Choosing Robot Feedback Style to Optimize Human Exercise Performance*

Roshni Kaushik
Robotics Institute
Carnegie Mellon University
Pittsburgh, USA
roshnika@andrew.cmu.edu

Rayna Hata
Robotics Institute
Carnegie Mellon University
Pittsburgh, USA
rhata@andrew.cmu.edu

Aaron Steinfeld
Robotics Institute
Carnegie Mellon University
Pittsburgh, USA
steinfeld@cmu.edu

Reid Simmons
Robotics Institute
Carnegie Mellon University
Pittsburgh, USA
rsimmons@andrew.cmu.edu

***Abstract*—Different people respond to feedback and guidance in different ways, and their preferences may change based on their mood, tiredness, etc. We present a robot exercise coach that provides verbal and nonverbal feedback in two different styles: firm and encouraging. We collect a dataset of people experiencing both feedback styles and show that the style that someone performs best with may not be the one they have the best subjective experience with or be the one that they state they prefer. To account for this, we present a contextual bandit approach that enables the robot coach to learn the best style to use over time to improve the human's performance, and show that this approach performs quite well in expectation on the real human data.**

***Index Terms*—personalization; adaptation; exercise; robot; coaching; feedback**

I. Introduction

One of the key objectives of human-robot interaction is to develop robots that can seamlessly integrate into human social environments. As robots are increasingly being deployed across various domains, transitioning from controlled factory settings to more intricate environments like healthcare and education, it becomes essential for them to be contextually-aware and tailor their interactions by using verbal and nonverbal modalities. A failure to personalize interactions and instead rely on a one-size-fits-all strategy may result in missed opportunities to enhance individual performance and experiences based on preferences and context.

A crucial aspect of this process involves enabling robots to understand individuals' preferences for feedback. People can respond differently to feedback, and these preferences can shift depending on factors like mood, personality, fatigue, etc.

In the domain of exercise, guidance and feedback play a crucial role, as a personalized coach can offer corrections to help ensure that exercises are performed correctly, minimizing the risk of injury and maximizing the exercise effectiveness. A coach can also provide motivation and encouragement to improve consistency and make the exercise experience more enjoyable. People may have different feedback style preferences for exercise; some individuals may favor a firmer approach, while others may prefer more encouragement. Exercise feedback preferences can be impacted by a variety of factors as shown in [1]. In this paper, we explore the following research question:

- How can an exercise coach robot determine which feedback style to use when in real-time?

To explore this question, we use an existing robot exercise coach framework that can analyze the human's exercising in real-time to provide specific feedback (e.g., corrections, encouragement). The coach can provide that feedback in different styles, specifically firm and encouraging, and our prior work has shown that people have different performances and preferences with these two styles [2]. We then present a contextual bandit approach for the robot to learn the appropriate feedback style to use for an individual. This approach incorporates a context at each time-step, predicts the best feedback style to use, uses that feedback style to react in a multi-modal way, and then observes the reward (human's performance) after the feedback to train the bandit.

To explore this approach with real human data, we created a dataset of people exercising with the two feedback styles. We first obtained data from the previous work where younger participants interacted with the robot [2]. To create a more diverse dataset, we conducted a new user study with older adults interacting with the robot, as we believe that they may respond differently to the two feedback styles.

On this combined dataset, we show that a robot cannot simply ask for a style preference or ask Likert-style questions about subjective experience with the styles to determine the best style to use to optimize for performance. We also show that older adults do respond to these styles differently, justifying our data collection with this population to create a more age-diverse dataset. Lastly, we use our contextual bandit approach with the human data in this dataset and show that our adaptive approach results in a better expected performance than simply choosing one style for each exercise round.

II. Related Work

A. Human Feedback

In human-human interactions, people use both verbal and nonverbal feedback and have different task performances and experiences. When speech was accompanied by gestures, people interpreted it differently as the nonverbal behavior helped

*NSF Grant #2112633

to disambiguate the verbal feedback [3]. This motivates our use of both verbal and nonverbal feedback for the robot exercise coach. In the exercise domain, nonverbal behavior was shown to be crucial in coach-athlete communication, which motivates the use of nonverbal robot behavior in our work [4].

Humans also can react differently to feedback given in different ways, motivating our use of multiple feedback styles. Researchers found that students had different preferences for how teachers should give different types of feedback [5]. Researchers also found that people generally prefer process-based criticism over personal criticism [6]. Choosing the correct feedback style can even have an impact on task performance. A group that received error feedback outperformed groups that received other types of feedback in learning how to vertical jump [7]. We attempt to leverage this effect in our work by determining how to choose the correct style of feedback for each person. In the exercise domain specifically, feedback preferences can be impacted by a variety of factors as shown in [1], including physical health status, educational level.

B. Robot Feedback Improving Performance

In this work, we explore how robot feedback can positively impact performance. Our prior work indicates that nonverbal robot behavior improves task performance in a sorting game task [8]. Robot nonverbal behavior was shown in improve task performance for difficult collaborative tasks [9]. Robot gestures reduced perceived workload and improved task performance, making difficult tasks feel easier [10]. In the domain of exercise, a robot coach was shown to reduce mistakes using a combination of verbal and nonverbal modalities [11]. Additionally, a robot's gaze improved the human's performance in a cooperative task [12].

C. Robot Feedback Improving Human Subjective Experience

Although the focus in our work is how to design robot feedback to improve performance, a robot coach's feedback can also have an impact on the human's subjective experience. Researchers developed a model that expressed different levels of competence and warmth with the NAO robot, changing its hand and body movements [13]. A robot programmed to have a positive mood increased participants' valence and arousal [14]. Additionally, the combination of verbal and nonverbal feedback was found to improve the human's experience [15].

