the robot is not equipped with an actual learning algorithm; thus, the engagement design cannot adapt to the robot's actual learning status. By contrast, in this paper, we deploy a policy-deriving algorithm on the robot and design *Learning Engagement* methods to dynamically integrate the robot's actual learning status with its engagement expression and then investigate the effects on users' perceptions from multiple perspectives.

2.2 Users' Mental Model of Intelligent Systems and the Need for Transparency

Norman's framework on human-centered design [127, 128] explains the dynamics between user mental modal and the actual status of a system/artifact. Simply speaking, when users interact with a system, they build internal conceptualizations of the systems that allow them to explain and predict the system's workings [47]. Norman [127] postulates that, "Good conceptual models are the key to understandable, enjoyable products: good communication is the key to good conceptual models." This idea is fundamental to the transparency demand of AI systems: to make optimal use of AI systems, it is critical that users maintain an up-to-date mental model of these systems. It has been shown that improved mental models contribute positively to user satisfaction and perceived control [88] as well as to overall trust in an intelligent system [107] and to better utilization of its decisions and recommendations [30, 156].

Communicating what status a system is in or explaining how the system works are effective transparency mechanisms to improve users' mental models of AI systems [88, 89]. In contrast, users are more likely to build flawed mental models when encountering an opaque system [128]. One way to enhance transparency in intelligent systems is the research of explainable AI (XAI) [53], which focuses on explaining and justifying the outcomes of AI-driven decisions or recommendations [83, 131]. Previous research has examined different types of explanation [115, 183], generating text or visualizations to offer local or global explanations of the AI models [116]. Another commonly studied approach to enhancing the transparency of intelligent systems is communicating the inner working status (e.g., uncertainty, confidence, and capability) of AI systems to users [41] through various designs such as icons [42], textual annotations [12], plots [148], and image/sentence highlighting [182]. One of the most commonly communicated information is the uncertainty of AI systems. Some prior works develop simplified, qualitative descriptions of uncertainty. For example, Ancker et al. [3] and Gkatzia et al. [49] use phrases like "very likely", "likely", and "unlikely" to describe how probable an outcome is. And some research has demonstrated more expressive representations of uncertainty, such as visualization, for non-expert users to understand. Jung et al. compare a gradient plot versus a point estimate to communicate the remaining battery level range for an electric vehicle [74], which reduces driver anxiety in a driving task. Hohman et al. deploy a "regions-of-error" technique showing the model uncertainty of predictions for ML experts [63]. These text or visualization-based approaches are effective for displaying the uncertainty of AI systems in some cases. However, they are not related to users' dynamic interaction behaviors and are inappropriate when a user focuses on inputting data into the system (e.g., demonstrating a skill to the robot). Thus, we propose to communicate the internal status of the AI system by expressing its engagement when users interact with the AI to fit users' dynamic interaction behaviors.

Note that "mental model" has different definitions in different interaction scenarios [48]. In this paper, we define the mental model of users as their perceived robots' learning status during the teaching process. And in our work, we take the robot-based AI system as an example and focus on communicating robots' internal status (i.e., its learning status – the progress of the learning algorithm) by expressing adaptive engagement during users' input process.

2.3 Transparency Design for Robot Internal Status

Researchers have proposed many feasible methods to express robots' internal status to human partners, such as through motions and trajectories [66, 92, 134], language and voice [17, 19, 81, 142, 157, 170], visual display [46, 67, 100, 137, 180], and social cues [21, 106, 132, 196].

First, robots' motions and trajectories have been widely used to externalize their states. For example, [66] proposed to convey uncertainty in a pick-and-place robot by extending its waiting time and reducing its moving speed. [134] exploited the speeds of motion to communicate a learning agent's uncertainties in taking action. Besides motion, robots can directly use natural language and voice to announce their status in human-robot tasks, which is shown to improve team performance [157, 170]. For example, [142] enabled robots to provide easy-to-understand feedback in natural language to users on tasks that cannot be achieved. [19] equipped social robots with the ability to express their curiosity verbally. [81] manipulated the tone of robots' voices to inform their confidence in suggestions. The visual display is another effective alternative to express robots' internal states. For example, visualization can illustrate changes in agents' uncertainty [100], and can summarize the results of scene detection and command recognition [137]. Besides, the visual display on robots' bodies has also been exploited, e.g., expressing robot emotion by changing the body's color luminosity [180] or skin texture [67]. Furthermore, programmable lights on robots can help express their intent [46]. Moreover, animation and visual metaphors can also serve as status cues. For example, [37] used non-semantic and semantic icons to suggest the confidence level of a robot. In addition, researchers proposed to integrate multi-modal signals together to convey robots' states, such as robots' emotions by combining multi-modalities, such as color, sound, vibration, and light [168, 169]. In humanoid robots, social cues, e.g., facial expressions [132], gaze [106], gestures [21], etc., are widely adopted to foster mutual understanding with human partners. For example, [21] had a Simon robot apply nonverbal gestures to query a human teacher about areas of uncertainty in the underlying model. [196] proposed to express a robot's confidence level via body orientation; if the robot is less confident, it will slowly turn to the user. [70] designed a mechanism to control the timing of a robot nodding to signal a humanoid robot's agreement, interest, and confidence when talking.

Existing research on robot confidence or uncertainty expression shares a similar idea with our work. However, we focus on the RLfD scenario, especially exploring how to show the internal learning status of a robot during the demonstration gathering period, when the human instructor expects to get feedback on the demonstration in real-time.

2.4 Engagement and Robot Engagement in Learning

Broadly defined, engagement is the process of initiating, maintaining, and terminating the interaction between humans and other interactive parties, which can be humans, computers, and robots [165]. More specifically, in HCI, engagement can be categorized into three main types: emotional engagement, behavioral engagement, and cognitive engagement (including attentional engagement) [109, 166]. Each type of engagement represents a different aspect of human participation and involvement in an interaction. Emotional engagement refers to humans' affective reactions towards the interaction, such as interest, anger, excitement, frustration, and boredom [95]. Behavioral engagement refers to humans' physical behaviors when participating in the interactive process [33]. Cognitive engagement refers to humans' psychological devotion to the interaction [103] such as thinking, reflection, attention allocation and redistribution, etc. [22]. When it comes to human-robot interaction, people have the tendency to regard the robot as another human partner, especially if it has a human-like appearance [117], and thus they may seek engagement cues from the robot as in human-human interaction. Inspired by human engagement expression, researchers explore the use of similar social signals such as eye gaze [2], and body posture [179] to indicate robot engagement.

In the context of learning, robot engagement can also be represented through the above three aspects: cognition, behavior, and emotion [133, 166]. i) Among all the cognitive states, the allocation of attention is one of the most important cognitive resources [133]. Thus we focus on the robot's attention expression. Based on human-human teaching and learning, attentional engagement is highly related to the learning process [147, 186]. Focal attention implies a positive state, while distraction such as divided attention [75] and mind wandering [112] are often associated with negative states. In HRI, a robot can signal its attention via different cues, e.g., gaze [2, 90, 106], head orientation [106, 176], and body postures [179]. ii) Behavioral engagement is usually represented by task-related activities, e.g., task attempts, efforts, active feedback, etc. [175]. Among all possible behavioral engagement cue candidates, mimicry and behavioral synchronization are two common ways to express one's engagement unconsciously, which refers to "non-conscious mimicry of the postures, mannerisms, facial expressions, and other behaviors of one's interaction partners" [23]. In both human-human and human-robot interaction, the imitation behavior can increase the likelihood of understanding [24], interpersonal coordination [25] and emotional contagion [56]. iii) Emotional engagement is associated with the affective status in the learning process, such as boredom, anxiety, confusion, etc. Despite its importance, emotional engagement is hard to generalize to most existing RLfD scenarios since the robotic systems lack the full ability to express emotions [175]. As a result, in this paper, we focus on communicating robots' attentional and behavioral engagement in the RLfD scenario.

In summary, although existing works have researched the expression of general robot engagement, few have focused on the RLfD interaction scenarios. To fill this gap, in this paper, we propose a novel method to integrate the learning status of the robot with the expressed engagement to adapt to the dynamic and reciprocal RLfD teaching process. It is worth mentioning that beyond the designed expression of engagement for the robot-shaped AI system, our proposed method can be adapted to other kinds of AI systems, such as GUI-based intelligent systems, by showing engagement through various forms, e.g., text [163], icons [55], colors [58, 164], cartoon and animations [114, 195], etc. Thus, engagement expression could be applicable in a wide range of human-AI interaction scenarios.

3 DESIGN OF LEARNING ENGAGEMENT

In this section, we first briefly introduce the representation of human body poses which forms the basis of robot learning from physical demonstrations by humans. Then we design two forms of engagement cues to communicate robots' learning status: one signals attentional engagement through *Gaze Following*, and the other indicates behavioral engagement via *Rhythm Synchrony*. Finally, we propose two *Learning Engagement* modes, the *Z2O-Mode* and the *D2O-Mode*, combining the robot's engagement expression with the actual progress of the underlying learning algorithm.

3.1 Representation of Human Body Pose

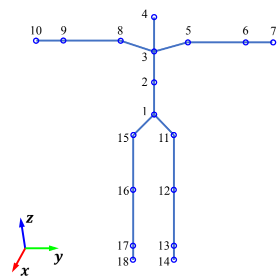
Demonstrations in RLfD, i.e., the expert's body poses [143], are usually represented in a tree-like structure, with pose joints as its nodes and pose bones as its edges. As shown in Figure 2, the body poses can be denoted in two ways: global position-based representation or local transformation-based representation.

Position-based representation describes human body pose in a single global sensor reference frame, as shown in Figure 2 (a). More specifically, the pose skeleton is denoted as $[J^{(1)}, J^{(2)}, ..., J^{(n)}]$, where $J^{(i)} \in \mathbb{R}$ is the position vector of the $i-th$ human joint in the skeleton, and n is the total number of joints. This representation allows us to get the global position of each joint on which the proposed attentional engagement cue *Gaze Following* and the behavioral engagement cue *Rhythm Synchrony* are based.

Transformation-based representation describes human body pose with a series of frames of reference [174], as shown in Figure 2 (b). It sets one of the joints (often the hip joint) in the skeleton as the root node, and this node is described in the sensor reference frame. The other joints have their own (right-handed) reference frames, and the links in the tree-like skeleton define the parent-child relationship between two connected joints, e.g., the left elbow joint is a child of the left shoulder joint. The pose of a non-root joint is then described by a translation (body-dependent, i.e., the bone length) and a rotation (body-independent, i.e., joint movement) in its parental reference frame. In this way, a human body movement can be denoted as $[T_1, T_2, T_3, ..., T_n]$, where T_i is the translation and rotation vector of the $i-th$ joint in the skeleton. This representation helps us to get the target orientation of the demonstrator's poses, which is used for the *reward* function and *state* representation in the learning algorithm design (detailed in Sec. 3.4).

3.2 Attentional Engagement Cue: *Gaze Following*

3.2.1 Design. In the context of learning, since the allocation of attention is an essential cognitive engagement [133, 166], we employ attentional signals as a cue to communicate robot engagement in RLfD. One of the most effective indicators of attention is eye gaze which is a key component of human cognition and guides attention to areas with high information value. On the one hand, gaze can signal a willingness to engage in the interaction and



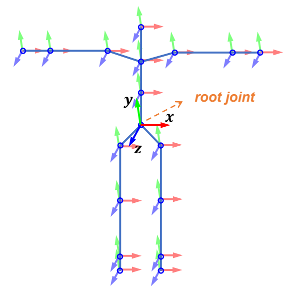


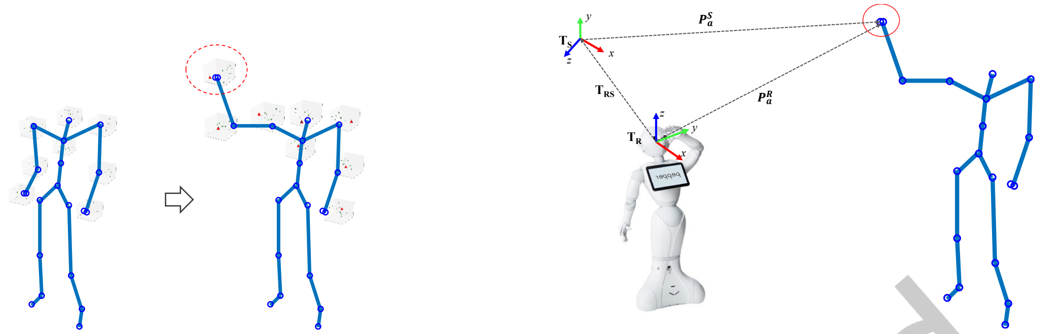
Fig. 2. Two representation forms of human body pose [175]. (a) The position form: all joints are described by their positions in a single sensor reference frame; (b) The transformation form: each joint has its own reference frame, and the skeleton defines parent-child structures and translations between frames; except that the root joint is referred in the sensor reference frame.

accelerates learning [113, 179]. On the other hand, gaze is instrumental to (visually guided) imitative learning, and its pattern helps discern whether a learner is focusing on learning goals and instructors' intentions [7, 113]. In human-robot interaction, robots' gaze following is recognized as an effective means to signal engagement in social learning [2, 106, 154]. In LfD settings, human instructors' goals and intentions are presented through their bodily actions, and thus the most informatively salient areas of their body pose should be the target of gaze following.

We propose the expression of attentional engagement for robots based on the cognitive theories of human attention. Generally speaking, a generation process of human visual attention involves two stages [73]: first, attention is distributed uniformly over the visual scene of interest; then, it is concentrated to a specific area for gaining information [43]. Therefore, the first challenge in designing the attentional engagement cue is to determine which part(s) of human movement the robot student should gaze at. A simple but effective way to measure the importance of each joint is by its extent of position change after a sub-movement. First of all, we use the aforementioned position-based representation to get the position of each joint in the global reference frame of a human body skeleton and utilize joint positions observed in the past several poses to model the temporal position distribution P_j of each joint j. Then we compare the distance between each joint's current and precedent position distributions. A larger distance suggests a bigger difference between the two observations, and thus it means the corresponding joint contains more information induced by the current motion.

In the current implementation, we measure the importance of joint points by the absolute magnitude of the position change. This works well in the skill-learning tasks in this paper. However, in some cases, slight movements may deserve more attention than big movements. We can deal with this situation by using the relative degree of position change to measure the joint importance. In this way, even for a slight movement, as long as the joint has a relatively large change compared to its past movement, it is reasonable to consider the movement worthy of attention.

Figure 3 (a) illustrates how to get the attention point. During a teaching process, when the human's body moves, the current position of each target joint is tracked, and we calculate the distance between the current position and the mean of the last n positions' distribution via Euclidean distance. The joint with the largest derivation from the past distribution (joints' past positions) will be regarded as the attention point P_a. Note that the attention point P_a is generated in the sensor reference frame (actually, now it is P_a^S) and needs to be further



(a) (b)

Fig. 3. An illustration of the *Gaze Following* method. (a) Finding the attention point by comparing the distance between the current position and the position distribution of each target joint. (b) Transformation of attention point P_a from the sensor reference frame T_S to the robot head reference frame T_R based on the transformation T_{RS} [175].

aligned to the robot's reference frame. We take the same strategy as in [175] to transform P_a^S by left-multiplying the transformation T_{RS}, i.e., $P_a^R = T_{RS} P_a^S$, as illustrated in Figure 3 (b), where T_{RS} refers to transformation from the sensor reference frame T_S to robot head reference frame T_R. Then we can control the orientation of the robot's head to direct to P_a^R.

To determine the number n of the last positions to track, we consider both sensitivity and stability. If too few past joint positions are taken into consideration, the method will be unstable due to noise from the Mocap demonstration data. On the contrary, if too many past joint positions are considered, the attention point P_a will react and update too slowly, which can lead to important actions being missed. To strike a balance between stability and sensitivity, the number of past joint positions to be tracked is empirically set to 10, which works well in the selected skills in this paper. For other skills/tasks, we suggest that the number of past joints to track should be determined mainly based on the frequency of skills. If the changes of skill actions are frequent, this number should be appropriately reduced because it is easy to capture changes at this time, and a smaller time step is beneficial to capture the latest changes; and when the movement changes of the skill are relatively slow, this number should be appropriately increased; otherwise, the obvious movement changes cannot be captured only with a small number of past joint positions. Therefore, the choice of this parameter can be set adaptively, roughly inversely proportional to the frequency. In addition, we set the refresh rate to ten frames, which means that we update the attention point every ten frames (choose the most frequently occurring attention point).

3.2.2 Verification. We did a pilot study to investigate whether the designed *Gaze Following* is appropriate and effective to express robots' attentional engagement in terms of four aspects: 1) perceived movement smoothness, 2) perceived movement stability, 3) perceived gaze accuracy, 4) perceived engagement expression. We deployed a Pepper robot and a human-skeleton avatar in a simulation environment (detailed in Sec. 4) to showcase the RLfD process and used a basketball-shooting motion as the target skill to learn. We used the motion capture data from the CMU Mocap dataset[1] to drive the human avatar to play and shoot the ball. During the human avatar's demonstration, the robot was driven by the *Gaze Following* method to direct its gaze on the human avatar in real time. We recorded the human-robot interaction process from a third-person view (the illustration can be

[1]http://mocap.cs.cmu.edu

found in the online Appendix [2]) and showed the videos to 10 participants (5 Female; age *mean* = 24.9, *SD* = 2.1) recruited from a local university. After participants watched the videos, we asked whether the robot's eye gaze was smooth, stable, and following the human's movement ("3 - Yes", "2 - Hard to say", and "1 - No"), respectively. We also invited them to rate their perceived level of robot engagement in the learning process on a 7-point Likert scale (1:Extremely disengaged, 7: Extremely engaged)). Nine out of ten participants considered the robot's gaze movement was smooth and stable (one person found it hard to say), and all of them positively confirmed the gaze-following accuracy. The average engagement level rating was 6.5 (SD = 0.7). These results suggested that the proposed *Gaze Following* cue was appropriate and effective in expressing the robot's attentional engagement.

3.3 Behavioral Engagement Cue: *Rhythm Synchrony*

In human-human interaction, behavioral engagement is often conveyed by one's physical participation [109]. Among all the social behaviors that happen in an interaction, mimicry and behavioral synchronization are two common ways to express one's engagement unconsciously [141]. Mimicry is typically defined as spontaneous, immediate imitation of gestures, postures, and the dynamics of movements of another person [91, 93], while behavioral synchronization is regarded as the mutual alignment of interaction partners' behavior on a larger time-scale [141]. There has been a lot of evidence in social science and psychological science showing that mimicry and synchronization can promote rapport [23], trust [110], altruistic behavior and liking between interacting partners [184], and can be interpreted as engagement [99]. Mimicry can be implemented without actual policy learning algorithms, as demonstrated by the method *Approximate Imitation* in [175], which allows the robot to directly reproduce to the best extent possible pose of each human joint on the corresponding robot joints. However, it may not be appropriate in our teaching scene because if it is visually too similar to the learning target, human instructors may misinterpret such engagement cues as the robot's actual learning outcomes [175, 179], especially in the motion learning task. Therefore, in this paper, we propose communicating robot learners' behavioral engagement through behavioral synchronization.

