2022 31st IEEE International Conference on Robot and
Human Interactive Communication (RO-MAN)
August 29 - September 2, 2022. Naples, Italy.

2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) | 978-1-7281-8859-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/RO-MAN53752.2022.9900654

Affective Robot Behavior Improves Learning in a Sorting Game*

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

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

***Abstract*—Nonverbal communication in the field of education can allow teachers to emotionally support their students and improve educational experience and performance. Robot nonverbal movements have been shown to improve both subjective experiences and task performance, and this work investigates whether affective robot behavior can improve human learning. This is tested using an online sorting game where players learn easy or difficult rules, aided by robot feedback videos that contain either neutral or affective movements. Results indicate that affective robot behavior improves learning of the sorting rules and reduces the perceived difficulty of the task. Extensions include expanding the features used to determine the robot feedback and increasing the possible robot motions to create a rich set of robot feedback options to personalize the education experience further for the student.**

***Index Terms*—emotion, learning, nonverbal, task performance**

I. Introduction

When developing social robots for education, there has been a focus on two types of outcomes: cognitive and affective [1]. The bulk of existing work ($\approx 66\%$ in this meta-analysis) has been focused on affective outcomes, or qualities such as attentiveness, engagement, and reflectiveness, while less focus has been given to cognitive outcomes, which include measures of knowledge.

One way robots could impact these cognitive and affective outcomes is through their nonverbal behaviors. People use nonverbal behavior to convey important information during communication [2]. In education, teachers can change their nonverbal immediacy, defined as the degree of perceived physical or psychological closeness between people, using different nonverbal behaviors. A teacher smiling and leaning in to explain a concept to a student will have a very different impact compared to a teacher crossing their arms and frowning while communicating the same information. Changing nonverbal immediacy can even affect the student's cognitive learning or performance, and having a teacher high in nonverbal immediacy was associated with an increase in test performance [3].

In particular, our work investigates learning gains (a cognitive outcome) due to robot nonverbal behavior, and specifically how affective robot behavior improves learning. This is different from behavior such as pointing or gazing at task-specific information, as affective behavior is not imparting any task-related information. As an example, teachers often encourage their students after they make mistakes and celebrate their successes [4], and the robot can emulate that behavior. We will refer to this type of affective behavior that matches the desired effect (e.g., happiness on student's success) as *matching*. The effect on student learning can also be impacted by the difficulty of the concept to learn, and we design an experiment that measures the interaction between task difficulty and matching affective robot behavior.

In this work, we use a simulation of the humanoid robot Quori [5] to generate neutral and affective feedback to the students' correct and incorrect moves in a card sorting task. The goal of the task is for the students to learn the underlying sorting rules, based on demonstrations and feedback by the robot. We conducted an online study to gauge the effect of the rule difficulty and robot affective movements on learning performance. Our hypotheses are the following:

H1: Participants will learn the rule better with the *matching affective* robot.

H2: Participants will have a more positive subjective experience of the game with the *matching affective* robot (more engaged, lower perceived difficulty, higher perceived learning).

H3: Participants will have a more positive experience of the *matching affective* robot (higher intelligence and animacy).

Our results show that participants learned the rules better with the *matching affective* robot compared to a *neutral* robot. Participants also perceived the difficulty of the task as lower with the *matching affective* robot, independent of the true task difficulty. Overall, the results indicate that the robot tutor's nonverbal behaviors do improve objective performance and subjective experience of an educational task.

II. Related Work

A. Human nonverbal behavior

Nonverbal communication is an important aspect of successful teaching. It can serve a variety of purposes including

*NSF Grant #IIS-1939047

supplementing verbal exchanges, revealing emotional states, and influencing the performance of others [6]. Student nonverbal behaviors such as attentiveness and use of space can also help teachers be more successful in teaching [7].