Our work shows how changing the way the robot gives feedback can change the human's performance and their experience. Our prior work showed that people do respond differently to two different feedback styles (firm and encouraging) on the robot exercise coach [2].

D. Personalization

Researchers have explored how to personalize robot behavior based on many different factors, including personality. Participants in a study varied their preferred distance to the robot based on various personality traits, such as proactivity [16]. Researchers adapted a robot's verbal and nonverbal behavior based on the human's extroversion [17]. A robot personalized to the human maintained a higher level of engagement and motivation during the exercise session [18]. Additionally, students that interacted with a robot with personalized feedback showed a significant increase in emotional response [19]. These studies illustrate the benefit of personalization, but generally do not adapt the robot's behavior based on real-time feedback. Our work attempts to improve the human's performance using real-time information from the human to learn the best styles to use as they exercise.

III. Methods

We utilized Quori [20] as the exercise coach, where the human performed exercises in front of the robot (Figure 1) and the robot provided verbal and nonverbal feedback that could include corrections and encouragement. For the robot to appropriately provide feedback, it first needed to analyze the human's exercise in real-time and then provide feedback based on its exercise evaluation in the two different feedback styles: firm and encouraging.

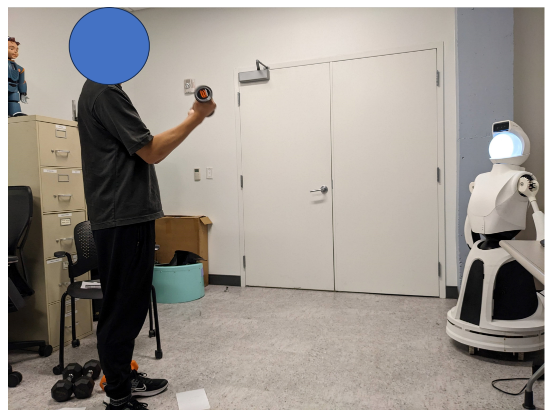

Fig. 1. Human exercising with the robot exercise coach

A. Exercise Evaluation and Feedback

Because we planned to combine the datasets, we utilized the same exercise evaluation and feedback methodology as our prior work with younger participants [2]. A summary of the approach is included in this section for completeness. For the robot to provide relevant, real-time feedback, it needs to analyze the human's exercises. We chose two exercises, bicep curls and lateral raises, because they are upper body exercises (allowing participants to sit while exercising and expanding the potential participant pool to those with limited lower body mobility). The utility of these exercises was also validated by two domain experts[1].

1) Segmentation: The first step to providing feedback is to determine where one repetition starts and another ends. We process the images from the camera mounted on the robot using the Mediapipe library[2] to get 3D joint positions, and

[1] Ayotoni Aroyo, ACSM-CPT (Exercise Physiologist and Physical Activity Lead at Emory University's Cognitive Empowerment Program) and Gustavo J Almeida, PT, Ph.D. (UT Health San Antonio)

[2] https://pypi.org/project/mediapipe/

TABLE I

EXAMPLES OF VERBAL AND NONVERBAL FEEDBACK TO DIFFERENT SITUATIONS FOR THE THREE FEEDBACK STYLES THE NEUTRAL STYLE DOES NOT HAVE VERBAL FEEDBACK WHILE THE HUMAN IS EXERCISING.

	Firm		Encouraging	
Evaluation	**Verbal**	**Nonverbal**	**Verbal**	**Nonverbal**
Last 2 reps slow	Try to speed up	50% sad, lean forward slightly	Nice job, can you speed up a little on the next few?	50% sad, lean forward slightly less than firm
Last 2 reps low range of motion	Focus on getting a full range of motion in your elbows	50% sad, lean forward slightly	You are doing great, try to get a full range of motion in your elbows.	50% sad, lean forward slightly less than firm
Last 2 reps good speed, previous 2 were slow	Nice speed, keep going	60% happy, small upward arms, small backward torso	Nice job, great speed!	90% happy, large upward arms, large backward torso

we then compute angles of interest for our exercises: right/left shoulder and right/left elbow. Looking at the angles, we can determine conditions for when a rep ends. For example, when the gradient jumps and the value of the elbow angle changes sign, this could indicate that the upper arm is moving back up to start a new bicep curl. Figure 2 illustrates the segmentation applied to an example set of bicep curl angles.

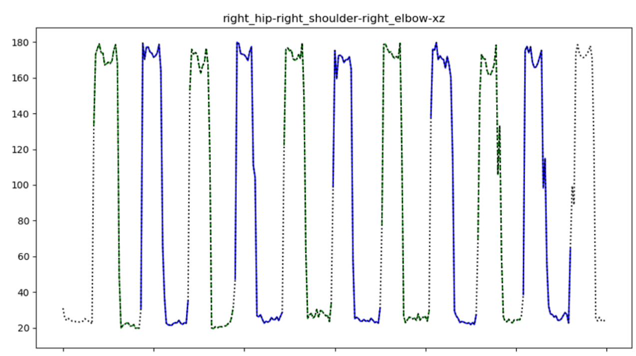

Fig. 2. One of the bicep curl angles segmented into reps with the alternating solid blue and dashed green colors indicating where the segmentation method found the end of one rep and the beginning of another

2) Comparison: The second step is to compare what the person has done to previously recorded demonstrations of the exercises, allowing us to determine *how well* a rep was performed. We record both properly done demonstrations and examples of common mistakes for each exercise (such as not raising your arm to 90° on a lateral raise). We then compare the rep to the recorded demonstrations using the Dynamic Time Warping [21] distance to determine which demonstration is closest. If the rep is 'close enough' to one of the recorded demonstrations, we assign the rep the evaluation of that demonstration (e.g. good form, low range of motion). If the rep does not match closely enough to any of the demonstrations, then the evaluation is bad form. This evaluation is then used to provide multi-modal feedback to the person.