3.3.1 Design. Rhythm synchrony is essential in behavioral synchronization [35]. Although implicit, it is a critical signal in human-human interaction [123, 150]. Also, there is evidence showing that robots performing synchronization behavior, such as following human movement's rhythm, can lead to a pleasant sense of interaction with users [4, 121]. Inspired by these findings, we design *Rhythm Synchrony* as a kind of robot behavioral engagement cue in RLfD. The general idea is to let the robot make some slight bodily motions rhythmically according to the rhythm of human actions. For example, limb movement as simple as tapping and swinging is commonly used to show rhythm [61, 121]. Considering the humanoid robot Pepper, for instance, we can leverage its upper limbs, including its hand, elbow, and shoulder, to implement such a rhythm.

To leverage rhythm for synchronization of the human demonstration, the first problem to solve is how to obtain the rhythm pattern in a human demonstration. Here we introduce the concept of *period* in our tasks. We define the *period* as an interval in which a joint moves from a relatively extreme position to another in an action sequence. The duration of each period can be different, and the end position of a joint motion trajectory in a period is not necessarily the same as the start position [121]. For example, drawing a square with a hand in the air can be regarded as containing four *periods*, each corresponding to one stroke. We can show that a robot is following the rhythm of the human demonstration by making the robot act in the same period as the human demonstration action. Without loss of generality, to get the periods in a movement, we need to find temporal segmentation points which feature discriminable events such as local extrema of the movement with vanishing velocity as described in [85, 121]. And a period can be delimited by two segmentation points. More specifically, to get the periods of a demonstration movement, we need to go through two steps. i) Step 1, get the velocity of target joints at each time step. As only upper-limb movement skills are taught in our tasks, we can focus

[2]https://userstudy.link/OnlineAppendix/appendices.pdf

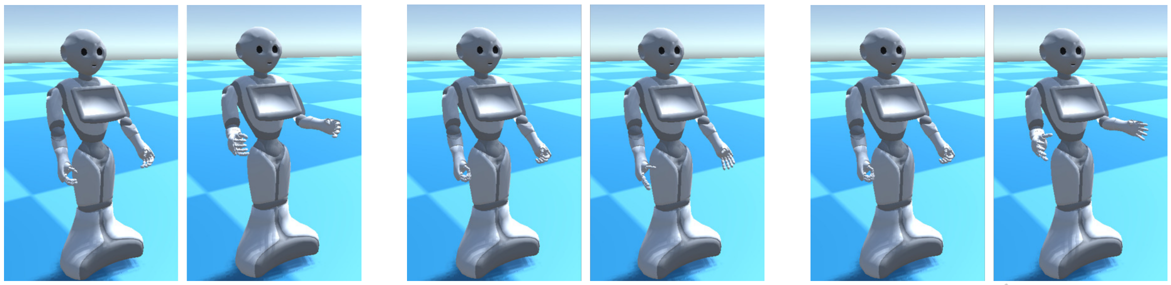


Fig. 4. Three candidates of rhythm synchrony behavior. a) The robot raises and lowers its forearms. b) The robot clenches and loosens its fists. c) The robot raises and lowers its forearms while clenching and loosening its fists.

on the two arms and then get the average velocity of joints in these two parts. The velocity value Vel_t of one joint at a given time t can be approximated by finite-difference $\mathrm{Vel}_t \approx \frac{|\mathrm{Pos}_{t+\Delta t} - \mathrm{Pos}_t|}{\Delta t}$, where $\mathrm{Pos}_{t+\Delta t}$ and Pos_t are the joint positions at time $t + \Delta t$ and t in the sensor reference frame respectively. ii) Step 2, get segmentation points based on velocity. After accumulated observing the demonstration over a series of movement steps, we can get a sequence of corresponding velocity values (which can form a velocity curve) of a given body part. Then, we adopt a filter to reduce noise interference. In this paper, we use Savitzky-Golay filter [139] with the advantage that it can retain the change information of a signal effectively while the filtering is smooth. Then, on the velocity curves, we can obtain the temporal segmentation points of the corresponding body part, with the criteria that the average joint velocity is close to zero and lasts for more than five timesteps (frames) which is a noticeable duration in our settings. Once we get the segmentation points, we treat the intervals between each pair of adjacent segmentation points as periods. After getting the periods of the left arm and right arm of the human demonstrator, we can synchronize the rhythm of the left and right arms of the robot accordingly. Note that there is a cold start problem with this approach, and we need the human instructor to perform the demonstration once in advance so that the robot can leverage this practice teaching round to get the initial rhythm of the movement and adjust it in the actual teaching rounds. Such a design is applicable in RLfD tasks, especially in multi-round skill teaching tasks, as the demonstration is easy to perform and adjacent two rounds of demonstrations are usually very similar in pace.

The second problem to solve is how to show robots' rhythmic behavior. Following some common human behaviors in displaying synchronization in real-world motion skill learning [50, 172], we identify two ways to express a robot's rhythmic movement in synchronization with human actions: raising/lowering the forearm(s) and clenching/loosening the fist(s). Note that the specific choice and the moving range of the robot joint(s) involved are not fixed and should be determined by the particular tasks for the robot to learn and by the specific physical form factors of the robot.

We take the rhythmic movement design of the robot's elbow joint as an example to introduce the detailed design of *Rhythm Synchrony*. First, we drive the robot's elbow by changing the Euler angle θ_{roll} of the joint, which will produce the visual effect of raising and lowering the forearm. Through empirical testing, we set the reasonable *movement range* of the elbow's θ_{roll} to be 0.2-0.9 (the two extreme joint states are shown in Figure 4 (a)). Then, we can map the obtained periods into the *movement range* of the elbow joint, by adjusting the value of its θ_{roll} at a uniform speed from one end to the other end in one period, and back in the next period (illustrated in Figure 5). Namely, the rhythm of the robot is represented by periodic raising-lowering of its elbow joint, and the movement is synchronized with the speed and timing of the human demonstration. Similarly, we design

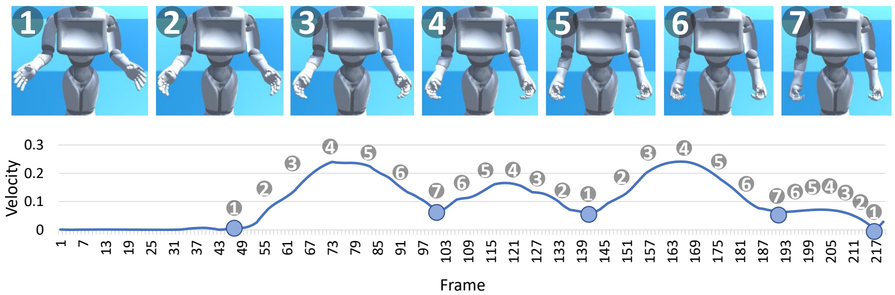


Fig. 5. An illustration of the *Rhythm Synchrony* method. The first row shows seven uniformly sampled positions of the robot's upper limb movement (the movement of the upper body is a continuous process, and we choose seven moments just as examples). The second row shows a curve of the velocity values of a demonstrator's movement, where the five blue points indicate the segmentation points. Two adjacent segmentation points determine a *period*. From each segmentation point to the next, we manipulate the robot's upper limb movement from one extreme state to another, e.g., from 1 to 7 or from 7 to 1.

another expression of rhythmic movement through the fist relaxing and clenching action of the robot (shown in Figure 4 (b)). In addition, we combine the two together as a third way to convey synchronization – *Rhythm with fist&elbow* (shown in Figure 4 (c)).

As suggested by [175], setting one second of delay for the robot's movement is more acceptable to users. If the robot generates the movement in a very responsive manner, users are likely to feel that it is acting on itself rather than following the demonstrator's movement. Consequently, we make the robot 1s slower than the human's movement in *Rhythm Synchrony*. However, since the eye-gaze in human-human interaction is adjusted in real-time, we do not set any delay for the robot in *Gaze Following*.

3.3.2 Verification. To get the most appropriate behavioral engagement expression for RLfD, we conducted a within-subjects pilot study to compare the following four behavioral engagement expressions: 1) approximate imitation (implemented following [175]), 2) rhythm following with elbow, 3) rhythm following with fist, and 4) rhythm following with fist and elbow. We invited 12 participants (5 Females) from a local university, with an average age of 24.6 (SD = 1.9). In the same virtual environment used in the previous pilot study for *Gaze Following* testing, we let the human avatar teach the Pepper robot the swimming action and had the robot express its behavioral engagement through the above four forms in the learning process. The participants watched the video recordings of the robot's four forms of behavioral engagement cues in random order and answered 5-point Likert scale questions (1: Strongly disagree, 5: Strongly agree) after each video. The questions were regarding whether behavioral engagement 1) is easy to be confused with the robot's learning outcome, 2) can make people feel that the robot is engaged in learning, 3) is easy to observe without too much attentional load, and 4) makes sense (looks natural and reasonable).

The results analyzed by Friedman test with Wilcoxon post-hoc are shown in Figure 6. We can find that the approximate imitation is significantly most likely to mislead users to think that the robot was showing its learning outcome (with all $p < .001$ compared with three rhythm synchrony cues), even though it also stimulates a strong sense of high engagement. Among the synchronization-based cues, the combined elbow and fist rhythm cue is considered the most appropriate and easier to observe among all. Based on these pilot study findings, we choose *rhythm with fist and elbow* as our final behavioral engagement cue for the later design and experiment.

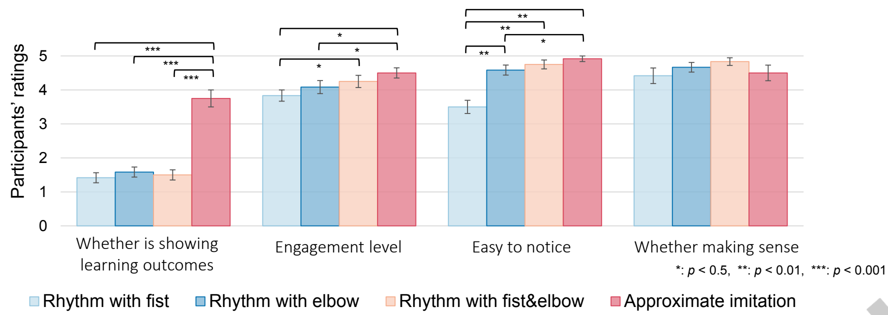


Fig. 6. Results of the pilot study (the error bar represents the standard error).

3.4 Learning Engagement

In this section, we first introduce the learning algorithm used in this paper, and then we propose two kinds of *Learning Engagement* that incorporate the learning status of the robot into the engagement expression.

3.4.1 Learning Algorithm and Learning Status Indicator. Mathematically, RLfD can be formulated as an infinite-horizon Markov Decision process (MDP) problem with a finite state space $\mathcal{S}$, a finite action space $\mathcal{A}$, a transition kernel $p : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$, a reward function $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$, a discount factor $\gamma \in [0, 1)$, and an initial state distribution p_0 from which the initial state s_0 is sampled. The goal of the learning task is to learn a policy $\pi \in \Pi$ ($\pi : \mathcal{S} \times \mathcal{A} \to [0, 1]$) such that the expected cumulative reward is maximized. Usually, a demonstration trajectory is denoted as $D = [(s_1^d, a_1^d), (s_2^d, a_2^d), ..., (s_n^d, a_n^d)]$, where (s, a) is the state-action pair. And the demonstration trajectories are usually regarded as generated by rolling out an expert policy to simulate an MDP.

We chose Q-Learning as our algorithm based on two reasons. First, due to its simplicity and effectiveness [118], Q-Learning (with its variations, such as DQN) is a widely adopted algorithm in LfD [59, 60]. Second, Q-learning is especially suitable for LfD scenarios where the reward function is already defined, which is in line with our case.

Q-Learning seeks to learn a policy that maximizes the expected action value, i.e., cumulative discounted rewards starting from the initial state distribution, where Q is the action value function. Before the learning begins, Q is initialized to a possibly arbitrary fixed value. Then, at each time t the agent selects an action a_t, observes a reward r_t, enters a new state s_{t+1}, and Q is updated. The core of the algorithm is a Bellman equation as a simple value iteration update:

$$Q^{\text{new}}(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\Big(r_t + \gamma \cdot \max_a Q(s_{t+1}, a) - Q(s_t, a_t)\Big) \tag{1}$$

where α is the learning rate which is set to 0.5 to balance the previous knowledge with the new reward, and γ is the discount factor which is set to 0.9 so that the algorithm can fully consider the external reward. Besides, to leverage the exploration-exploitation trade-off [34], we adopt the epsilon-greedy policy and set the epsilon to be 1 with an anneal decay to be 0.001 to control the percentage of exploration and exploitation. In the following, we introduce the definitions of action, state, and reward in the context of generating learning engagement.

Action. As suggested by [136], the action a can directly specify the torques of robots' joints. Since Q-Learning is only feasible for discrete action space, we need to discretize the action space. To simplify the question, we design the action as one of the three operations of the Euler angle of each joint's configuration, $+0.1/-0.1/keeping\ still$. Since our requirements for accuracy are not high, such a design method can narrow the action space and simplify the question without loss of generality. Considering Pepper as our example robot, we select skills that can be performed via upper limbs, so only the robot's upper limbs will be involved in the movement. Hence, the final action space is a ternary choice ($+0.1/-0.1/keeping\ still$) of these 8 correlated joint configurations, *Left/Right Shoulder's* θ_{pitch}, *Left/Right Shoulder's* θ_{roll}, *Left/Right Elbow's* θ_{yaw}, *Left/Right Elbow's* θ_{roll}, while satisfying the robot's physical constraints.

State. Following previous work [135], we define the state of the robot as the rotations of each corresponding link expressed in Euler angles and computed in the robot's local coordinate frame. With the same state feature transformation, we can also get the target state by extracting the same set of features from the demonstrator's pose. As suggested by [136], we augment the state features with a set of target features resulting in a combined state represented by the concatenation of the robot's features and the target (human demonstrator's) features, $\Phi(robot)||\Phi(target)$. We also conduct discretization for the state representation. Note that in our state design, to simplify the learning question, we only consider the robot's pose learning. While in a more comprehensive design, the height of the robot's root (pelvis) from the ground, the center of mass velocity of each link, the relative positions of each link with respect to the root, and other factors related to the environment should be taken into consideration.

Reward. In this task, we design a reward that encourages the robot to match the joint orientations of the demonstrated motion at each timestep, as we want the robot to perform the skill as similarly as the human instructor. The reward can be formulated as the difference between the joints' Euler rotation angles of the robot motion and those of the human demonstrated motion. The more similar the two are, the higher the reward value will be. Similar reward designs can be seen from locomotion imitation tasks [135, 136].

$$Reward = exp\left(-\sum_{i=1}^{N}(J_i^{\text{H}} - J_i^{\text{R}})\right),\ J_i^{\text{H}} \in Config^{\text{H}},\ J_i^{\text{R}} \in Config^{\text{R}} \tag{2}$$

where $J_i^{\text{H}}/J_i^{\text{R}}$ is one of the human's/robot's joint configurations (in Euler rotation angles form). The training workflow of our task is: once the first demonstration is given, the learning process gets started and goes through iterative training. If a new demonstration is received during training, the algorithm will use the newly demonstrated motions as the optimal reference for the following round of policy deriving.

Learning Status Indicator. After deciding on the learning algorithm of the robot, we need to find out a learning status indicator to represent the training status of the algorithm. In reinforcement learning, based on learning goals there are many learning status indicator candidates, such as the entropy [68, 162], the value of the loss/cost function [44, 105], the normalized cumulative reward (NCR) [136], etc. In this paper, we adopt entropy [146], which is widely used to measure a model's uncertainty [10], as our learning status indicator because it can reflect the stability of Q-Learning representing whether the underlying policy is well-trained or not.

We define the probability of one action as the exponent of its Q-value out of the sum of the exponents of all actions' Q-values (Eq. 3). And the entropy of each state is defined as the sum of all actions' entropy values (Eq. 4). Thus, for a current timestep t, the entropy of the policy can be calculated by averaging all the entropy values of each state in training steps from t_0 to t (Eq. 5).

$$P(a_i|s_i) = \frac{exp(\tau \cdot Q(s_i, a_i))}{\sum_j exp(\tau \cdot Q(s_i, a_j))} \tag{3}$$

$$E(s_i) = -\sum_{a_i} P(a_i|s_i) log P(a_i|s_i) \tag{4}$$

$$E = \frac{1}{|S_{t_0:t}|} \sum_{s_i \in S_{t_0:t}} E(s_i) \tag{5}$$

After getting the entropy, we design the *learning status indicator* as below:

$$LS_e = 1 - \frac{E_e - E_{min}}{E_{max} - E_{min}} \tag{6}$$

In this equation, we first normalize the entropy at any given episode E_e by Min-Max Normalization, where E_{max} is the maximum entropy of all training episodes (usually it occurs in the initial stage of the training process), and E_{min} is the expected minimum entropy of all training episodes (can be estimated through pre-training or prior training). Under such design, if the robot's policy is under-developed, E_e will be close to E_{max}, and the learning status indicator LS_e will be close to 0. In contrast, if the robot's policy is well-developed, E_e will be close to E_{min}, and the learning status indicator LS_e will be close to 1.

Note that we should be cautious about two issues when using entropy as the learning indicator. On the one hand, the minimum entropy needs to be estimated from the previous training, which may be slightly different from the minimum entropy that the current training process can achieve. On the other hand, the entropy calculation can sometimes be affected by outliers, which may cause fluctuation in different training episodes. However, in our target scenario, it is appropriate to choose entropy as the indicator for two reasons. First, under the same definition, the robot's action space and state space when learning similar skills are also similar, so the estimated minimum entropy from the previous training will not be much different from the minimum entropy in the current skill training process. In addition, the five rounds of learning we collected were selected over a long time span, which is also in line with the common practical scenarios of LfD. When calculating the entropy of a certain round, the influence of fluctuations on the current round is almost negligible, and it can still faithfully reflect the learning state of the robot in the current time period.

Once finishing the design of the learning algorithm and the calculation of the learning status indicator, the next step is to combine the learning status indicator with the previously designed two kinds of engagement expression, *Gaze Following* and *Rhythm Synchrony*. Next, we propose two forms of *Learning Engagement* that can reflect the internal learning status of a robot via engagement expression.

3.4.2 Learning Engagement 1: from zero to one, the Z2O-Mode. In human learning, we have this observation: one tends to be more active if he/she gains more confidence in learning [16]. Inspired by this, we designed the first learning engagement which reflects the learning status of robots by adjusting the intensity of robots' actions. Specifically, if a robot is in a good learning status, its actions can be designed to be quick and powerful, and if it is in a poor learning status, its actions can be designed to be slow and powerless. To achieve this effect, first of all, according to our previous design of attentional and behavioral engagement, each joint has a target configuration to control the robot's behavior at each moment, e.g., its head joint is configured to direct the eye gaze to the attention point, and its elbow joint is configured to raise arms to an exact height, etc. Thus, each time step has a set of *current configurations* of each joint and a set of *target configurations* of each joint based on the proposed engagement expression. If we want the robot to show full engagement, we can directly adjust the corresponding joints from the current configuration to the target configuration. Taking the robot's learning status into consideration, if the robot's learning status is poor, we will let the robot move slightly. On the contrary, if the robot's learning status is good, we will let the robot move more obviously, which can be achieved by setting a motion decay as follows:

$$config_i = config_i^C + (config_i^T - config_i^C) \times LS_e \tag{7}$$

where $config_i^C$ is the **c**urrent configuration of joint J_i, $config_i^T$ is the **t**arget configuration of joint J_i, and LS_e is the robot's learning status in the current episode, as defined in Eq. 6. Under such a design, at the beginning of learning, when the learning status indicator LS_e of the back-end policy is small (close to 0), the behavior of the robot will look slow and sluggish. With the gradual improvement of the policy, when the learning status indicator LS_e becomes large (close to 1), the behavior of the robot will become quick and flexible.