Teacher behavior has been shown to have a strong impact on student achievement [8]. Specifically, teaching characteristics such as perceived enthusiasm affect students' emotions [9]. Teachers' and students' emotions are closely related and are also tied to other variables such as student engagement and interpersonal relationships [10]. A robot teacher design, therefore, may benefit from emotional behavior. In our work, the robot plays the role of a teacher or tutor and provides feedback to the students' responses, using a combination of text and nonverbal modalities.

B. Robot affecting cognitive outcomes

We are interested in measuring how robot affective behavior impacts students' learning. Much of the work on improved task performance has focused on gestures or gaze to direct human attention. Robot deictic gestures were shown to improve task performance in difficult tasks [11], [12]. Gaze cues were shown to reduce human response times in a collaborative task [13]. Eye contact and iconic gestures improved retention of a message communicated by the robot [14]. Our work similarly uses a combination of arm gestures and gaze (accomplished by the robot turning its body), but our nonverbal behavior does not include task-related information and is purely affective.

Cognitive outcomes in studies designed to investigate the impact of a personalized robot generally involve comparing knowledge before and after an interaction. A robot personalized to learning differences improved the post-test performance compared to a non-personalized robot [15]. A socially supportive robot increased language test scores in [16], and higher nonverbal immediacy can improve learning in interactions between children and robot tutors [17]. In our work, we can measure learning incrementally, as the players sort each card, and their performance indicates how well they have learned the sorting rule.

C. Robot impacting affective outcomes

There has been extensive work into how nonverbal movements affect human's perception of the robot and subjective experience, such as increased enjoyment of an interaction [18] and evaluation of the robot as more likeable, active, and engaged [19]. Robots have been shown to increase engagement [20], especially with physically embodied robots [21], [22]. We are interested in how affective nonverbal behavior will impact subjective experience as players learn the sorting rule.

III. Sorting Game

Our educational task is a sorting game, where the learner must infer a rule that defines which cards belong in which of two bins. The students see one card at a time, guess which bin it belongs in, and then are told whether they guessed correctly. The cards (from the game Set[1]) have 4 properties (color, number, shading, and shape), and each property has 3 possible values, resulting in a total of 81 cards.

The set of all possible rules is very large, but only some of those rules are reasonable for a human to understand. For example, a rule that randomly sorts cards into the two bins would be technically possible, but there is no pattern for a human to infer. We defined rules of two different forms, which we call *easy* and *difficult*. An *easy* rule sorts cards based on a single property; for example, the rule "all diamonds go in the left bin and all squiggles/ovals go in the right bin" uses the shape property. A *difficult* rule sorts cards based on two properties; for example, "all green and purple diamonds go in the left bin and the rest (red diamonds and green/purple cards that are not diamonds) go in the right bin" uses both the shape and color properties. We chose these two forms of rules because we believe that their structure matches closely to how humans would model the possible rules in this task. Additionally, we believe it would be difficult for a human to infer a greater rule complexity due to the limited number of cards they would be presented with to see a pattern.

A. Example Rule Inference Process

For example, consider the rule "all green cards belong in the left bin, and all others in the right bin" and let's imagine how a human could infer this rule over time. We hypothesize, and our experimental evidences seems to confirm, that humans will begin by assuming the rule is *easy* (i.e., the simplest explanation for the information in front of them) and will only consider *difficult* rules if necessary. If the first card the player sees is green-one-diamond-solid (top row, Fig. 1), they have no prior information about the rule and will simply guess a bin randomly. From the rule, we know that the card belongs in the left bin, and the game will indicate that as the correct placement.