3) Multi-modal Feedback: Once the robot evaluates the quality of each rep, it reacts in a multi-modal way. One mode is to react verbally, providing congratulations of good form, suggestions to correct bad form, and positive reinforcement after correcting bad form. The robot, in general, reacts after it sees a pattern (e.g. three good form rep in a row), and our prior work explores how the frequency of feedback affects how it is viewed [22]. We used two feedback styles explored in our prior work that were developed in conjunction with domain experts: firm and encouraging [2]. The firm style has minimal encouragement, while the encouraging style includes positive statements to soften the blow of corrective feedback.

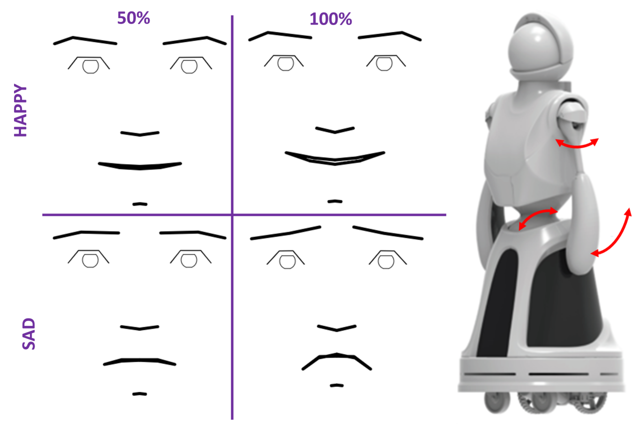

Fig. 3. Examples of robot's happy and sad facial expressions at different intensities (left) and its degrees of freedom for body movements (right)

In addition to the verbal mode, the robot also reacts non-verbally (Figure 3). In positive situations, the robot smiles and performs a happy body movement, while it slightly frowns and performs a sad body movement to accompany corrections. When using the encouraging style, the robot's positive reactions are more intense and its negative reactions are less intense, compared to the firm style. Our prior work explored which movement patterns on Quori were perceived as displaying different emotions [23], and we use the movements perceived as happy and sad for Quori's nonverbal reactions. Nonverbal feedback occurs in conjunction with the verbal feedback (e.g., a corrective verbal phrase paired with a negative nonverbal reaction) to augment the feedback signal of the verbal feedback. The robot also reacts nonverbally 50% of the time based on the rep immediately previous when there is no verbal utterance by the robot, which previous work found was perceived positively [22]. In general, people preferred a higher frequency of nonverbal feedback as nonverbal signals can be less overwhelming and require less mental energy to process, and the nonverbal robot reactions can enhance the human's subjective experience during the interaction.

Table I includes multiple examples of how the robot reacts in different situations in a multi-modal way and with the different feedback styles.

B. Adaptive Feedback Model

We introduce an adaptive feedback model using a contextual bandit for a robot to learn in real-time which feedback style to use when. The feedback style someone might prefer may be dependent on a variety of factors, and we include fatigue estimation in this work as an important contextual feature, since someone's preference for encouragement may vary based on how tired they are.

In this approach (Figure 4), the robot first observes a context from the human. For our work, we assume this context c is a fatigue estimate, where the robot estimates whether the human has low, moderate, or high fatigue. Next, the robot queries its policy Π for the best action (feedback style) given the current context. The chosen action a is then used to generate multi-model feedback in the chosen style. The human observes this feedback and performs the next rep of the exercise with either good or bad form. The robot can observe this reward (0 or 1) and trains on the combination of context, action, and reward to improve the policy. The goal of this approach is for the robot to choose the action that maximizes the human performance given the context.

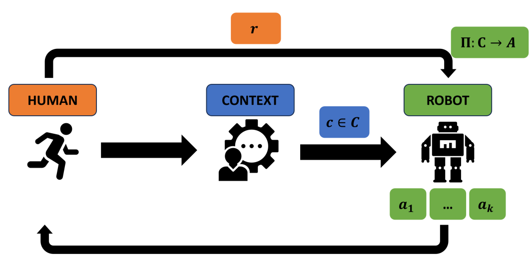

Fig. 4. The robot observes the context c from the human and chooses an action a. The human responds to the robot's action, and their performance r forms the reward for the robot to train policy Π.

When translating this approach to the robot exercise coach, we note that the robot does not provide verbal feedback at every repetition (only reacting when a pattern is observed, such as 3 good reps in a row). Reacting verbally every rep would be very overwhelming for the human, and they would not be able to process the feedback at such a high frequency. In practice, we assume that any reps following a verbal utterance with a particular style have the same style (e.g., an encouraging phrase has an effect for the few reps after that phrase is uttered), and the style of the robot can change when the conditions have been met for the next verbal utterance. This will allow the robot to adapt to the human's performances and will allow the human to react to the robot's feedback. In the simulation formulation, however, we assume the robot can change its feedback style every repetition.

We developed a human simulation and fatigue model to determine whether the contextual bandit can learn the feedback style with which the simulated human performs best.

First, we created a model of fatigue as the human is exercising; this is the context the robot will observe. We simulate an exercise session where the human performs a series of sets of 10 reps, where the fatigue is low for the first 5 reps, moderate for the next 3 reps, and high for the last 2 reps. This approximates a human whose fatigue increases as they perform each set and resets back to low fatigue after a rest period following each set.