3.4.3 Learning Engagement 2: from disorderly to orderly, the D2O-Mode. The second learning engagement reflects the learning status of robots by adjusting the degree of order/chaos of robots' actions. This design is inspired by the common phenomenon in human learning and especially in children learning, where a disorderly behavior is usually regarded as disengagement while an orderly behavior is often seen as engagement [144, 171]. This kind of phenomenon can be utilized to coincide with the learning progress of the robot. At the beginning of learning, when the entropy of the policy is relatively large, the robot's behavior can be designed to look chaotic and unpredictable. With the gradual improvement of the learning, when the entropy of the policy is becoming relatively small, the robot's behavior can be designed to become regular and predictable. Based on such a design, we can control the configuration of robot joints to realize the transformation from disorderly to orderly. The specific implementation is detailed as follows.

Before controlling the robot's joints to rotate disorderly, we need to get the maximum range (usually determined by the physical constraint) of motion of a robot joint to ensure that the robot does not make incredible movements. For each joint, we first get the upper boundary $bound_U$ and the lower boundary $bound_L$ within the robot's physical limit. At each time step, based on the designed attentional and behavioral engagement, we can get the target configuration of each joint, $config_i^T$. And we can then control the robot's joints to move randomly with the following equation.

$$config_i = \begin{cases} config_i^T + (bound_U - config_i^T) \times R, & if R \geq 0 \\ config_i^T + (config_i^T - bound_L) \times R, & if R < 0 \end{cases} \tag{8}$$

where the $(bound_U - config_i^T)$ and $(config_i^T - bound_L)$ are the feasible moving intervals around the target configuration of each joint $config_i^T$, and R is a random degree controller parameter to make the configuration of the robot move randomly around the target configuration without exceeding the motion limit of the robot joint. We define the random degree controller $R = rand \cdot (1 - LS_e)$, which consists of two parts: one is a random number $rand \in [-1, 1]$ aiming to produce a random and chaotic effect, the other is a degree parameter, which is based on the learning status LS_e of the training algorithm to control the level of randomness of robot's joint motion. If the random degree controller R is greater than or equal to 0, then we let the robot's configuration move randomly from the target configuration to the upper boundary, that is, choose a position randomly between the current target configuration and the upper boundary as the current robot's configuration. Similarly, when the random degree controller R is less than 0, we let the robot randomly move to a position between the target configuration and the lower boundary. We can adjust the random degree through the parameter LS_e. If the current learning status of the robot is good where the LS_e is large (close to 1), the random degree $(1 - LS_e)$ will be low (close to 0), and the action of the robot will be more orderly and predictable. On the other hand, if the current learning status of the robot is poor where the LS_e is small (close to 0), the random degree $(1 - LS_e)$ will be high (close to 1), and the robot's action will be more disorderly and unpredictable.

4 USER STUDY

We conducted a user study to investigate the effects of the proposed *Learning Engagement* designs on users – human instructors in RLfD. Specifically, we focused on two research questions (RQs):

RQ1. How will the proposed *Learning Engagement* designs affect users' mental model of robots' learning progress? Specifically, we want to explore (RQ1.1) users' perceptions of robots' engagement, (RQ1.2) users' perceptions of robots' learning status as well as the correlation between user-perceived robot learning engagement and user-perceived robot learning status, and (RQ1.3) users' expectation of robots' future learning outcomes. Moreover, we want to verify (RQ1.4) the accuracy of users' mental models of robot learning progress.

RQ2. How will the proposed *Learning Engagement* designs affect users' overall perceptions of the robot and the RLfD process? Specifically, on the one hand, we are interested in users' perceptions of the robot in terms of (RQ2.1) the acceptability of its behavior, (RQ2.2) its intelligence and ability, (RQ2.3) its potential to improve further given more demonstrations. On the other hand, we want to gain a better understanding of (RQ2.4) users' perceptions of the quality of their own demonstrations, (RQ2.5) their self-reflection during the teaching process, (RQ2.6) their feeling of tiredness, and (RQ2.7) their willingness to continue teaching the robot. Furthermore, we would like to investigate (RQ2.8) the effect on users' engagement, i.e., the re-demonstration and re-watching behaviors in the teaching process.

4.1 Hypotheses

The key independent variable (IV) in our user study is the engagement mode adopted by robots. We compare the proposed *Learning Engagement*, Z2O-Mode and D2O-Mode, with two baselines – the Full Engagement mode (Full-Mode) and the None Engagement mode (None-Mode). In the Full-Mode, robots always show the highest level of active engagement regardless of their actual learning status. In the None-Mode, robots always remain motionless during humans' teaching processes, which is common in RLfD settings. These two baselines enable us to compare the effects on users' perceptions with/without considering the actual learning status and with/without showing robot engagement.

We proposed a series of hypotheses. Prior studies suggest that non-verbal social signals such as gaze [133] and movement synchrony [99] are effective engagement cues for human learners. We thus hypothesized that:

H1. **(RQ1.1)** Users' perceptions of robot engagement intensity in Z2O-Mode and D2O-Mode will be significantly different from Full-Mode (*H1a*) and None-Mode (*H1b*). And users will perceive significant changes in robot engagement across five teaching rounds in Z2O-Mode and D2O-Mode (*H1c*), but not in Full-Mode and None-Mode (*H1d*).

According to the educational theory that the degree of students' engagement will influence teachers' perceptions of students' learning status and expectations of learning outcomes [18, 181], we raised the hypotheses *H2* and *H3*:

H2. **(RQ1.2)** Users' perceived robot learning progress in Z2O-Mode and D2O-Mode are significantly different from Full-Mode (*H2a*) and None-Mode (*H2b*). Moreover, users will perceive significant changes in robot learning status across the five teaching rounds in Z2O-Mode and D2O-Mode (*H2c*), but not in Full-Mode and None-Mode (*H2d*). Also, users' perceived robot learning status will be significantly correlated with users' perceived robot engagement (*H2e*).

H3. **(RQ1.3)** Users' expectations of robot future learning outcomes in Z2O-Mode and D2O-Mode will be significantly different from Full-Mode (*H3a*) and None-Mode (*H3b*). Further, users' expectations will significantly change across five teaching rounds in Z2O-Mode and D2O-Mode (*H3c*), but not in Full-Mode and None-Mode (*H3d*).

Since we adapted robots' engagement expression to their underlying learning status, we hypothesized that:

H4. **(RQ1.4)** Users' perceptions of robot learning progress in the proposed Z2O-Mode and D2O-Mode will be significantly more accurate than Full-Mode (*H4a*) and None-Mode (*H4b*).

As suggested by [175] that showing engagement will make users hold a more positive attitude toward the robot and their own demonstrations, we proposed the following two hypotheses:

H5. **(RQ2.1-RQ2.3)** Users will have a significantly higher perception of robot behavior acceptability in the proposed Z2O-Mode and D2O-Mode than Full-Mode (*H5a*) and None-Mode (*H5b*). Besides, robots in the proposed Z2O-Mode and D2O-Mode will be perceived as significantly more intelligent than Full-Mode (*H5c*) and None-Mode (*H5d*). Furthermore, if given more demonstrations, robots in the proposed Z2O-Mode and D2O-Mode will be perceived to have significantly greater potential to improve further than in Full-Mode (*H5e*) and None-Mode (*H5f*) if given more demonstrations.

H6. **(RQ2.4)** Users will perceive their demonstrations to have a significantly higher quality when teaching robots in the proposed Z2O-Mode and D2O-Mode than in the Full-Mode (*H6a*) and None-Mode (*H6b*).

It has been shown that self-reflection is a critical aspect in the teaching process [8] which is conducive to the adjustment of the teaching so as to achieve better teaching outcomes [177]. Thus, we hypothesized that:

H7. **(RQ2.5)** When teaching robots in the proposed Z2O- and D2O-Mode conditions, users will reflect significantly more on themselves than when teaching robots in the Full-Mode (*H7a*) and in the None-Mode (*H7b*).

In terms of users' experience in the teaching process, it has been shown that learners giving informative feedback will lead to a positive experience for teachers [25, 45, 94]. We thus hypothesized that:

H8. **(RQ2.6-RQ2.7)** First, users will feel significantly less tired when teaching robots in the Z2O- and D2O-Mode than in the Full-Mode (*H8a*) and in the None-Mode (*H8b*). Moreover, users will be significantly more willing to continue teaching robots in the Z2O- and D2O-Mode conditions than in the Full-Mode (*H8c*) and in the None-Mode (*H8d*).

Moreover, it has been shown that robots showing informative behaviors can increase users' engagement, which can be reflected by the humans' efforts put into the teaching process [100]. We hypothesized that:

H9. **(RQ2.8)** Users' re-demonstration behaviors in the proposed Z2O- and D2O-Mode will be significantly more frequent than in the Full-Mode (*H9a*) and in the None-Mode (*H9b*). Similarly, users' re-watching behaviors in the Z2O- and D2O-Mode will be significantly more frequent than in the (*H9c*) Full-Mode and in the (*H9d*) None-Mode.

4.2 Experimental Setup: Interface and Task Design

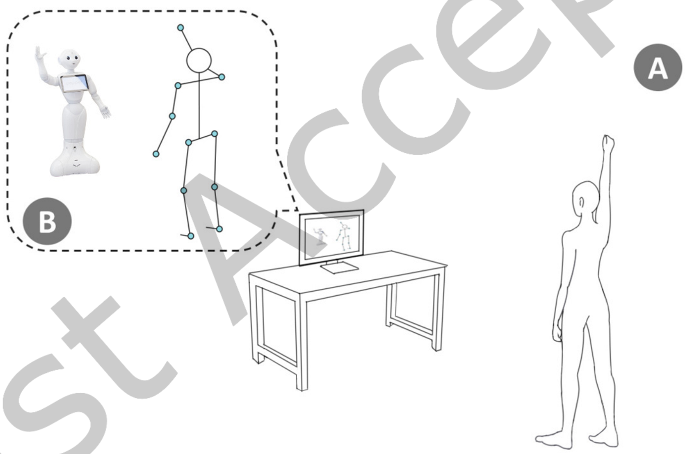


Fig. 7. An illustration of the user study. a) A participant shows her/his demonstration in front of a computer with a webcam. b) A skeleton-based avatar will teach the robot in a simulated environment.

For the user study, we developed an online human-robot teaching platform where participants performed a demonstration in front of a webcam, and the captured motions drove a skeleton-based avatar to replicate the demonstration in front of a virtual robot student in a simulated environment. The user study setup is illustrated in Figure 7. Besides the effects of COVID-19, we chose to carry out the study as an online simulation rather than a physical test due to the following considerations. First, we would like to minimize the possible interference introduced by the technical constraints of physical robots. Pepper, used in our experiment, often moves with undesirable noises, jerks, and vibrations, which could impact human participants' assessment of its ability. Second, Pepper has limited processing power, and thus we would need to send visual-audio signals captured by Pepper to an external server for processing and then send the output back to Pepper for engagement expression, which would lead to an undesirable delay in the engagement expression. Third, we want to mitigate the influence of inconsistent quality of demonstrations performed by different individuals. In a field test, an instructor's real-time demonstrations may contain noises introduced by environmental factors such as the accuracy of motion-capture devices [173], which would consequently affect robots' expressed learning engagement. Experiencing RLfD in simulation, on the contrary, can avoid all these side effects by providing a controllable and measurable environment to monitor and evaluate a robot's performance, which is commonly adopted as a promising alternative to real-world encounters in human-robot interaction studies [38, 175]. Prior studies have verified that a robot's internal status can be effectively conveyed to users via an on-screen display, and users' interpretations of an on-screen robot are consistent with those concerning a physical robot [82, 178, 179, 194].

We were well aware of the possible experience gap between interacting with a robot physically and online. To ensure the face and ecological validity to the best extent, we followed the guidelines proposed in [38] to carefully construct the simulation environment with key structural conditions transposed from the real world, implement the algorithm and design the experiment to avoid introducing unwanted variables and provide multi-perspective views to enhance participants' sense of presence in the simulation. Next, we will introduce these measures in detail.

Simulation Environment Construction. We used Robot Operating System (ROS)[3] for (virtual) robot control, and utilized Unity[4] for simulation rendering. As shown in the online Appendix[2], the simulated RLfD environment illustrates a standard skeleton model (from CMU Mocap[1]) of a human instructor avatar performing a skill in front of a virtual Pepper robot. We utilized the human avatar to transpose participants' physical motion into the virtual space as a reference for the robot to simultaneously express its engagement and to enable participants to embody themselves in the virtual space. We took a skeleton-form avatar to facilitate users' sense of presence in the virtual environment by reducing the mismatch of avatar appearance to theirs (e.g., gender, age, and skin color) and to avoid the uncanny valley effect [119]. In a real-world deployment, the avatar can be controlled by projecting streams of motion data collected from an actual human through motion-capture devices or computer-vision-based sensors onto these joints on the fly. In our experiment, however, to ensure a robust and consistent quality of demonstrations against the possible impact of personal, environmental, and technical factors, we drove the avatar with the same set of pre-recorded demonstrations (detailed below). And we took some measures to ensure participants believed the avatar represented their own teaching.

Skill Selection and Algorithm Implementation. The target scenario of this work is humans teaching robots some daily skills, and the objective is for robots to eventually master these skills by learning from demonstrations. The task selection thus needs to satisfy the following criteria. First, the skills should not be too complicated for laypeople to teach, since it is a robot learning task rather than a human learning task. Otherwise, users' attention may shift from observing the robot's engagement behavior to checking whether their demonstration is successful or not. Second, the skills should be feasible for a humanoid robot – Pepper in this work – to perform, given the degree of freedom of its movable joints. Based on the above two considerations, two authors independently selected skill candidates deemed appropriate from existing Mocap datasets[1] widely used by existing work [135]. They then discussed together and finally chose four skills with potential applications in real-world RLfD tasks for the study: washing a window, playing piano, boxing, and moving a heavy box (the illustration of the four skills can be found in the online Appendix[2]). These skills were sampled at 120 HZ and lasted for 15 to 20 seconds. To control the progress of robot policy derivation in the experiment, we equipped robots with the same learning algorithm trained on the same set of pre-recorded Mocap data of the four skills. We deployed Q-Learning described in Sec. 3.4 as the learning algorithm and used a CPU (Intel i7-10750) for the training. For the actual learning status in

[3]https://www.ros.org/
[4]https://unity.com/

each skill-teaching task, we pre-trained the robot policy and sampled five episodes (i.e., the $10k_{th}$, $30k_{th}$, $50k_{th}$, $70k_{th}$, $90k_{th}$ episode) from the entire learning process to represent different learning stages before the final policy was learned. The learning status indicators (described in Sec. 3.4.1) in the five episodes formed a learning progress curve for each skill. For a fair comparison, we computed the averaged learning indicator value across the four skills at each episode to ensure that robots' learning status in the same learning stage was the same when learning different skills.

Multi-view Video Recording. Then, we recorded the robot's expression of engagement in the four engagement modes in real time responding to the avatar's demonstration. To avoid confusing users, we designed four separate versions of Pepper robots with different names, each equipped with one engagement mode. It thus appears to users that the robot in each mode is learning from scratch. To remove bias caused by the robots' "identity", we named them after four fruits, i.e., Apple, Banana, Orange, and Pear. More importantly, to improve the sense of immersion and facilitate users' observation, we provide recordings from two views: a fixed first-person view and a rotatable full-angel view (as shown in Figure 8(D)). In the first-person view, users would feel that they are watching the robot student face to face, enabling them to inspect the robot's behavior clearly in close proximity. The full-angle view allowed users to observe both the robot's and the instructor's behavior in a teaching environment simultaneously and explore the scene from different angles. In total, we captured 4 modes $*$ 5 iterations $*$ 4 skills $*$ 2 views = 160 videos.

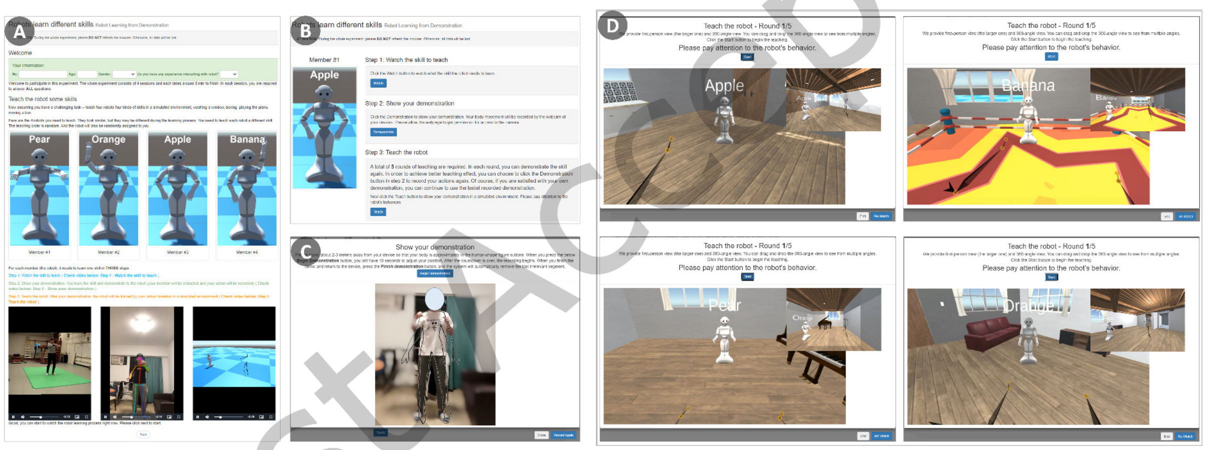


Fig. 8. The screenshots of our user study webpage. a) The introduction page where we introduce the task and procedure of the experiment and get participants familiar with the whole process. b) The main interface of the experiment where participants are required to teach a robot following three steps. c) The demonstration recording interface where participants show their demonstration in front of their device's webcam. d) The interface of teaching a robot in the simulated environment where both first-person view and full-angle view are provided for participants to watch the teaching process.

Experimental Website Development. We established an experimental website to conduct the user study online. Figure 8 shows the screenshots of the web interfaces of our study. The website consists of three key sections. 1) The introduction section introduces the task of the experiment to familiarize participants with the whole process (Figure 8(A)). 2) The main teaching section requires participants to watch a given skill (Figure 8(B)), demonstrate the skill in front of the webcam (Figure 8(C)), and then view an avatar teaching an assigned robot on their behalf in the simulated environment (Figure 8(D)). Through these steps, we hope to give participants the impression that they are actual instructors rather than bystanders. 3) The feedback collection section includes a set of in-task and post-task questionnaires for participants to fill out online. The user study website[5] can be accessed via common browsers.

4.3 Participants

After obtaining institutional IRB approval, we recruited participants through word of mouth and social media posters. A total of 48 participants joined and successfully completed the whole experiment with high quality. We also verified the face validity of our study that all 48 participants considered themselves to be actually teaching the robots. Among them, 20 self-identified as female and 28 as male; the average age was 25.2 (ranging from 22 to 32 years old); 26 were students and 22 were full-time employees. In addition, all participants own a Bachelor's degree or higher with diverse major backgrounds, including computer science, civil, chemistry, math, education, foreign language, etc. And 15 reported having prior experiences interacting with robots, such as cleaning and food delivery robots, in their daily lives. The entire study lasted for about one hour, and they received compensation of USD 10 each.