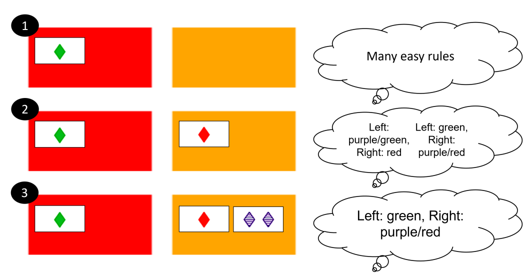

Fig. 1. Example of how a human could infer the rule that states "all green cards belong in the left bin, and all others in the right bin"

When the next card, for example, red-one-diamond-solid (middle row, Fig. 1), is shown to the player, they still do not have a clear idea as to the rule, but let us say that they guess that the rule is "all diamonds in the left bin, all others in the right bin." Using their guess of the rule, they will choose to put this card into the left bin, but the game will move it to the right bin, indicating the correct placement.

Now the player should have a clearer model of the rule. They will see from the first two cards that the only property

[1]https://www.setgame.com/welcome

that can define the rule is color. In fact there are only two possibilities as to the *easy* rule (middle row, Fig. 1).

If the next card presented to them is ‘purple-two-diamond-striped’ (bottom row, Fig. 1), the player still has an equal chance of getting this question correct, since the two possible rules sort the card into two different bins. However, after the game indicates the correct placement of the card (right bin), the player now can infer the rule, having eliminated the first possibility in the middle bubble of Fig. 1. They should now be able to sort every following card correctly, and only needed those three cards to correctly infer the rule.

There are two important takeaways from this example. First, when the rule is *easy*, it can be quick to eliminate all invalid rules. If the rule was *difficult*, it would take longer as all *easy* rules need to be eliminated first. Second is the importance of the order in which cards are presented to players. Certain cards are going to be more informative than others when it comes to eliminating possible rules, and uninformative cards can lead to it taking much longer to learn the rule. We developed an algorithm (details are beyond the scope of this paper) which presents cards in an order that will eliminate the greatest number of rules, given what has already been inferred from previous cards.

IV. Robot Feedback

We generated feedback videos that include a movement of the simulated robot with an overlay of text, indicating the robot’s speech. The robot, Quori [5], was simulated using Gazebo [23]. The robot videos are designed to react to the human sorting a card and provide information about the correctness as well as subtler feedback through movement. All feedback videos contain text that communicates correctness with ‘Hmm, not quite‘ or ‘Maybe think about the pattern in a different way‘ indicating an incorrect answer and ‘Good thinking!’ or ‘Nice work!’ for a correct answer. This text feedback was developed using guidelines for teacher praise and criticism from literature [4], [24].

We designed three types of nonverbal robots – *neutral*, *matching affective*, and *nonmatching affective* – each with their own type of movement. The *neutral* robot performs slight movements of its joints that is generated randomly without using any input from the human or game. This is our ‘baseline’ against which we will compare the other robots. The choice to have some movement for this baseline allows this robot to still appear ‘alive’ and provide a better comparison to the other moving robots.

Our previous work found correlations between specific movement patterns of Quori and the emotions it is seen to convey by human observers [25]. We found that happiness is correlated with a backward torso movement and symmetrical arm raising, and sadness is correlated with a forward torso movement and a slow lowering of the arms. The *matching affective* robot displays the “happiness” movement after correct guesses and the “sad” movement after incorrect guesses. Conversely, the *nonmatching affective* robot displays the opposite emotion – sadness for correct and happiness for incorrect.

When the participant chooses a bin to place the card, the robot turns toward the correct bin (turning only slightly in the neutral behavior), performs a nonverbal movement based on the type of robot, displays the text feedback, and returns to a neutral position. We generated a few slightly different movements for the correct and incorrect cases, for each robot, to have more variety in the movement options.

Fig. 2 shows example screenshots from feedback videos for the *neutral* and *matching affective* robots when the correct bin is the left bin. The text feedback appears, and the robot will turn towards the correct bin (left) in all cases.