Next we determine the feedback styles, or actions, the robot can choose between. We chose five different feedback styles: very firm, firm, neutral, encouraging, and very encouraging. We have thus far implemented two of these styles (firm and encouraging) on the actual robot, but we wanted to explore a larger range of styles in simulation to see if the bandit approach can still learn the best style in a more complicated scenario.

The bandit learns a policy on each of these feedback styles as a function of the context observed, and it treats each of the styles as independent and unrelated. The styles, however, are not truly independent, as very firm is a more extreme version of firm and very encouraging a more extreme version of encouraging. We did attempt to update the policy of related actions after viewing the reward, similar to the pseudo reward approach presented in [24]. For example, a reward after very firm feedback may tell us something about the firm action, but the pseudo rewards did not significantly improve our results. Therefore, we treat each of the styles as independent.

Lastly, we chose to construct a reward that is simply based on performance: 0 for bad form and 1 for good form. Our simulated human model performs the next rep correctly based on a probability $p(a_k, f)$, where a_k is the feedback style chosen by the robot after each rep. We construct the following equation to compute this probability:

$$p(a_k, f) = p(a_k)(1 - \gamma f e(a^r)) \quad (1)$$

- $p(a_k, f)$ is the probability the human performs the next rep correctly after viewing feedback style a_k with fatigue f
- $p(a_k)$ is the probability the human performs the next rep correctly with no fatigue after seeing feedback style a_k
- γ is 0 if there is no fatigue dependence in performance, and 1 if there is a large fatigue dependence
- f is 0, 0.3, or 0.6 for low, medium, and high fatigue, respectively
- $e(a_k)$ is an action-dependent fatigue factor (e.g., whether high fatigue reduces performance with a specific feedback style)

For example, if a human performs very well with the very firm style, then perhaps $p(a_0) = 0.8$, meaning without fatigue, they perform 80% of their reps with good form after seeing very firm feedback. If they have a high fatigue dependence

($\gamma = 1$), then the final term in the equation plays a large factor in their true performance with very firm feedback. If their fatigue was high ($f = 0.6$) but they had no action dependent fatigue factor ($e(a^0) = 1$), their performance $p(a_0, f)$ with high fatigue would become $0.8(1-0.4) = 0.48$, which is a reduction to 48% of their reps correct with good form after seeing very firm feedback with high fatigue. However, if they had an action dependent fatigue factor of ($e(a_0) = 0.5$), which means their preference for very firm feedback increases with higher fatigue, their new performance $p(a_0, f) = 0.8(1-(0.4 \times 0.5)) = 0.64$. This means that since their preference for very firm feedback increases with fatigue, their drop in performance with high fatigue is less severe (64% instead of 48%).

C. Simulation Results

We ran this simulation in three different scenarios. Experiment 1 explores no fatigue dependence, where the human's performance does not change with fatigue ($\gamma = 0, p(a^r, f) = p(a^r)$). Experiment 2 explores basic fatigue dependence, where the human's performance reduces with fatigue, but the bandit's actions do not change with fatigue ($\gamma = 0.5, e = [1, 1, 1, 1, 1]$). Experiment 3 is the most complicated scenario of complex fatigue dependence, where the action the bandit should choose changes with the different fatigue levels ($\gamma = 0.5, e = [4, 2, 1, -0.25, -2]$). This will demonstrate the efficacy of this approach in learning the optimal feedback style in increasingly complicated scenarios. We implement the contextual bandit model using the *bayesianbandits*[3] package.

For all these experiments, let us consider someone who performs best with very firm feedback and performs worst with very encouraging feedback. Specifically, $a^r = [0.8, 0.6, 0.5, 0.4, 0.3]$, where the human performs 80% of their reps correctly without taking fatigue into account after viewing very firm feedback, 60% with firm, etc.

1) Experiment 1 (No Fatigue Dependence): We first ran an experiment over 20 sets of 10 reps each where the human's performance is not fatigue dependent ($\gamma = 0, p(a^r, f) = p(a^r)$). The bandit's optimal action for all reps is the very firm action. Figure 5 shows that the bandit chose the very firm action 77% of the time over the 20 sets, and learned by the later sets to choose very firm for all the reps.

2) Experiment 2 (Basic Fatigue Dependence): We next ran an experiment over 20 sets of 10 reps each where the human's performance reduces with fatigue, but the bandit's actions do not change with fatigue ($\gamma = 0.5, e = [1, 1, 1, 1, 1]$). The human's fatigue does reduce their performance, but it reduces it equally across all actions and the optimal action for the bandit is still very firm across all fatigue levels. In this experiment, the bandit chose the very firm action 62% of the time time over the 20 sets, which is less than the no fatigue case, but still very frequent.

3) Experiment 3 (Complex Fatigue Dependence): We lastly ran an experiment over 20 sets of 10 reps each where the

[3]https://bayesianbandits.readthedocs.io/en/latest/index.html

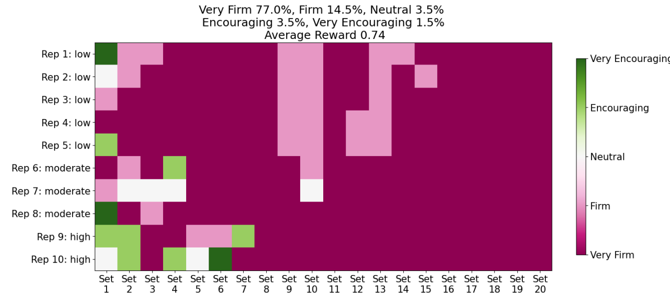

Fig. 5. Experiment 1: Simulation results with no human fatigue dependence in performance. The human's performance is the same across all fatigue levels, and the optimal action for the bandit is very firm across all reps.

action the bandit should choose changes with the different fatigue levels ($\gamma = 0.5, e = [4, 2, 1, -0.25, -2]$), with the probabilities bounded between $[0.05, 0.95]$. The bandit should optimally choose very firm for low fatigue, is relatively indifferent between all styles for moderate fatigue (probabilities range from 39% - 43%), and should choose encouraging or very encouraging for high fatigue. This extreme case of fatigue dependence illustrates how the style the robot should choose could completely change as the fatigue changes.