4.4 Procedure

Figure 9 illustrates the procedure of the user study. We adopted a within-subjects design with the engagement mode as the independent variable. And each mode was assigned a different skill to teach. We counterbalanced the order effect via the Latin Square design, to shuffle the mode-skill assignment and the order of engagement modes. This led to a total of $4 * 4 = 16$ combinations of engagement mode and skill. Each combination was randomly assigned to three participants.

After obtaining informed consent, we first introduced the task of this experiment to participants. Next, we presented portraits of four robots under different names and in different postures to participants (as shown in the top row of Figure 8(A)), to strengthen participants' impressions that these were four independent robots. We told participants that "robots may have different reactions in the learning process, so please pay attention to them". Then we showcased how to teach a robot a skill by taking teaching a driving skill as an example (selected from the same Mocap dataset with the same requirements). More specifically, we provided videos (as shown at the bottom of Figure 8(A)) to help participants understand the three steps to teach a robot. Participants were allowed to watch the example videos and try to demonstrate the skill repeatedly until they could follow the course of action well. After participants confirmed their understanding of each step's task requirements, they proceeded to the main experiment.

Once the main experiment started, participants were assigned four skills to teach in a specified order. Each skill required five rounds of teaching and each round involved three steps. In the first step, participants were asked to watch the Mocap demonstration video of the given skill to learn how to perform it. They were allowed to view it as many times as needed until they could reproduce the skill. We then asked participants a simple attention check question "What is the skill to teach?" to check their understanding and seriousness about the study. They had to choose the correct answer to advance to the next step.

In the second step, participants were asked to grant webcam access to the experimental website. Once done, their body movement would be captured and streamed to a video feed window on the study webpage in real-time. An outline of a translucent human figure was drawn on top of the video images to help participants adjust their distance and position to the webcam (shown in Figure 8(C)), which would keep the most appropriate shooting angle and foreground-background layout. After participants clicked the *Start Recording* button, a 10-second countdown appeared, giving them enough time to get to the filming position indicated by the outline and prompting them that the recording was about to begin. As soon as the countdown ended, participants could start their demonstration. To ensure consistent demonstration quality, we set a fixed length of demonstration

[5]https://userstudy.link/

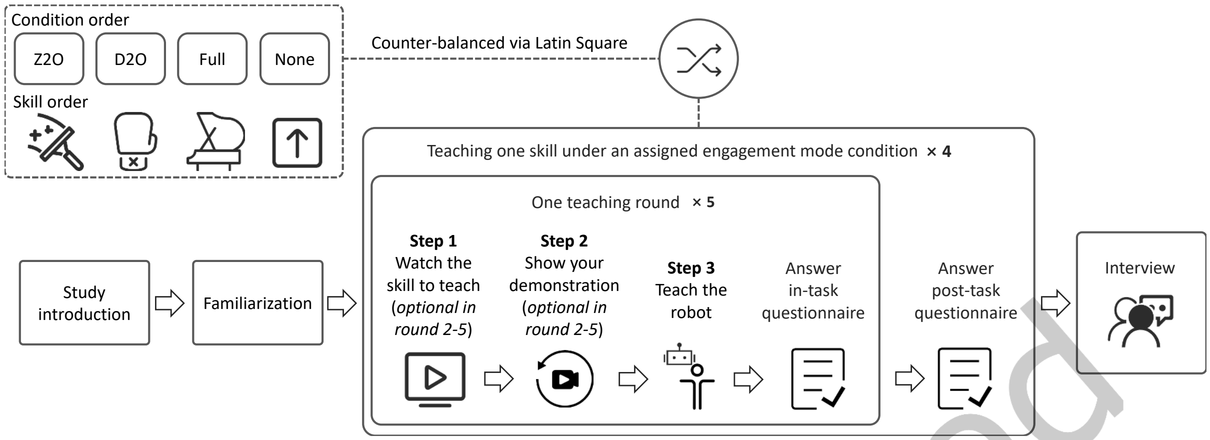


Fig. 9. The procedure of the user study. First, we introduce the task to participants and get them familiar with the teaching process. Then, four skill-teaching tasks are conducted one by one. In each skill-teaching task, participants are asked to go through five teaching rounds step by step (note that in the second to the fifth round, steps 1&2 are optional). Once finishing one teaching round, they are required to answer an in-task questionnaire, and once completing one skill-teaching task, they are required to fill in a post-task questionnaire. After all four skill-teaching tasks are finished, we conduct a final semi-structured interview with participants.

recording for each skill based on the corresponding Mocap data and pilot tests. When the recording was about to end, another 10-second countdown would appear on the interface to prompt people that the recording would automatically terminate in 10 seconds. Usually, participants could complete all the actions before the end of the second countdown; in such cases, they were instructed to return to a natural standing posture and keep it until the recording was done. We told them that an action recognition algorithm running behind the webpage would automatically segment and cut off their standing period after the demonstration. And we allowed participants to record as many times as necessary by clicking the *Record Again* button if they would like to adjust their demonstrations. It should be noted that in this step, we did not actually record participants' demonstrated actions for training the robot; instead, the goal of this step was to give participants a real sense of participation, and thus we simulated the actual recording process in a strict control as required in actual RLfD teaching process.

In the third step, we told participants that the demonstration they had just recorded had been transposed to their avatar in the virtual space and would be replayed in front of a robot student. In this way, they got to watch how the robot student would react to their teaching in real-time through the webpage (as shown in Figure 8(D)). We provided a *Re-Watch* button for them to playback the teaching process as many times as they want. After completing the above three steps of the first teaching round, we asked participants to answer an in-task questionnaire related to the robot's reactions during the teaching process just now. After answering all questions, they could click the *Next-Round of Teaching* button to enter the subsequent round of teaching tasks, if any.

Participants needed to repeat the above three steps in the second to the fifth teaching round, but the first and the second steps were both optional. We told participants that if they wanted to review the standard demonstration of the skill, they could go through the first step again. Besides, if participants would like to improve their demonstration for any reason, they could go through the second step and record their demonstration again. Otherwise, they could start a new teaching round directly from the third step. In this case, participants were informed that the latest demonstration (from the previous teaching round) would be presented to the robot by the avatar. After the completion of each teaching round, the same in-task questionnaire would be given to participants.

Upon completing all five rounds of teaching, we asked participants to fill out a post-task questionnaire consisting of questions related to their perceptions of the RLfD process. And all four skill-teaching tasks followed the same procedure and protocol as described above. Finally, at the end of the entire study, we conducted a semi-structured interview and collected participants' feedback on their feelings and thoughts during the experiment.

The whole study was conducted online, and one of the authors who served as the experimenter accompanied participants throughout the process via Zoom – a video conferencing software, providing non-interference operation guidance when necessary. Due to the complexity of the experiment, we allowed participants to ask the experimenter questions whenever needed and encouraged them to express their ideas verbally at any time during the experiment. Participants agreed to share their screens with us on Zoom, and with their permission, we recorded the entire study.

We took two measures to verify the face validity of our experiment design. First, during the study, the experimenter carefully observed the user's reaction to determine whether the user realized or suspected that this was a pre-recording. If the user showed any sign of doubt, we would exclude the data from this user. Second, since we had recorded the whole process of the experiment with the consent of the user, two authors carefully reviewed all the feedback from the user during the experiment, including the thinking aloud and all the recordings of the experiment process, and then strictly judged whether the user perceived that it was pre-recorded. As long as one of the two authors perceived that the user had doubts, it was judged that the user did not pass the verification.

4.5 Measurements

We measured participants' perceptions of robots' reactions in each round of teaching and their overall perceptions and experiences of the teaching process by the in-task and post-task questionnaires respectively, shown in Table 1. Through the questions, we can directly access participants' perceptions to investigate a set of hypotheses (***H1-H3, H5-H8***). Besides, we can explore further findings from their ratings. To verify ***H4***, we measured the accuracy of users' mental model of robots' learning progress by calculating the similarity between participants' perceived robot learning progress and the actual robot learning progress via the Mean Square Error (MSE) measure. In addition, we recorded the number and timestamps of click events of the two buttons *Record Again* and *Re-Watch*, which can provide more objective information about participants' behavior during the teaching process and help us investigate the hypothesis ***H9***. Furthermore, transcriptions of participants' verbal feedback – thinking aloud [71] throughout the study and of the final exit interviews could provide more in-depth insights. Specifically, in the semi-structured interview, we first invited participants to provide the reasons behind their answers to each in-task and post-task question. Then, we collected their opinions to some other open-ended questions, such as i) "Which robot do you want to teach most, why?", ii) "What's your feeling when teaching the [name] robot?", iii) "How do you feel after completing the task of teaching the four robots?", etc. Participants' feedback helped us comprehensively understand their perceptions and further contributed to valuable design implications.

5 RESULT

In this section, we present results from the user study according to our two research questions. As a Shapiro-Wilk test did not show evidence that the data fits the normality assumption, we ran non-parametric statistical tests. In general, we ran Friedman Test with post-hoc Wilcoxon signed rank tests on subjective ratings of 7-point Likert questions. In addition, we carried out thematic analysis [14] to qualitatively explore participants' perceptions from the verbal feedback during the study and in the interview. One author transcribed participants' in-task verbal recordings and post-task interview recordings into words. After another author checked the correctness,

Table 1. Purposes (related hypotheses), questions, and answer options of the in-task and post-task questionnaires.

Purpose	Question	Answer option
	In-task Questionnaire	
Manipulation Check	**Q1**. Do you think the robot's eye gaze was following your demonstration?	(7: Strongly agree, 1: Strongly disagree)
	Q2. Do you think the robot's upper limb was following the rhythm of your demonstration?	(7: Strongly agree, 1: Strongly disagree)
H1	**Q3**. How do you feel that the robot is engaged in the teaching process?	(7: Strongly engaged, 1: Strongly disengaged)
H2	**Q4**. To what extent do you think the robot has mastered the skill? Please give a score to the robot (7 points: fully master the skill).	(7: 6-7 point, 1: 0-1 point)
H3	**Q5**. What do you think is the likelihood that the robot will master the skill in the future?	(7: Very likely, 1: Very unlikely)
	Post-task Questionnaire	
H5	**Q1**. Do you agree that the robot's behaviors in the learning process are acceptable/reasonable?	(7: Strongly agree, 1: Strongly disagree)
	Q2. Do you agree that the robot is intelligent and has great learning ability?	(7: Strongly agree, 1: Strongly disagree)
	Q3. Do you agree that if given further demonstration, the robot will learn much better?	(7: Strongly agree, 1: Strongly disagree)
H6	**Q4**. Do you agree that your demonstration is appropriate for robots to learn based on the robot's reactions?	(7: Strongly agree, 1: Strongly disagree)
H7	**Q5**. Have you ever reflected on the quality of your demonstration or whether your demonstration works? Please rate how much you reflect on yourself.	(7: Reflected a lot, 1: Did not reflect at all)
H8	**Q6**. Do you feel tired when teaching the robot?	(7: Strongly agree, 1: Strongly disagree)
	Q7. Do you want to continue to teach such a robot?	(7: Strongly agree, 1: Strongly disagree)
	Q8. Do you have any other thoughts? Please feel free to write it down.	Any comment

the two authors independently open-coded the transcriptions and resolved the conflicts together. In the rest of this section, based on the proposed research questions and hypotheses, we present the results accordingly.

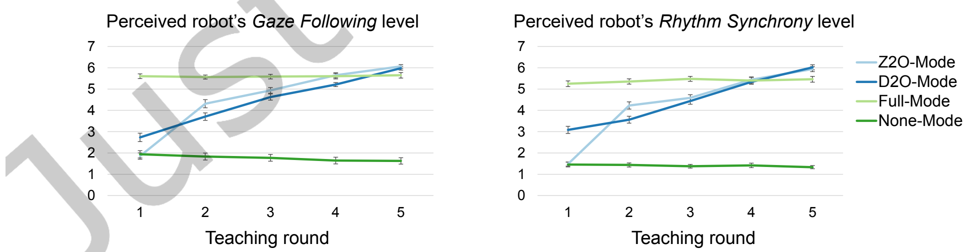


Fig. 10. Participants' perceptions of correspondence level of robots' *gaze following* and *rhythm synchrony* with the given demonstration (the error bar represents the standard error).

Before further exploring, we first verify the effectiveness of our manipulations on robots' learning engagement expression. They are measured by **Q1** and **Q2** of the in-task questionnaire, as shown in Figure 10. Friedman test results show significant differences among the four engagement modes in terms of participants' perceived robot's attentional engagement level (i.e., *Gaze Following* level, $\chi^2(3)$=430.187, p<0.001) and behavioral engagement level (i.e., *Rhythm Synchrony* level, $\chi^2(3)$=451.430, p<0.001). Moreover, participants' perceived attentional engagement levels among five teaching rounds are significantly different in *Z2O-Mode* ($\chi^2(4)$=154.378, p<0.001) and *D2O-Mode* ($\chi^2(4)$=150.308, p<0.001); so are their perceived behavioral engagement levels (in *Z2O-Mode*, $\chi^2(4)$=160.402, p<0.001; in *D2O-Mode*, $\chi^2(4)$=140.113, p<0.001). By contrast, for *Full-Mode* and *None-Mode*, there are no significant differences among the five teaching rounds in either manipulation check measure. These results meet our expectations since we did not alter the robots' attentional and behavioral expressions in different teaching rounds in these two engagement modes. Based on the above results, we can conclude that our manipulations are effective.

5.1 How Will the Proposed *Learning Engagement* Affect Users' Mental Model of Robots' Learning Progress? (RQ1)

To answer RQ1, we gather participants' ratings on their perceptions of robots' engagement level, perceptions of robots' learning status, and expectations of robots' future learning outcomes after each round of teaching through an in-task questionnaire (results are shown in Figure 11). Overall, we found that users held significantly different perceptions of robots' learning process, which is mainly caused by the distinct engagement the robots expressed.

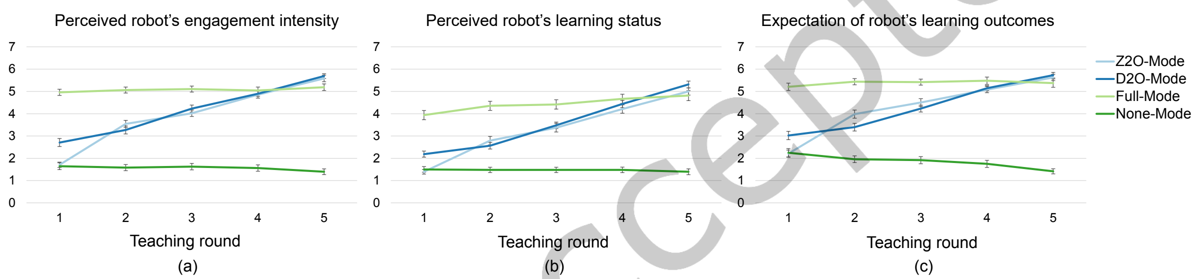


Fig. 11. Participants' ratings of the questions in the in-task questionnaire (the error bar represents the standard error). a) Participants' perceptions of robots' engagement intensity in the teaching process. b) Participants' perceptions of robots' learning status in the teaching process. c) Participants' expectations of robots' future learning outcomes in the teaching process.

5.1.1 *Effects on Users' Perception of Robots' Engagement in the Teaching Process.*

Friedman test shows significant differences among the four engagement modes, $\chi^2(3)$=431.638, p<0.001 (Figure 11(a)). Specifically, users' perceived robot engagement in Z2O-Mode and D2O-Mode are significantly different from Full-Mode (Z=-8.892, p<0.001; Z=-7.769, p<0.001) and None-Mode (Z=-12.132, p<0.001; Z=-12.446, p<0.001), thus ***H1a*** and ***H1b*** are both accepted. And participants' ratings of robot engagement in the different teaching rounds vary significantly in Z2O-Mode ($\chi^2(4)$=169.302, p<0.001) and D2O-Mode ($\chi^2(4)$=148.353, p<0.001), ***H1c*** accepted, but not in Full-Mode or None-Mode (***H1d*** accepted).

In general, participants indeed relied on the dynamics of a robot's gaze and/or motion synchronicity to discern whether and the extent to which it was engaged with their demonstrations. Specifically, in Z2O-Mode, all participants perceived the robot to be increasingly engaged mainly because of the robot's gradually focused gaze and active following. For example, P2 (F, 26) said *"At first, the robot's eye-gaze was static and it did not look at my motions, but gradually it began to move and focus on my demonstration."* In D2O-Mode, 45 out of 48 participants raised their assessment of the robot's engagement due to the robot's increasingly orderly gaze and behaviors. For example, P38 (M, 25) said *"Gradually, the robot's eyes were no longer looking around, and its arms' movement became regular, which showed it could participate in my teaching more and more actively."* While in Full-Mode, the reason why most (38/48) participants consistently gave high ratings was the robot's consistently high intensity of gaze and behaviors. An opposite effect was observed in None-Mode, where all participants gave consistently low ratings of robots' engagement. 33 out of 48 participants thought robots' engagement was low because the robot *"did not show any movement during the teaching process"*, and 11 participants even doubted *"the robot was broken"*.

This finding reveals that users will naturally transfer social norms of human-human interaction into the process of perceiving AI systems without any additional guidance and learning [116, 124]. This also illustrates the feasibility of expressing the engagement of AI systems by drawing on common behaviors and habits in human-human interactions. Future HCI research can also draw on the interaction behavior model between people to a certain extent to design the behavior expression of the corresponding AI system [20].

5.1.2 Effects on Users' Perception of Robots' Learning Status in the Teaching Process. Friedman test shows significant differences in different engagement modes, $\chi^2(3)$=398.585, p<0.001 (Figure 11(b)). Specifically, users' perceived robot learning status in Z2O-Mode and D2O-Mode are significantly different from Full-Mode (Z=-7.969, p<0.001; Z=-6.410, p<0.001) and None-Mode (Z=-11.421, p<0.001; Z=-12.051, p<0.001) (***H2a*** and ***H2b*** accepted). And participants' ratings in different teaching rounds vary significantly in Z2O-Mode ($\chi^2(4)$=167.008, p<0.001) and D2O-Mode ($\chi^2(4)$=161.979, p<0.001) (***H2c*** accepted), also in Full-Mode ($\chi^2(4)$=54.958, p<0.001) but not in None-Mode (***H2d*** partially accepted).

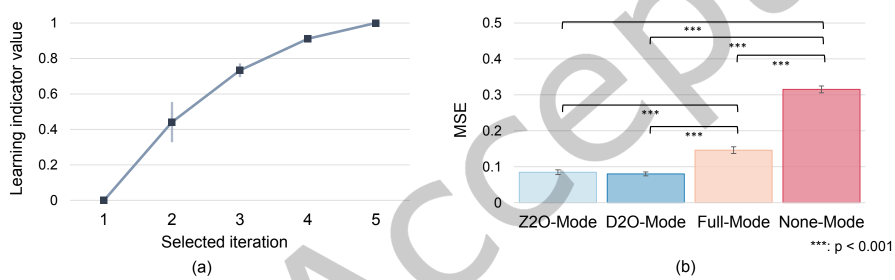


Fig. 12. The accuracy of participants' perceptions of robots' learning progress. (a) The averaged learning status indicators when robots learn the four skills, which are sampled in five different training iterations and serve as the ground truth. (b) The Mean Square Error (MSE) score of comparing participants' perceived robots' learning progress with robots' actual learning progress when teaching robots in different engagement modes (the error bar indicates mean standard error).