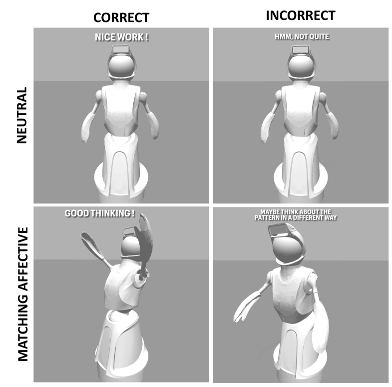

Fig. 2. Example screenshots of the *neutral* and *matching affective* conditions when the correct bin is the left bin.

V. Study Design

We designed an online study using this sorting game to test the effect of affective robot behavior on learning. We first discuss the two phases of the game: demonstration and trial, and then how we use the game to design the online experiment.

A. Demonstration and Trial Phases

In the demonstration phase, participants see two cards, one at a time. The participants are told that the robot will demonstrate where the two cards belong based on the rule. This gives them a head start in learning the rule and establishes the robot as a ‘teacher’ who will help them learn the rule.

In the trial phase (Fig. 3), the participants drag a card from the gray staging area to the bin they believe it belongs (they can move it from one bin to another if they change their mind). After clicking the “Submit Choice” button, the robot provides feedback (see Section IV) on their choice through a video, and then the next trial loads.

As the participant plays the game, the cards previously seen (in either the demonstrations or previous trials) remain visible in the correct bins, as we are not testing memory but ability to infer the rule from previously seen cards. Additionally, if a participant sorts a card incorrectly, it will be placed in the correct bin when the next card appears so mistaken assumptions about the rule do not persist.

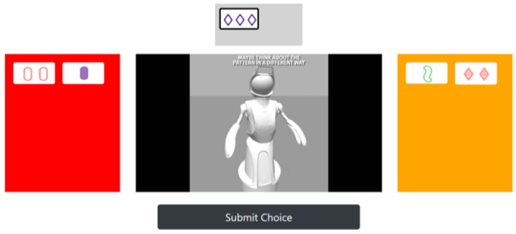

Fig. 3. The player has seen two demonstration cards and has sorted two cards themselves. They are now attempting to sort their third card.

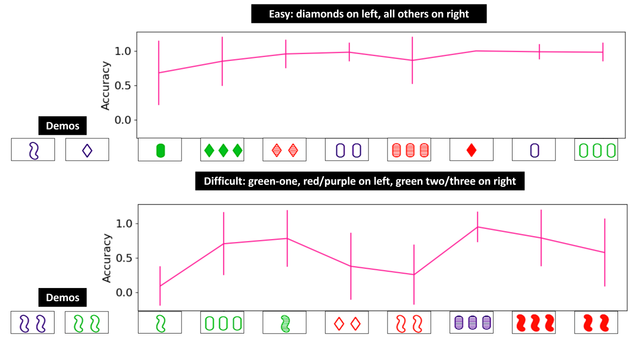

Fig. 4. Participant performance for the easy and difficult rule. Demonstration cards and trial cards are shown as well as an explanation of each rule.

B. Experimental Design

In this experiment, we want to test the impact of affective robot behavior on learning. Our first independent variable is **difficulty**, whether the rule is *easy* or *difficult*. Our second independent variable is the type of **feedback**. Our primary experiment is comparing the *neutral* and *matching affective* robots, though we did run this study comparing *neutral* and *nonmatching affective*. The reason for this is to determine whether the differences seen are truly because of the matching affect and not simply due to the increased movement from the affective movements.

During the experiment, each participant played two rounds of the game, once with an *easy* rule and once with a *difficult* rule (the order of the rules was randomized). Additionally, they were randomly assigned one of the **feedback** conditions for each round, allowing for seeing the same condition twice. The *easy* and *difficult* rules were the same for all players, and each had a fixed card order. We treat the two rounds each participant plays as independent trials.

When a participant begins the study, they complete a consent form and are presented with an explanation of the study format and sorting game. They then play their first round of the game based on the condition they were randomly assigned. We compute the accuracy from the trial phase as the average correctness of the 8 cards they have to sort.