Figure 6 shows that the bandit shows that the bandit is able still learn the optimal action in the complex fatigue scenario. It learns to choose very firm for low fatigue and very encouraging for high fatigue. Additionally, if the bandit did not see the context, it would perform much worse (average reward of 0.45 in addition to choosing the incorrect action for moderate and low fatigue compared to 0.64 when taking the context into account). This validates our use of a contextual bandit, rather than a simple bandit.

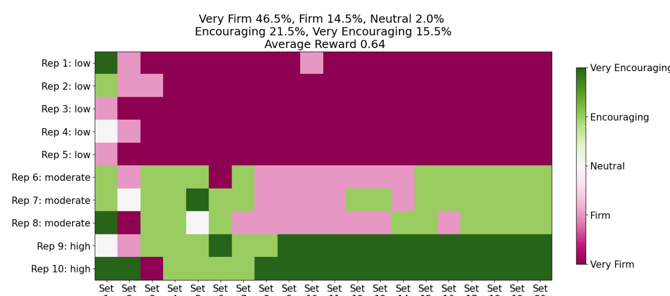

Fig. 6. Experiment 3: Simulation results with complex fatigue dependence in performance. The bandit should choose very firm for low fatigue, is relatively indifferent between all styles for moderate fatigue, and should choose encouraging or very encouraging for high fatigue.

In these experiments, we show that the contextual bandit can accurately learn the optimal feedback style even with complex fatigue dependence.

IV. User Study Dataset

A. Study Design

In our previous work conducted in a laboratory setting [2], 19 participants performed two rounds of exercise of 4 sets each (2 bicep curls and 2 lateral raises). After each round, the participant completed a short survey. The survey had questions taken from the Godspeed Questionnaire [25] to measure their perception of the robot in terms of animacy, likability, and perceived intelligence. Most of these 19 participants were young adults ($\mu = 28.5, \sigma = 14.7$, with one participant ≥ 65).

We implemented the same study procedures with our older adult study: two counterbalanced rounds of exercise, one with each feedback style, and surveys with the same questions. In this study, we took Quori to an assisted living facility (Vincentian Schenley Gardens in Pittsburgh, PA) and ran the study protocol with 8 older adults there. We also added two more participants in a laboratory setting, resulting in a total of 10 participants ($\mu = 77.4, \sigma = 10.6$) with 9 participants ≥ 65.

Combining the two sets of data, we have a total of 29 participants, 10 of which are 65 or older. We aimed to create a more age-diverse dataset as most traditional recruiting methods skew younger in age and exercise has value across all age ranges. Even with a specific effort to add older adults, we still have an unbalanced dataset, but it is much more balanced than the original dataset (34% older adults rather than 5%).

There are two major differences between these two studies. The first is that in the older adult study, we asked the participants before beginning the exercise rounds whether they preferred a firm or encouraging feedback style. As we will show in Section IV-B, the style that participants state before the interaction and the style that they prefer (rate higher on the surveys) do not always match the style they perform best with. This highlights the need for an adaptive approach to learn in an online way which style each person performs best with.

The second difference is that in the older adult study, we estimated fatigue using a heart rate monitor (Polar Verity Sense), calculating the heart rate reserve using the following equation[4]:

$$HRR = \frac{HR - RHR}{MaxHR - RHR} \tag{2}$$

- HRR is the heart rate reserve, where we set less than 0.2 to be low fatigue, 0.2-0.4 to be moderate fatigue, and 0.4 and above to be high fatigue after pilot testing
- HR is the participant's current heart rate in beats per minute
- RHR is the participant's resting heart rate, computed as an average during the introduction to the exercise session
- $MaxHR$ is the participant's estimated maximum heart rate, computed using $(220 - Age)$[5]

which allows us to determine how high an individual's heart rate is, proportional to their resting and max heart rates. [?] explores the relationship between physical fatigue and heart rate metrics, including heart rate reserve.

We do not have fatigue information for the 19 participants from the first study, so after observing the data from the older adult study, we estimated that participants have low fatigue for the first 90% of each set and moderate fatigue for the remainder of each set. There are, of course, individual variations that we observed, but this fatigue estimate was used in the contextual bandit approach in Section V.

[4]https://my.clevelandclinic.org/health/articles/24649-heart-rate-reserve

[5]https://www.heart.org/en/healthy-living/fitness/fitness-basics/target-heart-rates

B. Study Results

We first assigned a performance group to each of the 29 participants: those who performed better with the encouraging style, those who performed better with the firm style, and those who performed about the same with the two styles. To do this, we computed the difference d in performance (percentage of good form reps) between the two styles for each participant (encouraging - firm). We then calculated the mean and standard deviation of those differences. To compute a 95% confidence interval, we used the formula $\mu \pm (t \times \sigma)$, where t is the critical value from the t-distribution and σ is the standard error (standard deviation divided by $\sqrt{n}$).