Qualitative analysis shows that the increased engagement intensity in Z2O-Mode and the reduced engagement randomness in D2O-Mode led most participants (46/48) to improve ratings on the robot's learning status. For example, in Z2O-Mode, P26 (F, 24) said *"As the swing amplitude of the robot's arm becomes increasingly obvious, I felt it was more and more confident."* Similarly, in D2O-Mode, P27 (M, 25) mentioned *"The robot did not seem to pay attention to my teaching at the beginning, but it was more and more involved in learning. I think it must have mastered the knack of the skill"*. It is interesting that in Full-Mode, although the robot-expressed engagement does not change, there is still an upward trend in humans' perceptions of robot learning status although the trend is not as steep as in Z2O- and D2O-Mode. It turns out that the positive feedback from the robot makes 15 out of 48 participants think it is likely to make progress. For example, P30 (M, 24) said *"This robot was able to follow me in the beginning. I thought such a serious robot should make some progress."*. In contrast, in the None-Mode, 47 out of 48 participants did not consider the robot had made significant progress over five teaching rounds because the robot *"did not engage in my teaching at all"*.

We further analyze the correlation between participants' perceived robots' engagement and robots' learning status using the Spearman correlation. The overall coefficient value across the four engagement modes is 0.890,

p < 0.001 (***H3c*** accepted). Within each engagement mode, the correlation is also significant (Z2O-Mode: 0.903; D2O-Mode: 0.872; Full-Mode: 0.677; None-Mode: 0.628; all with p < 0.001). Both the qualitative feedback and quantitative results show that users' perceptions of robots' learning status are highly correlated with their perceptions of robots' engagement.

This finding reveals the effectiveness of implicitly conveying the internal state of AI systems by controlling the engagement expression. Our results suggest that users' perceptions of AI systems' internal status will be affected by their observation of the behaviors of AI systems, which is also consistent with users' *mental model* theory [128]. It is worth mentioning that the learning statuses behind the four modes of robots in our study are actually the same, but they are perceived to be different due to the different engagement expressions. Therefore, we should carefully design the expression of AI systems' behaviors to help people establish an accurate mental model of the AI system. Otherwise, a poor design that causes a disparity between user perception and the actual state of the AI system will affect people's subsequent behavior and decisions, which may further impair the outcome of subsequent tasks.

5.1.3 Effects on Users' Expectation of Robots' Future Learning Outcomes in the Teaching Process. Friedman test shows significant differences in different engagement modes, $\chi^2(3)$=418.758, p<0.001 (Figure 11(c)). Specifically, participants' expectations of robots' future learning outcomes in Z2O-Mode and D2O-Mode are significantly different from Full-Mode (Z=-8.109, p<0.001; Z=-8.273, p<0.001) and None-Mode (Z=-11.893, p<0.001; Z=-11.947, p<0.001) (***H3a*** and ***H3b*** accepted). And participants' ratings in the different teaching rounds vary significantly in Z2O-Mode ($\chi^2(4)$=151.698, p<0.001) and D2O-Mode ($\chi^2(4)$=-135.447, p<0.001) (***H3c*** accepted), and a significant drop is observed in None-Mode ($\chi^2(4)$=46.747, p<0.001) but no difference is found in Full-Mode, thus ***H3d*** is partially accepted.

Participants' qualitative feedback shows that robots' increased engagement strengthened participants' confidence in robots' future learning outcomes. Specifically, in Z2O-Mode and D2O-Mode, 46 out of 48 participants raised their expectations based on the robots' progress. For example, P11 (M, 25) mentioned *"From acting slow at the beginning to gradually being able to follow my demonstration, the robot kept refreshing and raising my expectation for it."* While in Full-Mode, 42 out of 48 participants' expectations were not raised because the robot *"seemed to be the same as it was at the last learning round"*. To our surprise, there was a significant decreasing trend in None-Mode. 20 out of 48 participants were *"discouraged"* by the lack of improvement in robots' reaction to the demonstration, and they *"gradually stop expecting more progress from it"*. For example, P4 (F, 25) mentioned *"At first, although it did not give me any feedback, I thought the skill was not difficult for the robot to learn, so I rated 3 (Slightly unlikely). But to my disappointment, the robot still remained motionless in the remaining rounds, so I gradually lowered my expectation of it"*.

This finding implies that an AI system's behavior (e.g, engagement expression) greatly affects users' expectations of it. Users' expectation of AI systems plays a very important role in human-AI interaction/collaboration, influencing different aspects of user experience and long-term use of interactive systems [80, 87]. Appropriate expectations of an AI system (i.e., the expectation matches AI's actual potential/capability) can help people properly divide tasks, and correctly decide whether, when, and to what extent to refer to the AI's suggestions [130]. On the contrary, inaccurate expectations will lead to disuse (underestimating the capabilities of AI) [98] or abuse (over-dependence on AI) of the AI system [130], which can cause negative interaction experiences and poor task outcomes [79, 151]. Therefore, designers must carefully design the behavior of the AI system to help people calibrate their expectations [199]. In addition, our findings suggest that users' expectations for the AI system can be updated continuously throughout the interaction process. Therefore, designers can assist users in updating their expectations of the AI system by dynamically adjusting the information communicated by the system.

5.1.4 Accuracy of Users' Mental Model of Robots' Learning Progress. To investigate whether users' mental model of robots' learning progress is accurate, we take robots' real learning progress as the ground truth (denoted by the *Learning Status Indicator* of the backend algorithms (detailed in Sec. 3.4.1)), ranging from 0 to 1, the higher, the better state (shown in Figure 12 (a)). Since we use a 7-point Likert scale (**Q4**) to infer participants' mental model, for easier comparison with the ground truth, we normalize the collected ratings by dividing them by 7. Then, we calculate the Mean Square Error (MSE), a common metric in regression, to measure the distance between the perceived learning progress (i.e., normalized subjective ratings) and the actual progress (i.e., *Learning Status Indicator*). Figure 12 (b) shows the results (the lower the better) in different engagement modes. Wilcoxon test results indicate that the MSE scores in Z2O-Mode and D2O-Mode are significantly smaller than that in Full-Mode (Z=-4.226, p<0.001; Z=-4.769, p<0.001) and None-Mode (Z=-6.031, p<0.001; Z=-6.031, p<0.001), thus ***H4a*** and ***H4b*** are accepted.

This result shows that it is feasible to convey a robot's internal learning status through appropriate engagement cues in the process of teaching the robot. This approach of encoding the internal state of an interactive system into its expressed engagement is expected to inspire novel human-AI communication design in broader scenarios.

5.2 How Will the Proposed *Learning Engagement* Affect Users' Overall Perceptions of the Robot and the RLfD Process? (RQ2)

To answer RQ2, we gather participants' ratings on their perceptions of the overall teaching process in the post-task questionnaire (results are shown in Figure 13). Generally, robots' communication of different engagement could lead to users' significantly different perceptions of robots and the teaching process.

5.2.1 Effects on Users' Perception of the Acceptability of Robots' Behavior. Overall, Friedman test shows significant differences among four engagement modes ($\chi^2(3)$=92.557, p<0.001). Specifically, there is no significant difference between Z2O-Mode and Full-Mode, or between D2O-Mode and Full-Mode, ***H5a*** rejected. The acceptability of robot behavior is significantly higher in Z2O-Mode and D2O-Mode than None-Mode (Z=-5.886, p<0.001; Z=-6.021, p<.001), ***H5b*** accepted. Significant differences also exist between Z2O-Mode and D2O-Mode (Z=-2.000, $p < 0.05$), between Full-Mode and None-Mode (Z=-5.874, $p < 0.001$).

Participants generally approved of robots' behaviors in Z2O-Mode, D2O-Mode, and Full-Mode because they could receive meaningful feedback from robot learners. Specifically, in Z2O-Mode, all participants thought the robot's increased engagement was positive and reasonable. For example, P19 (M, 23) said *"With more learning rounds, the robot's gaze and arm movements followed my demonstration actions more and more actively, which look as understandable as a human student."* And in D2O-Mode, 43 out of 48 participants agree with the robot's being gradually orderly. For instance, P33 (F, 29) mentioned *"The robot might not understand my action at the beginning, so it did not pay attention to me, but it gradually followed my motion and learned with me, and I slowly drew its attention back."* However, 5 participants did not appreciate the disorderly behavior of the robot, and thought the robot was not polite. For example, P11 (M, 25) said *"Even if the robot couldn't understand my teaching, it shouldn't show chaotic movements."* In Full-Mode, the robot's high engagement was welcomed by 39 participants. For example, P20 (F, 25) said *"The robot was very serious from the beginning and had been learning from me attentively until the end, just like a hard-working human student"*. But 13 participants felt that the robot's response was a little unusual, as the robot never showed a noticeable change. For None-Mode, nearly all participants think the robot's behavior is incomprehensible and unsatisfactory. For example, P8 (M, 24) said, *"How could a robot not move at all? I was teaching it seriously. At least it should show some reaction."*

Such feedback reflects that users tend to apply social heuristics and norms to judge the behavior of the humanoid robot used in the experiment, which is consistent with the theory of *computers are social actors (CASA)* [124]. This also reminds us to fully consider the possible impact of AI system behavior on people's perception and acceptance. Designers can refer to some common behavioral laws in humans' social interaction and principles in the process of human-human interaction [116] to make the representation of AI system status match people's established way of interpreting communicative cues.

5.2.2 Effects on Users' Perception of Robots' Intelligence and Learning Ability. Friedman test result shows significant differences in four engagement modes ($\chi^2(3)$=101.847, p<0.001). Specifically, participants considered robots to be significantly more intelligent and capable of acquiring new things in Z2O-Mode and D2O-Mode compared with Full-Mode (Z=-2.611, p<0.01; Z=-4.212, p<0.001; ***H5c*** accepted) and None-Mode (Z=-6.004, p<0.001; Z=-5.957, p<0.001; ***H5d*** accepted). And there is no significant difference between Z2O-Mode and D2O-Mode, while robots' intelligence is perceived significantly higher in Full-Mode than in None-Mode (Z=-5.901, p<0.001).

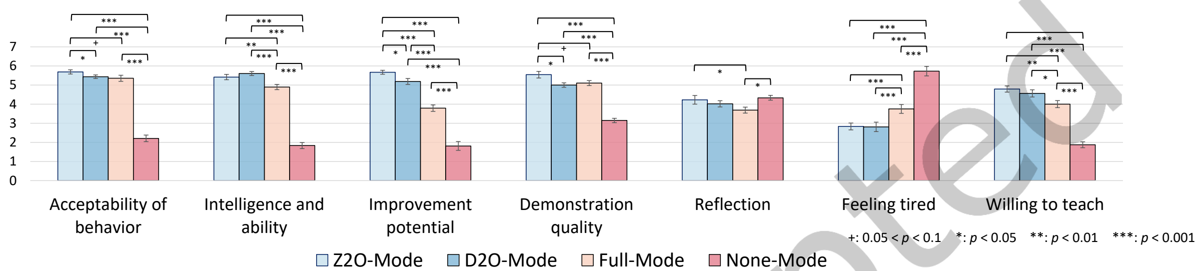


Fig. 13. Means and standard errors of participants' ratings of the questions (7-point Likert scale) in the post-task questionnaire.

Qualitative analysis suggests that all participants think that students who can show instant feedback especially adaptive feedback are more intelligent. For example, in Z2O-Mode, P19 (M, 23) said *"The robot could gradually learn from my demonstration, which showed that this robot had strong learning ability"* And in D2O-Mode, P22 (M, 27) mentioned *"This robot could gradually catch the rhythm of my movements, just like a smart but naughty child.'* As for Full-Mode, although the robot was highly engaged, its invariable feedback caused participants to lower their ratings. For example, P25 (M, 24) mentioned *"This robot's behaviors did not make much progress. Maybe its learning ability is not as high as the other two robots (in Z2O-Mode and D2O-Mode)"* And almost all participants thought the robot in None-Mode had little learning ability as it *"showed no obvious feedback"* and *"looked like a sculpture"*. It can be concluded that robots' positive reactions and adaptation to teachers' demonstrations are keys to people's perception of robots' learning capacity.

This finding reveals that first, users will have a positive attitude toward the ability of an AI system that can give feedback. Second, users have more positive perceptions of expressive behaviors that adapt dynamically to their interactions than fixed ones. Note that sometimes, it is not necessarily appropriate for an AI system to change its behavior just to seem smart, as it may result in an overestimation of the AI system [78, 199]. Instead, the design of AI system behavior should match its actual intelligence and capabilities to not mislead users.

5.2.3 Effects on Users' Perception of Robots' Potential of Further Improvement Given More Demonstration. Overall, Friedman test result shows significant differences among the four engagement modes ($\chi^2(3)$=115.025, p<0.001). Specifically, participants rated robots in Z2O-Mode and D2O-Mode to have significantly higher potentials than Full-Mode (Z=-5.404, p<0.001; Z=-5.243, p<0.001; ***H5e*** accepted) and None-Mode (Z=-6.000, p<0.001; Z=-6.012, p<0.001; ***H5f*** accepted). Furthermore, robots in Full-Mode were rated significantly higher than None-Mode (Z=-5.859, p<0.001), and the potential of robots in Z2O-Mode significantly exceeded those in D2O-Mode (Z=-2.593, p<0.05).

From qualitative feedback, most participants thought robots in Z2O-Mode and D2O-Mode had the potential for further improvement based on robots' increasingly higher engagement. For example, P33 mentioned *"The* *robot was gradually getting into a good engagement state, and I thought it was not far away from mastering the skill"*. However, as for the comparison between Z2O-Mode and D2O-Mode, 18 participants thought the robot [in D2O-Mode] was *"like a careless and naughty kid"* so *"it might not achieve as big progress as the serious one [in Z2O-Mode]"*. In Full-Mode, most participants held low expectations due to robots' little improvement in the past teaching rounds. As for None-Mode, 47 out of 48 participants gave negative feedback due to the lack of engagement. For example, P46 (M, 25) said *"It [the robot] can never master the skills I teach. No matter how many times you teach it, you will not be able to let a child who does not want to learn from you learn better."* The feedback is aligned with the motivation principle in education [167].

This result also shows that the engagement expression of AI systems not only affects users' perception of the current state but also affects users' perception of the potential of AI systems, such as the prediction of the performance of machine learning algorithms in machine teaching or interactive machine learning [64, 65, 190]. We argue that users' perceptions of the AI system's potential will affect the goal setting and the arrangement of future interactive tasks. Therefore, when designing the behavior of AI systems, we should consider the nuanced impact on users, so as to avoid users' wrong predictions of the potential and future outcomes of AI systems.

5.2.4 Effects on Users' Assessment of Demonstration Appropriateness. Friedman test shows significant differences among four engagement modes ($\chi^2(3)$=54.052, p<0.001). There are no significant differences comparing Z2O-Mode and D2O-Mode to Full-Mode; thus ***H6a*** rejected. But participants rated their teaching to be significantly more appropriate in Z2O-Mode and D2O-Mode than None-Mode (Z=-5.153, p<0.001; Z=-4.798, p<0.001; ***H6b*** accepted). We also find significantly higher ratings in Z2O-Mode than D2O-Mode (Z=-2.574, p<0.05), and higher in Full-Mode than None-Mode (Z=-5.003, p<0.001). Note that in different engagement modes, the demonstration performed in the simulated environment is from the same set of MoCap data with the same quality. Besides, we observed that almost all participants were able to complete the demonstration in front of the webcam with high quality and completeness. Nevertheless, participants' assessments of their own demonstrations were still significantly influenced by the robots' communication of engagement.

Most participants highly perceived their demonstration quality in Z2O-Mode, D2O-Mode, and Full-Mode because of the robots' high or increasingly high engagement. For instance, participants recognized their demonstration as it made the robot (in Z2O-Mode) *"more and more engaged in demonstrations"* and it made the robot (in D2O-Mode) *"gradually find the key rhythm of the skill"* and it *"caught the robot's (in Full-Mode) attention from the very beginning"*. While for None-Mode, 38 participants were not confident about their demonstration due to the robots' little engagement. For example, P20 (F, 25) said *"If it was not the problem of the robot itself, it must be that my demonstration was not good enough."* While 10 participants did not doubt their demonstration even the robot did not show any engagement. Overall, users seemed to mindlessly attribute robot students' feedback to their teaching strategies as in human-human teaching [158].

This result reflects that users' perception of their own performance can largely depend on the feedback given by the AI partner. Positive feedback from the AI system can lead to users' positive self-evaluation, while negative feedback may cause users to lower their self-evaluation, impair their confidence, and ultimately compromise task outcomes. Therefore, the feedback of the AI system should be carefully designed according to specific goals. For example, if we do not want to cause users' negative self-evaluation, we should pay attention to avoid introducing negative behaviors into the AI system's feedback. Conversely, in certain circumstances, if the AI system needs to raise users' awareness of their own poor performance, appropriate negative feedback can be leveraged to prompt users to make timely adjustments.

5.2.5 Effects on Users' Self-reflection on Teaching. Friedman test result suggests there are significant differences across four engagement modes ($\chi^2(3)$=9.519, p<0.05). Specifically, a significant difference can be found between Z2O-Mode and Full-Mode (Z=-2.404, p<0.05), but not between D2O-Mode and Full-Mode, ***H7a*** is partially accepted. Similarly, we do not observe any significant difference between Z2O-Mode and None-Mode, nor between D2O-Mode and None-Mode, ***H7b*** rejected. Also, the disparity between Z2O-Mode and D2O-Mode has no statistical significance. However, participants reflect themselves significantly less in Full-Mode than in None-Mode (Z=-2.266, p<0.05).

We originally hypothesized that participants would reflect more upon their demonstrations when teaching robots in Z2O-Mode and D2O-Mode. But it turns out that participants' self-reflection levels are all moderate in the four engagement modes. Generally, participants held moderate self-reflection in Z2O-Mode and D2O-Mode because the robots *"were more and more on track in learning"* thus they *"did not have to reflect themselves too much."* Interestingly, most participants had the least reflection in Full-Mode, as the robot expressed positive feedback from the beginning. For example, P29 (F, 25) mentioned *"The robot could follow the demonstration immediately in the first round, so I thought my teaching had no problem."* On the contrary, participants in None-Mode had the highest self-reflection because the robot surprisingly did not show any feedback. For example, P28 (M, 22) said *"I couldn't help suspecting that there was something wrong with my demonstration, because the robot didn't interact with me at all"*. In summary, human teachers tend to reflect on their teaching deeper if the robot student does not show the expected behavior.

This finding shows that properly expressing an AI system's inability (i.e., what and when the AI system cannot do [92, 142]) rather than blindly catering to users' expectations can promote users' critical thinking. Especially in the human-AI collaboration scenario [130], the final task outcomes depend on the performance of both team members. If the AI system always shows positive feedback, it is likely that users will reflect less on themselves and thus cannot make timely adjustments to achieve optimal task outcomes. Designers should make good use of the critical thinking brought about by the contrast with users' original expectations, and express the shortcomings of the AI system at an appropriate time to evoke people's reflection and agency.