The robot feedback video is placed in between the two bins (Fig. 3) and is chosen based on the **feedback** condition of that round. We generated 8 neutral feedback videos and 16 affective feedback videos of Quori using a Gazebo simulation, where each video is categorized by the feedback condition and correctness. After a participant chooses a bin to sort a card, the game will look at all videos for their feedback condition and that indicate the correctness of their choice (correct or incorrect). It will choose a video that meets those criteria that the participant has seen the least often so that they see a variety of videos.

After both the demonstration and trial phases, they complete a short survey consisting of 5-point Likert-style questions to evaluate their experience. Questions related to the game are listed below with scales from strongly disagree to strongly agree

- (Engagement) I enjoyed playing the game
- (Perceived Difficulty) I thought this game was difficult
- (Perceived Learning) I feel that I learned the game well.

Questions related to the robot were taken from the Godspeed Questionnaire [26] and included 3 questions related to animacy and 2 related to intelligence. We treat these qualitative measures as numeric data with the lowest level as 0 and highest level as 4. We average the 3 animacy questions to get an average animacy as well as the 2 intelligence questions to get an average intelligence score. We also included an optional area for free-form feedback.

After completing this survey, they complete their second round of the game, with demonstration/trial phases and survey, with **difficulty** and **feedback** determined by the condition they were sorted into.

VI. Results

Our primary study between *neutral* and *matching affective* was run on Prolific with 160 participants who identified as 70% female, 28% male, 2% other. Most were not at all, slightly, or moderately familiar with robots and had a bachelor's degree or less. Additionally, 66% identified as white, 23% as black, and 11% as other. We removed 4 outliers from the 320 rounds in which the participant sorted fewer than 2 out of 8 cards correctly (or 2.67 SD from the average accuracy).

Fig. 4 plots the average participants' performance for each of the 8 trials, separated by rule **difficulty**, with one SD around the mean shown. We can see that for the *easy* rule, participants seem to learn the rule at around trial card 6, as the accuracy approaches 1 at that trial. In contrast, for the *difficult* rule, participants seemed to improve performance over time, but did not reach the high accuracy achieved with the easy rule. In fact the performance seems to decline in the later trials, though the SD is quite large.

A. Difficulty and Feedback

We conducted a mixed-ANOVA, with **difficulty** treated as within subjects and **feedback** as between subjects. Not surprisingly, we found that the *easy* rule had higher accuracy, higher perceived user learning, lower perceived difficulty, and higher engagement – all significant at the 0.001-level.

More interestingly, Fig. 5 illustrates the results comparing the two **feedback** conditions, with significant differences marked. The results support hypothesis **H1**, as the accuracy

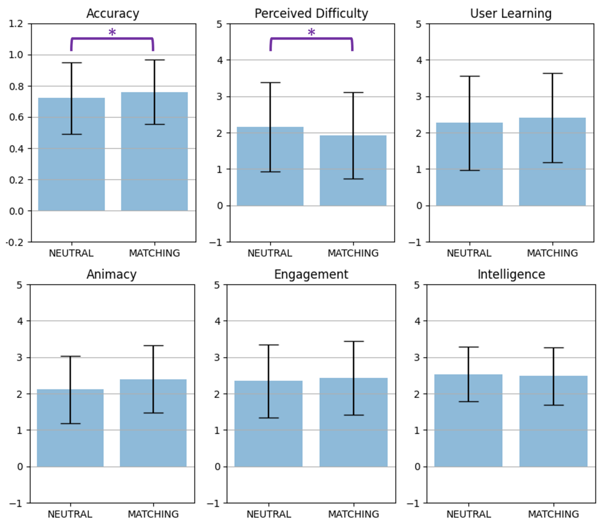

Fig. 5. Significant differences were found in accuracy and perceived difficulty at the $\alpha = 0.05$ level.

for participants viewing the *matching affective* robot was significantly higher than for the *neutral* robot ($F(1,75) = 6.48, p < 0.05, \eta^2 = 0.080$). However, when we ran this same study with 160 participants comparing *neutral* and *nonmatching affective*, there was no significant difference in accuracy (full results omitted due to space constraints).