We also computed a subjective measure for each participant from their survey responses after experiencing each feedback style. We averaged the 1-7 Likert scores that each participant completed for each robot style (lively, interactive, responsive, friendly, kind, pleasant, competent, intelligent). We subtracted the firm score from the encouraging one to get a single value where < 0 indicates a preference for firm and > 0 a preference for encouraging. We then performed the same grouping procedure as for the performance scores to determine which participants prefer the firm style, prefer the encouraging style, or have no style preference.

Table II includes the number of participants with each combination of performance and preference group. We can see that 76% of the participants lie off the diagonal, where the diagonal indicates agreement between the style someone prefers and the style they perform best with. Discounting the groups with no preference or performance difference, 70% of the remaining participants had complete disagreement in group assignment (e.g., preferring the firm style but performing best with the encouraging style).

TABLE II
PERFORMANCE AND PREFERENCE GROUPS FOR THE 29 PARTICIPANTS. THE DIAGONAL INDICATES AGREEMENT BETWEEN THE STYLE PREFERENCE AND THE STYLE THE PARTICIPANT PERFORMS BEST WITH.

	Prefers Firm	**Prefers Encouraging**	**No Style Preference**
Performs Best with Firm	1	2	6
Performs Best with Encouraging	5	2	2
Performs Equally with Both	3	4	4

For the 10 participants in the older adult study, we additionally have the participants' stated style to compare to the style they perform best with. Table III shows the performance groups separated out by what the participants stated as their style preference. We can see that 3/10 participants stated the style that they performed best with, but 4/10 of the participants stated the opposite style to the one they performed best with. This further supports our claim that simply asking people their preference does not always help the robot choose the feedback style to optimize performance. Some potential reasons for the discrepancy include limited exercise experience, unfamiliarity with feedback styles, or selecting responses based on perceived rather than actual preferences.

TABLE III
PERFORMANCE AND STATED STYLE PREFERENCE FOR 10 PARTICIPANTS IN THE OLDER ADULT STUDY

	Stated Firm	Stated Encouraging
Performs Best with Firm	2	4
Performs Best with Encouraging	0	1
Performs Equally with Both	0	3

We can also compared how older adults performed with different feedback styles compared to those less than 65 years of age. We can see in Table IV that the distribution of performance groups is statistically different ($p < 0.05$ after running an ANOVA) between adults (19) and older adults (10). In particular, a greater percentage of older adults seemed to perform better with the firm feedback style, demonstrating the importance of collecting data from multiple age groups.

TABLE IV
FEEDBACK STYLE THAT PARTICIPANTS PERFORMED BEST WITH, SPLIT BY AGE GROUP.

	Adult (<65)	Older Adult (>= 65)
Performs Best with Firm	15.8%	60%
Performs Best with Encouraging	36.8%	20%
Performs Equally with Both	47.4%	20%

V. ADAPTIVE FEEDBACK ON HUMAN DATA

Section III-C explored the use of a contextual bandit approach using a complex simulated human model, but we now want to test the efficacy of the approach on our dataset with real human data. Using the data from the 29 participants described in Section IV, we set the context in those data to be the estimated fatigue (low, moderate, or high). The model has the choice of two feedback styles the humans experienced: firm or encouraging. In our studies, participants experienced one round of exercise with the firm style and one with the encouraging style, and we additionally have their performance on each rep throughout the exercise session.

The goal of this test on real data is to see the actions the contextual bandit *would have* chosen given the observed context, and then estimate the reward it *would have* received based on the participants' performance. We can then compare that estimated reward to the participants' actual performance when they experienced one round of each feedback style.

Let us assume we have Participant X who performed two rounds of exercise, one with each style. They performed 40 repetitions with the firm style first, followed by 35 repetitions with the encouraging style. Note that the participants did not perform the same number of reps with each style since rounds are time-based. They performed 80% of reps with the firm robot correctly and 70% with the encouraging robot correctly.

We then start with the first repetition that they performed, which was with the firm style, and let us assume the first repetition had good form. We give the model the context, which for the first repetition was low fatigue. We query the model as to the action it thought it should take, given the context. Let us assume the model chooses the encouraging style. We now train the model on what the human actually experienced (on the true data, not the action the model chose); specifically, the (context, action, reward) of (low fatigue, firm, good performance). We then continue this process throughout the rest of the repetitions. For each repetition, we compute the expected reward that the robot would have received had it chosen the style outputted by the bandit. In this example, we set the expected reward received by the bandit for this rep as 0.7 since the human is expected to perform 70% of the repetitions correctly with the encouraging style (the action chosen by the bandit). Note we train the bandit on the true data (firm) and compute the bandit's expected reward based on the bandit's chosen action (encouraging).

We computed two different metrics throughout this approach. The first is the *actual reward* which is the total number of good performance reps the human actually performed through both rounds. The second is the *expected bandit reward* which is the sum of the expected rewards over all the rounds. We hypothesize that the *expected bandit reward* will be higher than the *actual reward*, since the model should be optimizing for the human's best performance.

Model Results

Figure 7 illustrates an example of this approach with one participant who performed much better with the encouraging style. The bandit quickly adapted to this performance difference and chose the encouraging style very frequently, which is optimal for this participant.