5.2.6 Effects on Users' Experience in the Teaching Process. Tiredness is a common measurement of instructors' teaching experience [52]. Friedman test result suggests that different engagement modes lead to significantly different levels of fatigue in teaching ($\chi^2(3)$=89.594, p<0.001). In particular, robots in the Z2O-Mode and D2O-Mode induce significantly lower sense of tiredness than Full-Mode (Z=-3.470, p<0.001; Z=-3.706, p<0.001) and None-Mode (Z=-5.899, p<0.001; Z=-5.911, p<0.001) (***H8a*** and ***H8b*** accepted). Moreover, participants felt significantly more tired in None-Mode than Full-Mode (Z=-5.718, p<0.001), but no significant difference between Z2O-Mode and D2O-Mode.

In addition, the willingness to teach a robot in the future is also a critical indicator of user experience [92]. Friedman results show that there are significant differences among the four engagement modes ($\chi^2(3)$=96.123, p<0.001). Specifically, participants expressed significantly stronger willingness to continue to teach in Z2O-Mode and D2O-Mode than in Full-Mode (Z=-2.986, p<0.01; Z=-2.303, p<0.05) and None-Mode (Z=-5.966, p<0.001; Z=-6.026, p<0.001) (***H8c H8d*** both accepted). Besides, participants are significantly more inclined to teach robots further in Full-Mode than in None-Mode, Z=-5.780, p<0.001; but no significant difference can be observed between Z2O-Mode and D2O-Mode.

Qualitative findings show that the reason why participants felt less tired and more willing to teach robots in Z2O-Mode and D2O-Mode is that most participants (40/48) got positive emotions during teaching and they prefer to teach students with progressive feedback rather than those without feedback or with unchanged feedback. For example, P11 (M, 25) in Z2O-Mode said *"The robot made me very excited to see that it little by little followed my demonstration."* And P14 (F, 32) in D2O-Mode mentioned *"I felt a great sense of achievement. I slowly drew an absent-minded student back to the learning state through my hard teaching."* However, if the robot showed non-adaptive engagement, 25 participants *"lacked willingness"* For example, in Full-Mode, P47 (F, 25) said *"This robot was quite hard-working. But as it had never shown any progress, I prefer to teach the other two robots [in Z2O-Mode and D2O-Mode]"*. As the robot in None-Mode did not show any engagement, participants *"felt like teaching in vain"* and *"did not want to teach this disengaged student any longer"*. For example (F, 27) mentioned

"The most tiring experience was teaching Orange [the robot in None-Mode] who didn't give me any feedback, which made me exhausted". To summarize, whether robot students could give feedback and/or whether the feedback is appropriate are crucial to a teacher's teaching experience.

This finding stresses that when designing human-AI interactions, on the one hand, it is not just about completing tasks as the design goal, but also enhancing the experience of the user interaction process as an important design consideration. Especially when user input takes a long time, the AI system should provide timely feedback to avoid user fatigue or other negative emotions due to lack of response. On the other hand, if the feedback provided remains constant over a period of time, it will also degrade the user's interaction experience with the AI system.

5.3 Frequency, Timing, and Motivation of Re-Demonstration and Re-Watching

As mentioned in Sec. 4.5, we collected users' button-clicking behaviors of *Record Again* (means re-demonstration) and *Re-Watch* buttons. We believe these click logs can, to some extent, objectively reflect the real perceptions and experiences of users in the teaching process. We analyze the frequency, timing, and motivation of these two kinds of button-click events.

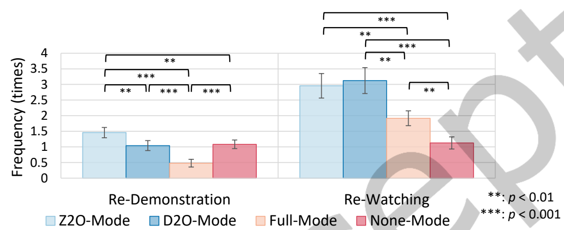


Fig. 14. Means and standard errors of the frequency of Re-Demonstration and Re-Watching in the four engagement modes.

5.3.1 Frequency of Re-Demonstration and Re-Watching. The re-demonstration and re-watching behaviors can be considered as participants' active engagement in the teaching process (results are shown in Figure 14).

Frequency of Re-demonstration Behavior. Friedman test results show significant differences among the four engagement modes ($\chi^2(3)$=34.921, p<0.001). Specifically, the frequencies in Z2O-Mode and D2O-Mode are significantly higher than Full-Mode (Z=-4.483, p<0.001; Z=-3.659, p<0.001), thus ***H9a*** is accepted. A significant difference can be found between Z2O-Mode and None-Mode (Z=-2.786, p<0.01) but not between D2O-Mode and None-Mode, thus ***H9b*** is partially accepted. Meanwhile, we log significantly more re-demonstration clicks in Z2O-Mode than D2O-Mode (Z=-2.814, p<0.01), and significantly fewer re-demonstration clicks in Full-Mode than None-Mode (Z=-3.896, p<0.001).

From the qualitative analysis, we find that in Z2O-Mode, the stillness of the robot at the early stage and its gradual improvement prompted 42 out of 48 participants to refine their demonstrations. For example, P19 (M, 23) said *"The robot might not see my demonstration clearly, so I re-demonstrated. Later, the robots' feedback became stronger, encouraging me to demonstrate a few more times to improve it."* In a similar fashion, the disorderliness at the early stage and the perceptible progress of the robot in D2O-Mode also encouraged most participants (28/48) to demonstrate more. However, the agility of the robots made 18 participants think that the robot did not need too many instructions. For example, P48 (M, 23) said *"From the beginning, this robot was able to flexibly look around and swing its arms, which made me feel that it should have started to learn. So I did not think it was necessary to demonstrate many times."* For a similar argument, the high engagement of robots in Full-Mode led 35 participants to feel that there was no need to adjust their demonstration. In None-Mode, the lack of response from the robots made 27 participants eager to demonstrate again; meanwhile, as the robot still had no feedback in the later stage, their enthusiasm seemed to be dampened.

Frequency of Re-watching Behavior. Friedman test shows significant differences among the four engagement modes ($\chi^2(3)$=23.756, p<0.001). Particularly, significantly more re-watching behaviors is logged in Z2O-Mode and D2O-Mode than in Full-Mode (Z=-2.678, p<0.01; Z=-3.431, p<0.01) and in None-Mode (Z=-3.988, p<0.001; Z=-4.504, p<0.001), thus ***H9c*** and ***H9d*** are both accepted. Moreover, there is no significant difference in re-watching frequency between Z2O-Mode and D2O-Mode, but the significance is found between Full-Mode and None-Mode (Z=-2.723, p<0.01).

From the qualitative analysis, we can identify that the gradual progress of robots in Z2O-Mode stimulated most participants' (41/48) interest in closer inspections. For example, P22 (M, 27) stated *"The robot's feedback seemed to be different from that of the previous round. I'd like to check it again"*. Similarly, in D2O-Mode, the enhancement of the robot's engagement and the disorderliness of action urge most participants (31/48) to review it several times. By contrast, the constant full engagement in the Full-Mode reduces people's desire to revisit the students' reactions again. For example, P27 (F, 25) stated *"As I expected, its behavior was the same as the previous round so there was no need to watch again"* Similarly, the static state of the robot in None-Mode seriously impaired most participants' interest in re-watching. For instance, P30 (M, 24) said *"I just wanted to finish this teaching process as soon as possible."*

This result indicates that users' engagement will be affected by the AI system's engagement. Therefore, designers can engage users by adapting the expression of AI engagement. We should also note that AI's engagement communication could seriously affect the interaction workflow. For example, users reduced the number of re-demonstrations due to their overestimation of the learning progress of the robot that showed a high engagement (e.g., robots in Full-Mode), which may prevent the robot from obtaining enough demonstrations. Hence, designers are advised to carefully consider the associated impact on user engagement, interaction dynamics, and task outcomes when designing AI system engagement.

5.3.2 Timing and Motivation of Re-Demonstration and Re-Watching Behaviors. We also analyze the timing and motivation of the re-demonstration and re-watching behaviors. Overall, there are different timing patterns of re-demonstration and re-watching in different engagement modes, and we can find corresponding motivations under participants' behaviors from their qualitative feedback. The details of the results are provided in the online Appendix[2].

5.4 Order Effect on Users' Perceptions

Finally, we would like to investigate further whether the order of the engagement mode will affect users' overall perceptions of the robot and the RLfD process (users' subjective ratings in the post-task questionnaire). Overall, from the results, we can find that the order of engagement mode does not have significant effects on participants' perceptions except for several aspects. Although we have carefully considered and balanced the order effect in our study design, we believe that our further explorations will benefit the future design of robots' behaviors.

Specifically, to explore the potential differences, we treated the engagement mode orders, *1234*, *2341*, *3412*, *4123* (1:Z2O-Mode, 2:D2O-Mode, 3:Full-Mode, 4:None-Mode) as the independent variable, and took participants' ratings of each engagement mode in each question as the dependent variable (e.g., ratings of Z2O-Mode in **Q1**). Kruskal-Wallis test results show that there is no significant difference in most questions, except for (1) users' ratings of None-Mode in **Q1** ($\chi^2(3)$=12.920, p<0.01), (2) users' ratings of Full-Mode in **Q3** ($\chi^2(3)$=15.162, p<0.01), (3) users' ratings of None-Mode in **Q3** ($\chi^2(3)$=12.418, p<0.01), and (4) users' ratings of None-Mode in **Q7** ($\chi^2(3)$=10.305, p<0.05). Overall, we found the differences were driven by the participants' comparisons between the models. If participants taught the robot in Z2O- or D2O-Mode first, followed by robots in Full- and None-Mode, they tended to have lower perceptions of the latter.

We can get more specific reasons from post-hoc comparisons using Bonferroni correction combined with participants' feedback. Post-hoc tests show that in **Q1: users' perceptions of the acceptability of robots'**

behaviors, user's ratings of None-Mode in order **4**123 were significantly higher than in order 123**4** (p<0.05), order 23**4**1 (p<0.01). Qualitative feedback showed that after seeing the robot in Z2O-Mode or D2O-Mode, participants thought the robot's behaviors in None-Mode were neither appropriate nor acceptable. For example, P14 (female, age: 32, order 23**4**1) mentioned "*The lack of feedback from this robot is incomprehensible compared to the previous ones.*"

Besides, post-hoc tests using Bonferroni correction show that in **Q3: users' perception of robots' potential of further improvement given more demonstration**, user's ratings of Full-Mode in order 12**3**4 were significantly lower than in order **3**412 (p<0.05), order 412**3** (p<0.001). For example, in order 12**3**4, P32 (F, 25) stated that *"After teaching students who can make progress (in Z2O-Mode and D2O-Mode), I felt that this student's learning ability was relatively weak, so I didn't think the possibility of its further improvement is high."* Also, users' ratings of None-Mode in order 23**4**1 were significantly lower than in order **4**123 (p<0.05). Besides, users' ratings of None-Mode in order 123**4** were significantly lower than in order **4**123 (p<0.05). (In order **4**123) P21 (M, 25) explained *"The first robot I taught was the motionless one (in None-Mode), and I thought that although this robot was not engaged, probably it might make progress. But after teaching the later robots, I regretted that I didn't give the lowest score to the first robot directly."* (In order 123**4**), P27 (M, 25) mentioned *"In the first round, I thought Banana (the robot in None-Mode) was the same as Apple (the robot in Z2O-Mode) that could gradually follow my motions in the next rounds although did not move at the first round. However, Banana still didn't show any movement after that, so I sharply lowered my expectation of it."*

In addition, post-hoc tests using Bonferroni correction show that in **Q7: users' willingness to teach the robot in the future**, user's ratings of None-Mode in order **4**123 were significantly higher than in order 23**4**1 (p<0.05). Participants were less willing to teach robots in None-Mode if they had taught a robot in other modes (e.g., D2O-Mode or Full-Mode). For example, P1 (M, 25, order 2341) said "*Teaching the first two active robot students gave me a sense of accomplishment in teaching. However, when teaching the last one, I just felt like I was teaching in vain.*"

These findings suggest that contrast in the behaviors of different AI systems can affect users' perception to a certain extent, which is consistent with the theory of order bias [145] or anchoring bias [27, 188]. On the one hand, when designing the interaction between humans and AI systems, the representation of AI system behavior should take contextual information into account (e.g., user feedback in the previous round of interaction, their past experience with a similar system, etc.), to avoid users' over- or under-estimation of the current AI services due to comparison [27]. On the other hand, designers can also take advantage of the subtle influence of contrast on user perception to implicitly calibrate users' perception of the AI system and improve their experience.

6 DISCUSSION

In this section, we first discuss some key implications for robots' engagement design in RLfD scenarios and summarize possible design considerations for the engagement expression of AI systems in human-AI interaction based on our results and findings. Then we identify several limitations of our work. Finally, we discuss future research opportunities in robot engagement design.

6.1 Implications

6.1.1 RLfD should be a mutual process instead of a monodrama. Our findings reveal that in RLfD, robots lacking instant reciprocal feedback can mislead and demotivate their instructors. In human-human teaching, appropriate feedback can help teachers master students' learning status and make corresponding adjustments [108]. Otherwise, they may lack passion and motivation in teaching [111, 149]. Our experiments reveal similar findings in human-robot teaching. Results in Sec. 5.2 show that robots in None-Mode will lead to human teachers' incorrect perceptions of robots' intelligence and potential for further improvement. Moreover, not seeing robots' feedback can lower humans' confidence in their demonstrations and make participants bored and not willing to give further teaching. In comparison, robots with engagement expressions are more appreciated by participants. Therefore, it is recommended to integrate appropriate robot feedback, e.g., engagement, during the demonstration stage for human instructors.

6.1.2 Robot feedback design should avoid introducing misperception. From the results of our user study, we have learned that if the robot feedback is not appropriately designed, it may lead to users' misperceptions of the robot. Specifically, based on our findings, we summarize several insights into generating and expressing proper feedback while a robot is learning a skill from humans. First, it would be better not to show behavioral cues very similar to the target action as feedback signals. Our pilot study (Sec. 3.3.2) results show that people are likely to misinterpret the robot's engagement as its actual learning outcome otherwise. Second, it is essential to help users maintain accurate mental models of the robots by responsible engagement expression design. Designers should match the expression of the engagement of a robot with the actual state of its underlying algorithm/model to calibrate users' perceptions and expectations. On the one hand, designers are recommended to be careful about the possible impact of excessively positive robot engagement on users. For example, if the engagement level displayed by the robot is significantly higher than its actual internal state (e.g., the Full-Mode), the user might overestimate the ability of the robot with overly high expectations, which may cause insufficient demonstration. On the other hand, designers are also suggested to avoid the impact of excessively negative (or no) engagement. For example, if the engagement level of the robot is significantly lower than the level of its actual state (e.g., the None-Mode), users will underestimate the capabilities of the robot and even give up further teaching, resulting in negative consequences. Third, the design should allow users to draw visual/mental associations between the robot's feedback and the task at hand. If the robot exhibits completely unrelated behaviors in reaction to people's teaching, it may be deemed impolite and consequently evoke negative emotions in humans [57, 152]. For example, from the interview in Sec. 5.2, some users said that the random movement of the robot (in D2O-Mode) at the beginning made them feel uncomfortable. In summary, when designing appropriate robot learning feedback in RLfD, it is necessary to take the task to teach, the actual learning state behind the robot, and users' experience into consideration.

6.1.3 Engagement as a design material for transparently communicating the internal status of AI systems. Some previous work has found that users will update their mental models based on the directly displayed status of the AI system in the form of visualization or text [198]. Our findings extend this understanding and demonstrate that users will also update their mental models of robots' learning status based on the conveyed level of their engagement – a common social cue in human-human interaction – linked to the learning progress of their backend AI models. The finding is in line with several well-known HCI theories and frameworks [126–128], indicating that not only for conveying the internal status of robots, but engagement can also serve as a potential design material to expand the existing design space for improving the transparency of other types of AI systems. The application and effect of engagement design for general AI systems beyond the RLfD setting can be explored in future work.

6.1.4 Design recommendations for engagement expressions of AI systems. In general, the design of engagement expressions needs to be determined according to the specific user interaction modality, the interaction behavior, the form of the AI system, and the contexts. Based on our key findings learned from robot engagement design, we outline some design considerations for general engagement communication of AI systems. First, the engagement displayed by the AI system should be closely related to the user's interactive behavior under the ongoing task (e.g., humans' body movements in our case). Engagement with irrelevant information may confuse or distract users. Second, the AI system should maintain the role of *"receiver"* or *"listener"* to avoid causing interference when the user is inputting into the system. It, however, could provide active, non-interruptive feedback in the process (e.g., robots' gaze following and rhythm synchrony in our case). Third, the engagement expression of an AI system should be adaptive based on the system's dynamic inner status. Static engagement expressions could impair the update of users' mental models. Finally, the engagement of AI systems can be designed to simulate the exchange of engagement cues as in human-human interactions. It has been shown that people will intuitively ascribe intentionality to systems as they do to humans, and naturally apply the same social heuristics used for human interactions to systems [32, 124]. According to the participants' feedback in our user study, it is natural for them, especially those non-experts in robotics and AI, to treat robots as human students. Based on this phenomenon and HCI theories such as CASA [124], we can design appropriate engagement expressions for AI systems according to the common social behaviors in interpersonal interaction.

6.2 Limitations

There exist several limitations in our current implementation and experiment setting. First, our study was conducted in a simulated environment. Although it has been verified that a robot's internal status could be effectively conveyed to users via on-screen display and users' interpretations from on-screen robot were consistent with the physical robot [82, 178, 179, 194], there are still some differences between interacting with a physical robot and seeing it on a screen [77]. We plan to carry out an in-the-wild RLfD study with physical robots in the future and will take real-life factors, e.g., distraction caused by the noise of robot movements, into design consideration. Second, we experimented with the most common learning process of robots in RLfD, i.e., the learning state is getting better and better with more training iterations. However, it should be noted that not all training processes guarantee improvement toward the intended goal, and we will explore users' perceptions under different situations of learning processes, such as failure in teaching. Third, we adopt a linear mapping between algorithmic learning progress and engagement expression, which is intuitive for human teachers. However, we also see the possibility of non-linear mapping between learning status and engagement level, which can be user-specific based on modeling of users' patience, users' experience in teaching robots, users' teaching interest, etc. In the future, we plan to design experiments to test user-adaptive non-linear mapping between learning status and engagement level. Moreover, this paper only focuses on the communication of learning status during the demonstration gathering stage. As shown in Figure 1, users could also get a sense of the learning performance by having the robot showcase its learning outcome at the end of the policy-deriving stage. We will study the means to combine these two types of transparency communications.

6.3 Future Research Opportunities

This paper is the first step to computationally design *Learning Engagement* to improve transparency in robot learning from human demonstrations. There are several valuable directions for further investigation.

Exploring other Learning Engagement Design. The proposed two *Learning Engagement* modes, Z2O-Mode and D2O-Mode, are only two representatives of many possible design alternatives to show the robot learning status. Besides eye gaze and rhythm, there are other engagement cues worth exploring, such as nodding [97], scratching head, leaning, and changing the distance between robots and humans, as long as the design is consistent with the common behaviors in human-human teaching and learning [29].

Extending the proposed method to other types of robots and LfD tasks. In this paper, we leverage gaze and upper body movements to convey learning engagement given the form factor of the Pepper robot. And this paper emphasizes low-level, skill-oriented robot learning tasks. High-level, goal-oriented LfD tasks that aim to solve a given problem (e.g., order coffee using mobile app [102]) are also common in our daily life. Compared with skill-oriented tasks, in goal-oriented tasks, people care more about the completion of the final goal where it may not be enough to express engagement only through the gaze and body movements of the robot. Next step, we plan to extend the proposed methods to non-humanoid-shaped robots and goal-oriented tasks.