There was also a significant interaction between **feedback** and **difficulty** ($F(1,75) = 4.36, p < 0.05, \eta^2 = 0.055$). We performed further analysis with a Tukey test, which found a significant difference in accuracy between the two **feedback** conditions when the rule was *difficult* ($p < 0.05$), but no significant difference when the rule was *easy*. This indicates that accuracy is higher in the *matching affective* condition when the rule is *difficult*.

H2 is partially supported as the perceived difficulty for the *matching affective* robot was significantly lower compared to the *neutral* robot ($F(1,75) = 4.54, p < 0.05, \eta^2 = 0.057$). This same significant difference was not present between *neutral* and *nonmatching affective*. The average user learning and engagement measures were higher for the *matching affective* robot, but these differences were not significant.

Finally, **H3** is not supported – while the average animacy measure was higher for the *matching affective* robot, there was no significant difference ($p > 0.05$).

B. Excerpts from Free-response

At the end of each round, the participants could provide optional free-form feedback about their experience playing the game and their opinion of the robot. A few excerpts of the responses for the three **feedback** conditions are shown below:

Neutral:

- the robot was still too boring and didn't help at all
- I wish the robot had a cute voice or did a dance so when you get an encouraging phrase you feel even happier.
- The robot was quiet and besides the responses you got from after making your move, the robot is kind of a non factor.

Matching Affective:

- the robot seemed to be celebrating when the answer was right.
- The robot is okay because it's responsive and active
- Robot felt more engaging this time
- The robot seemed nice and friendly (comments were not judgemental, seemed excited when you got one right, etc.)

Nonmatching Affective:

- The robot was having reactions to my actions but it looks like it's just a pre-programmed video, not that it is lively.
- I found it quite hard to understand the signals of the robot
- the robot patterns are somewhat confusing.
- The robot was kinda distracting.

We cannot generalize that these opinions were held by all participants, as we did specifically choose these comments because they further supported our hypotheses about the positive impact of the *matching affective* robot. However, the existence of these opinions does show that some participants did hold those beliefs.

VII. Discussion

Our results show that providing *matching affective* nonverbal behavior can be helpful for learning. Specifically, the accuracy during the sorting game is higher with this nonverbal behavior and the perceived difficulty of the rule is lower. This is perhaps a surprising result as the nonverbal behavior is not providing new task-related information as the human is learning the sorting rule. All **feedback** conditions told the human whether they sorted each card correctly through text. The only difference was the nonverbal behavior provided. Additionally, we did not find the same significant increase in accuracy or decrease in perceived difficulty between the *neutral* and *nonmatching affective* cases. This indicates that the improvement in learning is not due to the increase in movement between the *neutral* and affective robots, and that the matching affect is having an impact.

The significant interaction effect is also an interesting result, as this indicates that the impact of the *matching affective* movement is context-dependent: it has a significant impact when the task is difficult. A reason for this could be that the encouragement provided by the robot reduces frustration felt by the difficult task. When the task is easier, this encouragement is not needed as the human is learning the rule quicker (as illustrated by the varying performance in Fig. 4).

We can get some insight into the players' thought processes by looking at the free response feedback they provide. For the *neutral* robot, they saw the robot as boring and quiet, and one participant wanted the robot to be more encouraging with its movements. In contrast, participants thought the *matching affective* robot was celebrating correct answers and friendly.

If the participants were simply responding to the movement of the robot, they might see both the *matching* and *nonmatching* affective robots as similar – both more lively and engaging than the *neutral*. Some participants, however, saw the *nonmatching affective* robots as distracting or confusing and found it difficult to understand the robot's signals. This is intuitive since the robot's affective behavior was not matching the contextual situation of the game. These participant comments further validate the utility of *matching affective* feedback during this sorting game.