We calculated the difference in the *expected bandit reward* and the *actual reward* for each of the 29 participants (Figure 9). Since participants performed one round of exercise with each style, we can assume that we are comparing the expected reward of the bandit to a robot choosing each style for roughly half the time. For most participants (19/29) the contextual bandit had a higher expected reward. For the participant in Figure 7, the bandit had an expected reward that was 25% higher than the human's actual performance, which illustrates how an adaptive approach can learn the style someone performs best with and use that style more frequently to optimize for the human performance. Additionally, if we compare the bandit's performance with no context (simple bandit) and fatigue as context (contextual bandit), we observe that the contextual bandit performed better than the simple bandit for

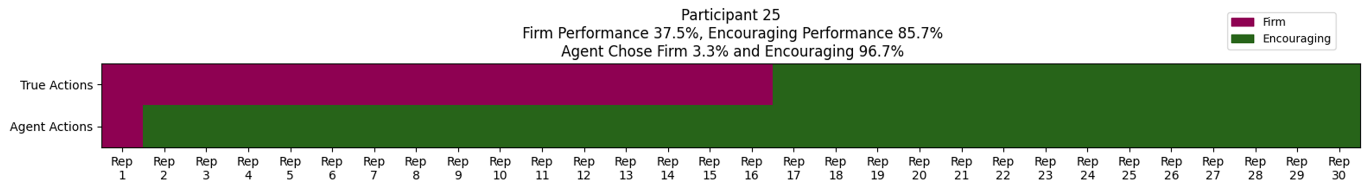

Fig. 7. Example of model results for one participant. This participant performs much better with the encouraging style (37.5% with firm, 85.7% with encouraging), which the model quickly learns by choosing the encouraging style 96.7% of the time.

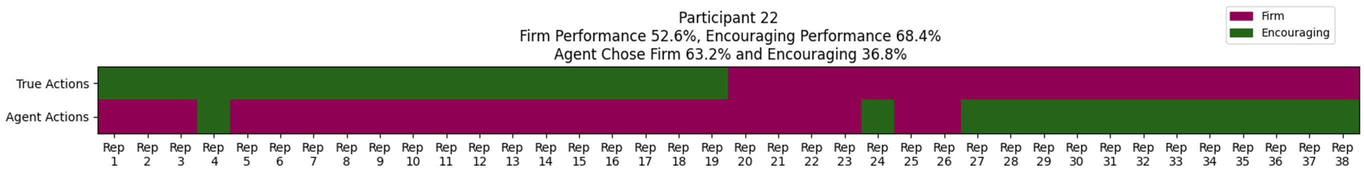

Fig. 8. Example of model results for one participant, who performed approximately equally with the two styles (52.6% with firm and 68.4% with encouraging). The model does not learn their preferences as quickly since their performance with the two styles is quite close.

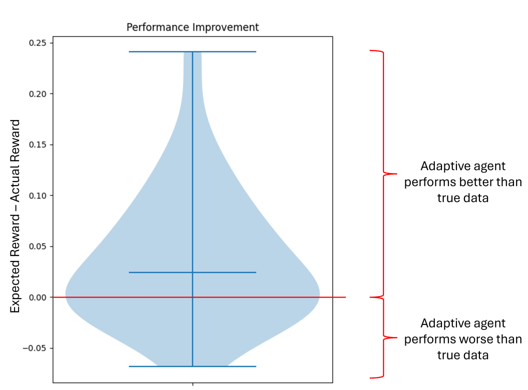

Fig. 9. Expected bandit reward - actual reward for all 29 participants. The contextual bandit outperforms the true data for most participants.

19 of the 29 participants. For the remaining 10 participants, the contextual bandit performed slightly worse or about the same, with the maximum difference being 2.5%. This further illustrates the need for informative context for the bandit to accurately choose the performance-optimizing feedback style.

For 10 participants, the contextual bandit performed slightly worse than the true data, but with the worst performance only 7% worse than the true data. Examining these participants to investigate the reasons for the bandit not performing as well, we can see that this occurs when the human's performance with the two styles is very close. This means that the expected reward the bandit received at each trial is approximately the same, and that combined with a limited number of iterations over which to learn (e.g., 38 reps instead of the 200 we performed in simulation) caused the bandit to not learn the best style to use for the human fast enough. For example, for Participant 22 (Figure 8), their performance was 53% with the firm style and 68% with the encouraging style. At each rep, the bandit was receiving 0.53 or 0.68, and for the first 20 reps, it kept switching between the two styles. For the last 12 reps, however, it began to choose the encouraging style more frequently, which is the optimal choice for this participant.

VI. Conclusion

We presented a contextual bandit approach which we test in complex performance difference situations to illustrate how the model can learn which feedback style to use when. We also presented a robot exercise coach that evaluates the human's exercises in real-time and provides multi-modal feedback in two different styles: firm and encouraging.

Next, we introduced a user study to collect data on how older adults responded to different feedback styles and combined with previous data collected to form a more age-diverse data set. We demonstrated that the style that someone **performs best** with is not always the style that they **say they prefer** and is also not always the style that they **rate the highest after experiencing**. We then showed how our contextual bandit approach for a robot exercise coach optimizes performance by learning which style someone performs best with in an online fashion, and our results indicated that this approach performs quite well in expectation.

In our future work, we will conduct a user study with both younger and older adults to test our contextual bandit approach against a static, nonadaptive baseline to verify that a coach that adapts its style to improve performance does indeed outperform a static baseline. This result will further cement the need for robots to personalize their behavior and have multiple ways of responding to the same stimuli, as humans are not a monolith and have different preferences and performances with different styles of feedback. We also want to further understand how human subjective experience will be impacted by this approach of optimizing for performance, since our dataset indicates that there can be trade-off between preference and performance.

Acknowledgment

Thanks to Nikki Kode for helping to run this study.

References

[1] N. Hirvonen, H. Enwald, P. A. Bath, R. Pyky, R. Korpelainen, and M.-L. Huotari, "Individual Factors Affecting Preferences for Feedback Message Tactics in the Contexts of Physical Activity," *Journal of Health Communication*, vol. 20, no. 2, pp. 220–229, Feb. 2015. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/10810730.2014.925015

[2] R. Kaushik and R. Simmons, "Effects of Feedback Styles on Performance and Preference for an Exercise Coach," in *2024 33st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2024.