Considering the diversity of user perception. Our experiment shows that, although users tend to have a similar mental model about a robot learner under the same engagement mode, their attitudes and reactions can be quite different. For example, some participants think that the random movement of the robot in *D2O-Mode* is a sign of the robot's liveliness and intelligence, while some participants worry about the robot's "disobedience". This reminds us of the existence of individual differences in user beliefs and preferences in the teaching process. In other words, one should consider users' differences by modeling their personality and cognitive processes, when designing behaviors of AI systems.

Extending to more general human-AI interaction scenarios. This paper has shown the potential of communicating AI agents' inner states through engagement expression. In the next step, we plan to apply the method proposed in this paper to the more general human-AI interaction scenarios where users interact with more common AI systems (such as GUI-based systems). In addition, we also hope to explore more engagement expression methods under different interaction scenarios and different forms of AI systems based on the theoretical support of sociology and psychology.

7 CONCLUSION

In this paper, we investigated a new approach to making robots' learning status transparent by modeling adaptive expressions of robot engagement. We proposed and verified the design of attentional and behavioral engagement expressions, *Gaze Following* and *Rhythm Synchrony*. And we computationally incorporated the robot's internal learning status into these engagement cues, which results in the *Z2O-Mode* and *D2O-Mode* of *Learning Engagement* for communicating dynamic learning progress. We conducted an online user study in a simulated environment to investigate the effects of different engagement modes on users' mental models of robots' learning process and users' perception of the overall teaching process. Quantitative and qualitative results showed that different engagement expressions led to different human perceptions of robots' learning status/progress, expected learning outcomes, intelligence, and acceptability, and affected users' self-assessment, self-reflection, teaching experience, and engagement. Based on our findings, we provided implications for both human-robot teaching and broader human-AI interaction. We hope this work serves as a starting point for making the inner status of AI systems transparent for non-expert users via engagement expression.

REFERENCES

[1] Pieter Abbeel and Andrew Y Ng. 2004. Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the twenty-first international conference on Machine learning*. 1.

[2] Henny Admoni and Brian Scassellati. 2017. Social eye gaze in human-robot interaction: a review. *Journal of Human-Robot Interaction* 6, 1 (2017), 25–63.

[3] Jessica S Ancker, Yalini Senathirajah, Rita Kukafka, and Justin B Starren. 2006. Design features of graphs in health risk communication: a systematic review. *Journal of the American Medical Informatics Association* 13, 6 (2006), 608–618.

[4] Pierre Andry, Arnaud Blanchard, and Philippe Gaussier. 2010. Using the rhythm of nonverbal human–robot interaction as a signal for learning. *IEEE Transactions on Autonomous Mental Development* 3, 1 (2010), 30–42.

[5] 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.

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

[7] Arkar Min Aung, Anand Ramakrishnan, and Jacob R Whitehill. 2018. Who Are They Looking At? Automatic Eye Gaze Following for Classroom Observation Video Analysis. *International Educational Data Mining Society* (2018).

[8] Jan Bengtsson. 1995. What is reflection? On reflection in the teaching profession and teacher education. *Teachers and teaching* 1, 1 (1995), 23–32.

[9] Jesse Josua Benjamin, Arne Berger, Nick Merrill, and James Pierce. 2021. Machine Learning Uncertainty as a Design Material: A Post-Phenomenological Inquiry. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[10] Umang Bhatt, Javier Antorán, Yunfeng Zhang, Q Vera Liao, Prasanna Sattigeri, Riccardo Fogliato, Gabrielle Melançon, Ranganath Krishnan, Jason Stanley, Omesh Tickoo, et al. 2021. Uncertainty as a form of transparency: Measuring, communicating, and using uncertainty. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 401–413.

[11] Tapomayukh Bhattacharjee, Gilwoo Lee, Hanjun Song, and Siddhartha S Srinivasa. 2019. Towards robotic feeding: Role of haptics in fork-based food manipulation. *IEEE Robotics and Automation Letters* 4, 2 (2019), 1485–1492.

[12] Roi Blanco, Diego Ceccarelli, Claudio Lucchese, Raffaele Perego, and Fabrizio Silvestri. 2012. You should read this! let me explain you why: explaining news recommendations to users. In *Proceedings of the 21st ACM international conference on Information and knowledge management*. 1995–1999.

[13] Michael E Bratman. 1992. Shared cooperative activity. *The philosophical review* 101, 2 (1992), 327–341.

[14] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. (2012).

[15] Tim Brys, Anna Harutyunyan, Halit Bener Suay, Sonia Chernova, Matthew E Taylor, and Ann Nowé. 2015. Reinforcement learning from demonstration through shaping. In *Twenty-fourth international joint conference on artificial intelligence*.

[16] Colin Bryson and Len Hand. 2007. The role of engagement in inspiring teaching and learning. *Innovations in education and teaching international* 44, 4 (2007), 349–362.

[17] Maya Cakmak and Andrea L Thomaz. 2012. Designing robot learners that ask good questions. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 17–24.

[18] Robert M Carini, George D Kuh, and Stephen P Klein. 2006. Student engagement and student learning: Testing the linkages. *Research in higher education* 47, 1 (2006), 1–32.

[19] Jessy Ceha, Nalin Chhibber, Joslin Goh, Corina McDonald, Pierre-Yves Oudeyer, Dana Kulić, and Edith Law. 2019. Expression of curiosity in social robots: Design, perception, and effects on behaviour. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[20] Jeanie Chan and Goldie Nejat. 2012. Social intelligence for a robot engaging people in cognitive training activities. *International Journal of Advanced Robotic Systems* 9, 4 (2012), 113.

[21] Crystal Chao, Maya Cakmak, and Andrea L Thomaz. 2010. Transparent active learning for robots. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 317–324.

[22] Peter McFaul Chapman. 1997. *Models of engagement: Intrinsically motivated interaction with multimedia learning software*. Ph. D. Dissertation. University of Waterloo.

[23] Tanya L Chartrand and John A Bargh. 1999. The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology* 76, 6 (1999), 893.

[24] Tanya L Chartrand, William W Maddux, and Jessica L Lakin. 2005. Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. *The new unconscious* (2005), 334–361.

[25] Sonia Chernova and Andrea L Thomaz. [n.d.]. Robot learning from human teachers. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 8 ([n.d.]).

[26] Sonia Chernova and Manuela Veloso. 2009. Interactive policy learning through confidence-based autonomy. *Journal of Artificial Intelligence Research* 34 (2009), 1–25.

[27] Isaac Cho, Ryan Wesslen, Alireza Karduni, Sashank Santhanam, Samira Shaikh, and Wenwen Dou. 2017. The anchoring effect in decision-making with visual analytics. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 116–126.

[28] Felipe Codevilla, Matthias Müller, Antonio López, Vladlen Koltun, and Alexey Dosovitskiy. 2018. End-to-end driving via conditional imitation learning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 4693–4700.

[29] Scotty Craig, Arthur Graesser, Jeremiah Sullins, and Barry Gholson. 2004. Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of educational media* 29, 3 (2004), 241–250.

[30] Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten Van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. 2008. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-adapted interaction* 18, 5 (2008), 455.

[31] Yuchen Cui and Scott Niekum. 2017. Active learning from critiques via bayesian inverse reinforcement learning. In *Robotics: Science and Systems Workshop on Mathematical Models, Algorithms, and Human-Robot Interaction*.

[32] Maartje MA De Graaf and Bertram F Malle. 2019. People's explanations of robot behavior subtly reveal mental state inferences. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 239–248.

[33] Luis De-Marcos, Adrián Domínguez, Joseba Saenz-de Navarrete, and Carmen Pagés. 2014. An empirical study comparing gamification and social networking on e-learning. *Computers & education* 75 (2014), 82–91.

[34] Richard Dearden, Nir Friedman, and Stuart Russell. 1998. Bayesian Q-learning. In *Aaai/iaai*. 761–768.

[35] Emilie Delaherche, Mohamed Chetouani, Ammar Mahdhaoui, Catherine Saint-Georges, Sylvie Viaux, and David Cohen. 2012. Interpersonal synchrony: A survey of evaluation methods across disciplines. *IEEE Transactions on Affective Computing* 3, 3 (2012), 349–365.

[36] Miha Deniša, Andrej Gams, Aleš Ude, and Tadej Petrič. 2015. Learning compliant movement primitives through demonstration and statistical generalization. *IEEE/ASME transactions on mechatronics* 21, 5 (2015), 2581–2594.

[37] Munjal Desai, Poornima Kaniarasu, Mikhail Medvedev, Aaron Steinfeld, and Holly Yanco. 2013. Impact of robot failures and feedback on real-time trust. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 251–258.

[38] Lorin Dole and Wendy Ju. 2019. Face and Ecological Validity in Simulations: Lessons from Search-and-Rescue HRI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–8.
[39] Gabriel Dulac-Arnold, Daniel Mankowitz, and Todd Hester. 2019. Challenges of real-world reinforcement learning. *arXiv preprint arXiv:1904.12901* (2019).
[40] Upol Ehsan, Q Vera Liao, Michael Muller, Mark O Riedl, and Justin D Weisz. 2021. Expanding explainability: Towards social transparency in ai systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–19.
[41] Malin Eiband, Daniel Buschek, and Heinrich Hussmann. 2021. How to support users in understanding intelligent systems? Structuring the discussion. In *26th International Conference on Intelligent User Interfaces*. 120–132.
[42] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing transparency design into practice. In *23rd international conference on intelligent user interfaces*. 211–223.
[43] Charles W Eriksen and James E Hoffman. 1972. Temporal and spatial characteristics of selective encoding from visual displays. *Perception & psychophysics* 12, 2 (1972), 201–204.
[44] Amir-massoud Farahmand, Andre Barreto, and Daniel Nikovski. 2017. Value-aware loss function for model-based reinforcement learning. In *Artificial Intelligence and Statistics*. PMLR, 1486–1494.
[45] Jonas Flodén. 2017. The impact of student feedback on teaching in higher education. *Assessment & Evaluation in Higher Education* 42, 7 (2017), 1054–1068.
[46] Kotaro Funakoshi, Kazuki Kobayashi, Mikio Nakano, Seiji Yamada, Yasuhiko Kitamura, and Hiroshi Tsujino. 2008. Smoothing human-robot speech interactions by using a blinking-light as subtle expression. In *Proceedings of the 10th international conference on Multimodal interfaces*. 293–296.
[47] Dedre Gentner and Albert L Stevens. 2014. *Mental models*. Psychology Press.
[48] Katy Ilonka Gero, Zahra Ashktorab, Casey Dugan, Qian Pan, James Johnson, Werner Geyer, Maria Ruiz, Sarah Miller, David R Millen, Murray Campbell, et al. 2020. Mental Models of AI Agents in a Cooperative Game Setting. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.
[49] Dimitra Gkatzia, Oliver Lemon, and Verena Rieser. 2016. Natural language generation enhances human decision-making with uncertain information. *arXiv preprint arXiv:1606.03254* (2016).
[50] Alexander Z Golbin, HM Kravitz, and LG Keith. 2004. *Periodic and rhythmic parasomnias*. London.
[51] Tara Gray and Laura Madson. 2007. Ten easy ways to engage your students. *College Teaching* 55, 2 (2007), 83–87.
[52] Lexie Grudnoff. 2011. Rethinking the practicum: Limitations and possibilities. *Asia-Pacific Journal of Teacher Education* 39, 3 (2011), 223–234.
[53] David Gunning and David Aha. 2019. DARPA's explainable artificial intelligence (XAI) program. *AI Magazine* 40, 2 (2019), 44–58.
[54] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. 2019. XAI—Explainable artificial intelligence. *Science Robotics* 4, 37 (2019).
[55] Chris Harrison, Gary Hsieh, Karl DD Willis, Jodi Forlizzi, and Scott E Hudson. 2011. Kineticons: using iconographic motion in graphical user interface design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1999–2008.
[56] Elaine Hatfield, John T Cacioppo, and Richard L Rapson. 1993. Emotional contagion. *Current directions in psychological science* 2, 3 (1993), 96–100.
[57] Michael Haugh. 2013. Im/politeness, social practice and the participation order. *Journal of Pragmatics* 58 (2013), 52–72.
[58] Florian Hawlitschek, Lars-Erik Jansen, Ewa Lux, Timm Teubner, and Christof Weinhardt. 2016. Colors and trust: the influence of user interface design on trust and reciprocity. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*. IEEE, 590–599.
[59] Todd Hester, Matej Vecerik, Olivier Pietquin, Marc Lanctot, Tom Schaul, Bilal Piot, Dan Horgan, John Quan, Andrew Sendonaris, Ian Osband, et al. 2018. Deep q-learning from demonstrations. In *Thirty-second AAAI conference on artificial intelligence*.
[60] Todd Hester, Matej Vecerik, Olivier Pietquin, Marc Lanctot, Tom Schaul, Bilal Piot, Andrew Sendonaris, Gabriel Dulac-Arnold, Ian Osband, John Agapiou, et al. 2017. Learning from demonstrations for real world reinforcement learning. (2017).
[61] Guy Hoffman and Gil Weinberg. 2010. Shimon: an interactive improvisational robotic marimba player. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*. 3097–3102.
[62] Guy Hoffman, Oren Zuckerman, Gilad Hirschberger, Michal Luria, and Tal Shani-Sherman. 2015. Design and evaluation of a peripheral robotic conversation companion. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 3–10.
[63] Fred Hohman, Andrew Head, Rich Caruana, Robert DeLine, and Steven M Drucker. 2019. Gamut: A design probe to understand how data scientists understand machine learning models. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.
[64] Jonggi Hong, Kyungjun Lee, June Xu, and Hernisa Kacorri. 2019. Exploring Machine Teaching for Object Recognition with the Crowd. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.
[65] Jonggi Hong, Kyungjun Lee, June Xu, and Hernisa Kacorri. 2020. Crowdsourcing the Perception of Machine Teaching. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[66] Julian Hough and David Schlangen. 2017. It's Not What You Do, It's How You Do It: Grounding Uncertainty for a Simple Robot. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*. IEEE, 274–282.

[67] Yuhan Hu and Guy Hoffman. 2019. Using skin texture change to design emotion expression in social robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2–10.

[68] Honglan Huang, Jincai Huang, Yanghe Feng, Jiarui Zhang, Zhong Liu, Qi Wang, and Li Chen. 2019. On the improvement of reinforcement active learning with the involvement of cross entropy to address one-shot learning problem. *PloS one* 14, 6 (2019), e0217408.

[69] 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*. 377–384.

[70] Carlos T Ishi, ChaoRan Liu, Hiroshi Ishiguro, and Norihiro Hagita. 2010. Head motion during dialogue speech and nod timing control in humanoid robots. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 293–300.

[71] Riitta Jääskeläinen. 2010. Think-aloud protocol. *Handbook of translation studies* 1 (2010), 371–374.

[72] Dylan Jennings and Miguel Figliozzi. 2019. Study of sidewalk autonomous delivery robots and their potential impacts on freight efficiency and travel. *Transportation Research Record* 2673, 6 (2019), 317–326.

[73] John Jonides. 1983. Further toward a model of the mind's eye's movement. *Bulletin of the Psychonomic Society* 21, 4 (1983), 247–250.

[74] Malte F Jung, David Sirkin, Turgut M Gür, and Martin Steinert. 2015. Displayed uncertainty improves driving experience and behavior: The case of range anxiety in an electric car. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2201–2210.

[75] Daniel Kahneman. 1973. *Attention and effort*. Vol. 1063. Citeseer.

[76] Jacques Kaiser, Svenja Melbaum, J Camilo Vasquez Tieck, Arne Roennau, Martin V Butz, and Rudiger Dillmann. 2018. Learning to reproduce visually similar movements by minimizing event-based prediction error. In *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*. IEEE, 260–267.

[77] Sara Kiesler, Aaron Powers, Susan R Fussell, and Cristen Torrey. 2008. Anthropomorphic interactions with a robot and robot–like agent. *Social Cognition* 26, 2 (2008), 169–181.

[78] René F Kizilcec. 2016. How much information? Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2390–2395.

[79] Bart P Knijnenburg and Martijn C Willemsen. 2016. Inferring capabilities of intelligent agents from their external traits. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 6, 4 (2016), 1–25.

[80] Rafal Kocielnik, Saleema Amershi, and Paul N Bennett. 2019. Will you accept an imperfect ai? exploring designs for adjusting end-user expectations of ai systems. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.

[81] Takanori Komatsu, Seiji Yamada, Kazuki Kobayashi, Kotaro Funakoshi, and Mikio Nakano. 2010. Artificial subtle expressions: Intuitive notification methodology of artifacts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1941–1944.

[82] Takanori Komatsu, Seiji Yamada, Kazuki Kobayashi, Kotaro Funakoshi, and Mikio Nakano. 2011. Effects of different types of artifacts on interpretations of artificial subtle expressions (ASEs). In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. 1249–1254.

[83] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. 2019. Personalized explanations for hybrid recommender systems. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 379–390.

[84] Robert M Krauss, Yihsiu Chen, and Purnima Chawla. 1996. Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? *Advances in experimental social psychology* 28 (1996), 389–450.

[85] Thomas Kreuz, Florian Mormann, Ralph G Andrzejak, Alexander Kraskov, Klaus Lehnertz, and Peter Grassberger. 2007. Measuring synchronization in coupled model systems: A comparison of different approaches. *Physica D: Nonlinear Phenomena* 225, 1 (2007), 29–42.

[86] Klas Kronander, Mohammad Khansari, and Aude Billard. 2015. Incremental motion learning with locally modulated dynamical systems. *Robotics and Autonomous Systems* 70 (2015), 52–62.

[87] Sari Kujala, Ruth Mugge, and Talya Miron-Shatz. 2017. The role of expectations in service evaluation: A longitudinal study of a proximity mobile payment service. *International Journal of Human-Computer Studies* 98 (2017), 51–61.

[88] Todd Kulesza, Simone Stumpf, Margaret Burnett, and Irwin Kwan. 2012. Tell me more? The effects of mental model soundness on personalizing an intelligent agent. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1–10.

[89] Todd Kulesza, Simone Stumpf, Margaret Burnett, Weng-Keen Wong, Yann Riche, Travis Moore, Ian Oberst, Amber Shinsel, and Kevin McIntosh. 2010. Explanatory debugging: Supporting end-user debugging of machine-learned programs. In *2010 IEEE Symposium on Visual Languages and Human-Centric Computing*. IEEE, 41–48.

[90] Yoshinori Kuno, Kazuhisa Sadazuka, Michie Kawashima, Keiichi Yamazaki, Akiko Yamazaki, and Hideaki Kuzuoka. 2007. Museum guide robot based on sociological interaction analysis. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1191–1194.

[91] Zeno Kupper, Fabian Ramseyer, Holger Hoffmann, and Wolfgang Tschacher. 2015. Nonverbal synchrony in social interactions of patients with schizophrenia indicates socio-communicative deficits. *PLoS One* 10, 12 (2015), e0145882.

[92] Minae Kwon, Sandy H Huang, and Anca D Dragan. 2018. Expressing robot incapability. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 87–95.

[93] Jessica L Lakin and Tanya L Chartrand. 2003. Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological science* 14, 4 (2003), 334–339.

[94] Michael Laskey, Caleb Chuck, Jonathan Lee, Jeffrey Mahler, Sanjay Krishnan, Kevin Jamieson, Anca Dragan, and Ken Goldberg. 2017. Comparing human-centric and robot-centric sampling for robot deep learning from demonstrations. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 358–365.

[95] Brenda Laurel. 2013. *Computers as theatre*. Addison-Wesley.

[96] Clemente Lauretti, Francesca Cordella, Eugenio Guglielmelli, and Loredana Zollo. 2017. Learning by demonstration for planning activities of daily living in rehabilitation and assistive robotics. *IEEE Robotics and Automation Letters* 2, 3 (2017), 1375–1382.

[97] Christopher Lee, Neal Lesh, Candace L Sidner, Louis-Philippe Morency, Ashish Kapoor, and Trevor Darrell. 2004. Nodding in conversations with a robot. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems*. 785–786.

[98] John D Lee. 2008. Review of a pivotal Human Factors article:"Humans and automation: use, misuse, disuse, abuse". *Human Factors* 50, 3 (2008), 404–410.

[99] Hagen Lehmann, Joan Saez-Pons, Dag Sverre Syrdal, and Kerstin Dautenhahn. 2015. In good company? Perception of movement synchrony of a non-anthropomorphic robot. *PloS one* 10, 5 (2015), e0127747.

[100] 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*. 909–916.

[101] Mao Li, Tim Brys, and Daniel Kudenko. 2018. Introspective Reinforcement Learning and Learning from Demonstration.. In *AAMAS*. 1992–1994.

[102] Toby Jia-Jun Li, Amos Azaria, and Brad A Myers. 2017. SUGILITE: creating multimodal smartphone automation by demonstration. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 6038–6049.

[103] Wei Li, Tovi Grossman, and George Fitzmaurice. 2014. CADament: a gamified multiplayer software tutorial system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3369–3378.

[104] Q Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[105] Bingyu Liu, Weihong Deng, Yaoyao Zhong, Mei Wang, Jiani Hu, Xunqiang Tao, and Yaohai Huang. 2019. Fair loss: Margin-aware reinforcement learning for deep face recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10052–10061.

[106] Andrea Lockerd and Cynthia Breazeal. 2004. Tutelage and socially guided robot learning. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, Vol. 4. IEEE, 3475–3480.

[107] Joseph B Lyons, Garrett G Sadler, Kolina Koltai, Henri Battiste, Nhut T Ho, Lauren C Hoffmann, David Smith, Walter Johnson, and Robert Shively. 2017. Shaping trust through transparent design: theoretical and experimental guidelines. In *Advances in human factors in robots and unmanned systems*. Springer, 127–136.

[108] Shuai Ma, Taichang Zhou, Fei Nie, and Xiaojuan Ma. 2022. Glancee: An Adaptable System for Instructors to Grasp Student Learning Status in Synchronous Online Classes. In *CHI Conference on Human Factors in Computing Systems*. 1–25.

[109] Xiaojuan Ma. 2018. Towards Human-Engaged AI.. In *IJCAI*. 5682–5686.

[110] William W Maddux, Elizabeth Mullen, and Adam D Galinsky. 2008. Chameleons bake bigger pies and take bigger pieces: Strategic behavioral mimicry facilitates negotiation outcomes. *Journal of Experimental Social Psychology* 44, 2 (2008), 461–468.

[111] Chutaphon Masantiah, Shotiga Pasiphol, and Kamonwan Tangdhanakanond. 2018. Student and feedback: Which type of feedback is preferable? *Kasetsart Journal of Social Sciences* (2018).

[112] Malia F Mason, Michael I Norton, John D Van Horn, Daniel M Wegner, Scott T Grafton, and C Neil Macrae. 2007. Wandering minds: the default network and stimulus-independent thought. *Science* 315, 5810 (2007), 393–395.

[113] Andrew N Meltzoff, Rechele Brooks, Aaron P Shon, and Rajesh PN Rao. 2010. "Social" robots are psychological agents for infants: A test of gaze following. *Neural networks* 23, 8-9 (2010), 966–972.

[114] Benedikt Merz, Alexandre N Tuch, and Klaus Opwis. 2016. Perceived user experience of animated transitions in mobile user interfaces. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 3152–3158.

[115] Martijn Millecamp, Nyi Nyi Htun, Cristina Conati, and Katrien Verbert. 2019. To explain or not to explain: the effects of personal characteristics when explaining music recommendations. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 397–407.

[116] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* 267 (2019), 1–38.

[117] Noriaki Mitsunaga, Christian Smith, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2008. Adapting robot behavior for human–robot interaction. *IEEE Transactions on Robotics* 24, 4 (2008), 911–916.

[118] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-level control through deep reinforcement learning. *nature* 518, 7540 (2015),

529–533.

[119] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine* 19, 2 (2012), 98–100.

[120] Christina Moro, Goldie Nejat, and Alex Mihailidis. 2018. Learning and personalizing socially assistive robot behaviors to aid with activities of daily living. *ACM Transactions on Human-Robot Interaction (THRI)* 7, 2 (2018), 1–25.

[121] Alexander Mörtl, Tamara Lorenz, and Sandra Hirche. 2014. Rhythm patterns interaction-synchronization behavior for human-robot joint action. *PloS one* 9, 4 (2014), e95195.

[122] Omar Mubin, Muneeb Imtiaz Ahmad, Simranjit Kaur, Wen Shi, and Aila Khan. 2018. Social robots in public spaces: A meta-review. In *International Conference on Social Robotics*. Springer, 213–220.

[123] Jacqueline Nadel, Ken Prepin, and Mako Okanda. 2005. Experiencing contingency and agency: First step toward self-understanding in making a mind? *Interaction studies* 6, 3 (2005), 447–462.

[124] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 72–78.

[125] Andrew Y Ng, Daishi Harada, and Stuart Russell. 1999. Policy invariance under reward transformations: Theory and application to reward shaping. In *Icml*, Vol. 99. 278–287.

[126] Donald A Norman. 1986. Cognitive engineering. *User centered system design* 31 (1986), 61.

[127] Donald A Norman. 1988. *The psychology of everyday things.* Basic books.

[128] Donald A Norman. 2014. *Some observations on mental models.* Psychology Press.

[129] Mahsan Nourani, Chiradeep Roy, Jeremy E Block, Donald R Honeycutt, Tahrima Rahman, Eric Ragan, and Vibhav Gogate. 2021. Anchoring Bias Affects Mental Model Formation and User Reliance in Explainable AI Systems. In *26th International Conference on Intelligent User Interfaces*. 340–350.

[130] Kazuo Okamura and Seiji Yamada. 2020. Adaptive trust calibration for human-AI collaboration. *Plos one* 15, 2 (2020), e0229132.

[131] Denis Parra and Peter Brusilovsky. 2015. User-controllable personalization: A case study with SetFusion. *International Journal of Human-Computer Studies* 78 (2015), 43–67.

[132] Tomislav Pejsa, Dan Bohus, Michael F Cohen, Chit W Saw, James Mahoney, and Eric Horvitz. 2014. Natural communication about uncertainties in situated interaction. In *Proceedings of the 16th international conference on multimodal interaction*. 283–290.

[133] Reinhard Pekrun and Lisa Linnenbrink-Garcia. 2012. Academic emotions and student engagement. In *Handbook of research on student engagement*. Springer, 259–282.

[134] Bei Peng, James MacGlashan, Robert Loftin, Michael L Littman, David L Roberts, and Matthew E Taylor. 2016. A need for speed: Adapting agent action speed to improve task learning from non-expert humans. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems*.

[135] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel van de Panne. 2018. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–14.

[136] Xue Bin Peng and Michiel van de Panne. 2017. Learning locomotion skills using deeprl: Does the choice of action space matter?. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. 1–13.

[137] Leah Perlmutter, Eric Kernfeld, and Maya Cakmak. 2016. Situated Language Understanding with Human-like and Visualization-Based Transparency.. In *Robotics: Science and Systems*, Vol. 12. 40–50.

[138] Affan Pervez, Yuecheng Mao, and Dongheui Lee. 2017. Learning deep movement primitives using convolutional neural networks. In *2017 IEEE-RAS 17th international conference on humanoid robotics (Humanoids)*. IEEE, 191–197.

[139] William H Press and Saul A Teukolsky. 1990. Savitzky-Golay smoothing filters. *Computers in Physics* 4, 6 (1990), 669–672.

[140] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as mechanisms for supporting algorithmic transparency. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[141] Stéphane Raffard, Robin N Salesse, Catherine Bortolon, Benoit G Bardy, José Henriques, Ludovic Marin, Didier Stricker, and Delphine Capdevielle. 2018. Using mimicry of body movements by a virtual agent to increase synchronization behavior and rapport in individuals with schizophrenia. *Scientific reports* 8, 1 (2018), 1–10.

[142] Vasumathi Raman, Constantine Lignos, Cameron Finucane, Kenton CT Lee, Mitchell P Marcus, and Hadas Kress-Gazit. 2013. Sorry Dave, I'm Afraid I Can't Do That: Explaining Unachievable Robot Tasks Using Natural Language.. In *Robotics: Science and Systems*, Vol. 2. Citeseer, 2–1.

[143] Harish Ravichandar, Athanasios S Polydoros, Sonia Chernova, and Aude Billard. 2020. Recent advances in robot learning from demonstration. *Annual Review of Control, Robotics, and Autonomous Systems* 3 (2020), 297–330.

[144] Robert Reid. 1999. Attention deficit hyperactivity disorder: Effective methods for the classroom. *Focus on exceptional children* 32, 4 (1999), 1–20.

[145] William E Remus and Jeffrey E Kottemann. 1986. Toward intelligent decision support systems: An artificially intelligent statistician. *Mis Quarterly* (1986), 403–418.

[146] Alfréd Rényi et al. 1961. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California.

[147] Ralph E Reynolds and Larry L Shirey. 1988. The role of attention in studying and learning. In *Learning and study strategies*. Elsevier, 77–100.

[148] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.

[149] Paul W Richardson and Helen MG Watt. 2010. Current and future directions in teacher motivation research. In *The decade ahead: Applications and contexts of motivation and achievement*. Emerald Group Publishing Limited.

[150] Philippe Rochat. 2003. Five levels of self-awareness as they unfold early in life. *Consciousness and cognition* 12, 4 (2003), 717–731.

[151] Heleen Rutjes, Martijn Willemsen, and Wijnand IJsselsteijn. 2019. Considerations on explainable AI and users' mental models. In *CHI 2019 Workshop: Where is the Human? Bridging the Gap Between AI and HCI*. Association for Computing Machinery, Inc.

[152] Maha Salem, Micheline Ziadee, and Majd Sakr. 2013. Effects of politeness and interaction context on perception and experience of HRI. In *International Conference on Social Robotics*. Springer, 531–541.

[153] Claude Sammut, Scott Hurst, Dana Kedzier, and Donald Michie. 1992. Learning to fly. In *Machine Learning Proceedings 1992*. Elsevier, 385–393.

[154] Akanksha Saran, Srinjoy Majumdar, Andrea Thomaz, and Scott Niekum. 2018. Real-time human gaze following for human-robot interaction. In *HRI'18*.

[155] Stefan Schaal et al. 1997. Learning from demonstration. *Advances in neural information processing systems* (1997), 1040–1046.

[156] James Schaffer, Prasanna Giridhar, Debra Jones, Tobias Höllerer, Tarek Abdelzaher, and John O'donovan. 2015. Getting the message? A study of explanation interfaces for microblog data analysis. In *Proceedings of the 20th international conference on intelligent user interfaces*. 345–356.

[157] Matthias Scheutz, Paul Schermerhorn, and James Kramer. 2006. The utility of affect expression in natural language interactions in joint human-robot tasks. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 226–233.

[158] Peter Seldin. 1989. Using student feedback to improve teaching. *New Directions for Teaching and Learning* 37 (1989), 89–97.

[159] Aran Sena, Yuchen Zhao, and Matthew J Howard. 2018. Teaching Human Teachers to Teach Robot Learners.. In *ICRA*. 1–7.

[160] Emmanuel Senft, Paul Baxter, James Kennedy, and Tony Belpaeme. 2015. Sparc: Supervised progressively autonomous robot competencies. In *International Conference on Social Robotics*. Springer, 603–612.

[161] Emmanuel Senft, Séverin Lemaignan, Paul E Baxter, Madeleine Bartlett, and Tony Belpaeme. 2019. Teaching robots social autonomy from in situ human guidance. *Science Robotics* 4, 35 (2019).

[162] Pierpaolo Sgroi, G Massimo Palma, and Mauro Paternostro. 2021. Reinforcement learning approach to nonequilibrium quantum thermodynamics. *Physical Review Letters* 126, 2 (2021), 020601.

[163] Yang Shi, Xin Yan, Xiaojuan Ma, Yongqi Lou, and Nan Cao. 2018. Designing emotional expressions of conversational states for voice assistants: Modality and engagement. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–6.

[164] Maria Shugrina, Wenjia Zhang, Fanny Chevalier, Sanja Fidler, and Karan Singh. 2019. Color Builder: A direct manipulation interface for versatile color theme authoring. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[165] Candace L Sidner, Christopher Lee, Cory D Kidd, Neal Lesh, and Charles Rich. 2005. Explorations in engagement for humans and robots. *Artificial Intelligence* 166, 1-2 (2005), 140–164.

[166] Chaklam Silpasuwanchai, Xiaojuan Ma, Hiroaki Shigemasu, and Xiangshi Ren. 2016. Developing a comprehensive engagement framework of gamification for reflective learning. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. 459–472.

[167] Patricia Smittle. 2003. Principles for effective teaching. *Journal of Developmental Education* 26, 3 (2003), 10–16.

[168] Sichao Song and Seiji Yamada. 2017. Expressing emotions through color, sound, and vibration with an appearance-constrained social robot. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*. IEEE, 2–11.

[169] Sichao Song and Seiji Yamada. 2018. Designing expressive lights and in-situ motions for robots to express emotions. In *Proceedings of the 6th International Conference on Human-Agent Interaction*. 222–228.

[170] Aaron St. Clair and Maja Mataric. 2015. How robot verbal feedback can improve team performance in human-robot task collaborations. In *Proceedings of the tenth annual acm/ieee international conference on human-robot interaction*. 213–220.

[171] Robert A Stebbins. 1971. The meaning of disorderly behavior: Teacher definitions of a classroom situation. *Sociology of Education* (1971), 217–236.

[172] Dagmar Sternad and William J Dean. 2003. Rhythmic and discrete elements in multi-joint coordination. *Brain research* 989, 2 (2003), 152–171.

[173] Mingfei Sun and Xiaojuan Ma. 2019. Adversarial imitation learning from incomplete demonstrations. *arXiv preprint arXiv:1905.12310* (2019).

[174] Mingfei Sun, Yiqing Mou, Hongwen Xie, Meng Xia, Michelle Wong, and Xiaojuan Ma. 2019. Estimating Emotional Intensity from Body Poses for Human-Robot Interaction. *arXiv preprint arXiv:1904.09435* (2019).

[175] Mingfei Sun, Zhenhui Peng, Meng Xia, and Xiaojuan Ma. 2022. Investigating the Effects of Robot Engagement Communication on Learning from Demonstration. *International Journal of Social Robotics* 14, 3 (2022), 789–806.
[176] Mingfei Sun, Zhenjie Zhao, and Xiaojuan Ma. 2017. Sensing and handling engagement dynamics in human-robot interaction involving peripheral computing devices. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 556–567.
[177] Wei Sun, Yunzhi Li, Feng Tian, Xiangmin Fan, and Hongan Wang. 2019. How Presenters Perceive and React to Audience Flow Prediction In-situ: An Explorative Study of Live Online Lectures. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–19.
[178] Dag Sverre Syrdal, Nuno Otero, and Kerstin Dautenhahn. 2008. Video prototyping in human-robot interaction: Results from a qualitative study. In *Proceedings of the 15th European conference on cognitive ergonomics: The ergonomics of cool interaction*. 1–8.
[179] Leila Takayama, Doug Dooley, and Wendy Ju. 2011. Expressing thought: improving robot readability with animation principles. In *Proceedings of the 6th international conference on Human-robot interaction*. 69–76.
[180] Kazunori Terada, Atsushi Yamauchi, and Akira Ito. 2012. Artificial emotion expression for a robot by dynamic color change. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 314–321.
[181] Vicki Trowler. 2010. Student engagement literature review. *The higher education academy* 11, 1 (2010), 1–15.
[182] Chun-Hua Tsai and Peter Brusilovsky. 2019. Evaluating visual explanations for similarity-based recommendations: User perception and performance. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. 22–30.
[183] Chun-Hua Tsai and Peter Brusilovsky. 2019. Explaining recommendations in an interactive hybrid social recommender. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 391–396.
[184] Rick B Van Baaren, Rob W Holland, Kerry Kawakami, and Ad Van Knippenberg. 2004. Mimicry and prosocial behavior. *Psychological science* 15, 1 (2004), 71–74.
[185] Jur Van Den Berg, Stephen Miller, Daniel Duckworth, Humphrey Hu, Andrew Wan, Xiao-Yu Fu, Ken Goldberg, and Pieter Abbeel. 2010. Superhuman performance of surgical tasks by robots using iterative learning from human-guided demonstrations. In *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2074–2081.
[186] Jennifer J Vasterling, Lisa M Duke, Kevin Brailey, Joseph I Constans, Albert N Allain Jr, and Patricia B Sutker. 2002. Attention, learning, and memory performances and intellectual resources in Vietnam veterans: PTSD and no disorder comparisons. *Neuropsychology* 16, 1 (2002), 5.
[187] Lev Semenovich Vygotsky. 1980. *Mind in society: The development of higher psychological processes*. Harvard university press.
[188] Emily Wall, Leslie Blaha, Celeste Paul, and Alex Endert. 2019. A formative study of interactive bias metrics in visual analytics using anchoring bias. In *IFIP Conference on Human-Computer Interaction*. Springer, 555–575.
[189] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y Lim. 2019. Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–15.
[190] Daniel Weitekamp, Erik Harpstead, and Ken R Koedinger. 2020. An interaction design for machine teaching to develop AI tutors. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–11.
[191] Eric Wiewiora, Garrison W Cottrell, and Charles Elkan. 2003. Principled methods for advising reinforcement learning agents. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*. 792–799.
[192] Grant Wiggins. 1998. *Educative Assessment. Designing Assessments To Inform and Improve Student Performance.* ERIC.
[193] Katie Winkle, Emmanuel Senft, and Séverin Lemaignan. 2021. LEADOR: A Method for End-to-End Participatory Design of Autonomous Social Robots. *arXiv preprint arXiv:2105.01910* (2021).
[194] Sarah Woods, Michael Walters, Kheng Lee Koay, and Kerstin Dautenhahn. 2006. Comparing human robot interaction scenarios using live and video based methods: towards a novel methodological approach. In *9th IEEE International Workshop on Advanced Motion Control, 2006.* IEEE, 750–755.
[195] Ziming Wu, Yulun Jiang, Yiding Liu, and Xiaojuan Ma. 2020. Predicting and diagnosing user engagement with mobile ui animation via a data-driven approach. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
[196] Seiji Yamada, Kazunori Terada, Kazuki Kobayashi, Takanori Komatsu, Kotaro Funakoshi, and Mikio Nakano. 2013. Expressing a robot's confidence with motion-based artificial subtle expressions. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 1023–1028.
[197] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–13.
[198] Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. 2019. Understanding the effect of accuracy on trust in machine learning models. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–12.
[199] Yunfeng Zhang, Q Vera Liao, and Rachel KE Bellamy. 2020. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 295–305.