[3] S. D. Kelly, D. J. Barr, R. B. Church, and K. Lynch, "Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory," *Journal of memory and Language*, vol. 40, no. 4, pp. 577–592, 1999, iSBN: 0749-596X Publisher: Elsevier.

[4] T. Dobrescu, "The role of non-verbal communication in the coach-athlete relationship," *Procedia-Social and Behavioral Sciences*, vol. 149, pp. 286–291, 2014, iSBN: 1877-0428 Publisher: Elsevier.

[5] P. C. Burnett, "Elementary students' preferences for teacher praise," *The Journal of Classroom Interaction*, pp. 16–23, 2001, iSBN: 0749-4025 Publisher: JSTOR.

[6] M. L. Kamins and C. S. Dweck, "Person versus process praise and criticism: Implications for contingent self-worth and coping." *Developmental psychology*, vol. 35, no. 3, p. 835, 1999, iSBN: 1939-0599 Publisher: American Psychological Association.

[7] T. Niznikowski, J. Sadowski, and A. Mastalerz, "The effectiveness of different types of verbal feedback on learning complex movement tasks," *Human movement*, vol. 14, no. 2, pp. 148–153, 2013, iSBN: 1899-1955 Publisher: Termedia.

[8] R. Kaushik and R. Simmons, "Affective Robot Behavior Improves Learning in a Sorting Game," in *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2022, pp. 436–441.

[9] H. Admoni, T. Weng, B. Hayes, and B. Scassellati, "Robot nonverbal behavior improves task performance in difficult collaborations," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2016, pp. 51–58.

[10] M. Lohse, R. Rothuis, J. Gallego-Pérez, D. E. Karreman, and V. Evers, "Robot gestures make difficult tasks easier: the impact of gestures on perceived workload and task performance," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 1459–1466, tex.ids= lohseRobotGesturesMake2014a.

[11] N. Salomons, T. Wallenstein, D. Ghose, and B. Scassellati, "The Impact of an In-Home Co-Located Robotic Coach in Helping People Make Fewer Exercise Mistakes," in *31st IEEE International Conference on Robot & Human Interactive Communication*, 2022, tex.ids= salomonsImpactInHomeCoLocated2022a.

[12] J.-D. Boucher, U. Pattacini, A. Lelong, G. Bailly, F. Elisei, S. Fagel, P. F. Dominey, and J. Ventre-Dominey, "I reach faster when I see you look: gaze effects in human–human and human–robot face-to-face cooperation," *Frontiers in neurorobotics*, vol. 6, p. 3, 2012, iSBN: 1662-5218 Publisher: Frontiers.

[13] R. Peters, J. Broekens, and M. A. Neerincx, "Robots educate in style: The effect of context and non-verbal behaviour on children's perceptions of warmth and competence," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2017, pp. 449–455.

[14] J. Xu, J. Broekens, K. Hindriks, and M. A. Neerincx, "Effects of bodily mood expression of a robotic teacher on students," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 2614–2620.

[15] M. Salem, K. Rohlfing, S. Kopp, and F. Joublin, "A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction," in *2011 Ro-Man*. IEEE, 2011, pp. 247–252, tex.ids= salemFriendlyGestureInvestigating2011a.

[16] M. L. Walters, K. Dautenhahn, R. Te Boekhorst, K. L. Koay, C. Kaouri, S. Woods, C. Nehaniv, D. Lee, and I. Werry, "The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment," in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* IEEE, 2005, pp. 347–352.

[17] A. Aly and A. Tapus, "Towards an intelligent system for generating an adapted verbal and nonverbal combined behavior in human–robot interaction," *Autonomous Robots*, vol. 40, no. 2, pp. 193–209, 2016, iSBN: 0929-5593 Publisher: Springer.

[18] B. Irfan, N. Céspedes, J. Casas, E. Senft, L. F. Gutiérrez, M. Rincon-Roncancio, C. A. Cifuentes, T. Belpaeme, and M. Múnera, "Personalised socially assistive robot for cardiac rehabilitation: Critical reflections on long-term interactions in the real world," *User Modeling and User-Adapted Interaction*, pp. 1–48, 2022, iSBN: 1573-1391 Publisher: Springer.

[19] G. Gordon, S. Spaulding, J. K. Westlund, J. J. Lee, L. Plummer, M. Martinez, M. Das, and C. Breazeal, "Affective personalization of a social robot tutor for children's second language skills," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[20] A. Specian, R. Mead, S. Kim, M. Matarić, and M. Yim, "Quori: A community-informed design of a socially interactive humanoid robot," *IEEE Transactions on Robotics*, vol. 38, no. 3, pp. 1755–1772, 2021, publisher: IEEE.

[21] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intelligent Data Analysis*, vol. 11, no. 5, pp. 561–580, 2007, iSBN: 1088-467X Publisher: IOS Press.

[22] R. Kaushik and R. Simmons, "Older Adults' Preferences for Feedback Cadence from an Exercise Coach Robot," in *2024 33st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2024.

[23] ——, "Perception of Emotion in Torso and Arm Movements on Humanoid Robot Quori," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021, pp. 62–66.

[24] K.-C. Chou, H.-T. Lin, C.-K. Chiang, and C.-J. Lu, "Pseudo-reward algorithms for contextual bandits with linear payoff functions," in *Asian Conference on Machine Learning*. PMLR, 2015, pp. 344–359.

[25] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, no. 1, pp. 71–81, 2009, iSBN: 1875-4805 Publisher: Springer.