VIII. Conclusion

We designed a sorting game in which players infer a sorting rule using feedback provided by a robot. We tested the impact of affective nonverbal robot behavior that provided encouragement (without more task-specific information) on learning the rule. We found having these emotional behaviors did improve the sorting accuracy, especially with greater rule difficulty, and lowered the perceived difficulty of the task. These results clearly point to the efficacy of emotional robot behaviors for not just more positive subjective experiences, but also for objective learning gains.

This work can be extended in two different ways. First, the robot's feedback currently takes into account only a single piece of information, the correctness of the previous question. We can extend this by analyzing student facial expressions and including historical performance to personalize the chosen feedback further. Second, the diversity of robot feedback can be increased by including more degrees-of-freedom, such as facial expressions. Further, performing this study in an in-person setting would allow for real-time feedback and the use of features not available in simulation (e.g., proximity to the student). Personalizing the education experience with rich robot feedback can further improve the students' learning performance and experience.

Acknowledgment

This work was supported in part by NSF Grant #IIS-1939047, and the user study was IRB approved.

References

[1] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, and F. Tanaka, "Social robots for education: A review," *Science robotics*, vol. 3, no. 21, 2018.

[2] 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.

[3] M. J. Harris and R. Rosenthal, "No more teachers' dirty looks: Effects of teacher nonverbal behavior," *Applications of nonverbal communication*, pp. 157–192, 2005.

[4] P. C. Burnett, "Elementary students' preferences for teacher praise," *The Journal of Classroom Interaction*, pp. 16–23, 2001.

[5] A. Specian, N. Eckenstein, M. Yim, R. Mead, B. McDorman, S. Kim, and M. Matarić, "Preliminary system and hardware design for Quori, a low-cost, modular, socially interactive robot," in *Workshop on Social Robots in the Wild*, 2018.

[6] A. E. Woolfolk and D. M. Brooks, "Chapter 5: Nonverbal communication in teaching," *Review of research in education*, vol. 10, no. 1, pp. 103–149, 1983.

[7] T. P. Mottet and V. P. Richmond, "Student nonverbal communication and its influence on teachers and teaching," *Communication for teachers*, pp. 47–61, 2002.

[8] J. E. Brophy, *Teacher Behavior and Student Achievement*. No. 73, Institute for Research on Teaching, Michigan State University, 1984.

[9] T. Goetz, O. Lüdtke, U. E. Nett, M. M. Keller, and A. A. Lipnevich, "Characteristics of teaching and students' emotions in the classroom: Investigating differences across domains," *Contemporary Educational Psychology*, vol. 38, no. 4, pp. 383–394, 2013.

[10] G. Hagenauer, T. Hascher, and S. E. Volet, "Teacher emotions in the classroom: Associations with students' engagement, classroom discipline and the interpersonal teacher-student relationship," *European journal of psychology of education*, vol. 30, no. 4, pp. 385–403, 2015.

[11] 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)*, pp. 51–58, IEEE, 2016.

[12] 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*, pp. 1459–1466, 2014.

[13] 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.

[14] E. T. Van Dijk, E. Torta, and R. H. Cuijpers, "Effects of eye contact and iconic gestures on message retention in human-robot interaction," *International Journal of Social Robotics*, vol. 5, no. 4, pp. 491–501, 2013.

[15] D. Leyzberg, S. Spaulding, and B. Scassellati, "Personalizing robot tutors to individuals' learning differences," in *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 423–430, IEEE, 2014.

[16] M. Saerbeck, T. Schut, C. Bartneck, and M. D. Janse, "Expressive robots in education: Varying the degree of social supportive behavior of a robotic tutor," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1613–1622, 2010.

[17] J. Kennedy, P. Baxter, E. Senft, and T. Belpaeme, "Higher nonverbal immediacy leads to greater learning gains in child-robot tutoring interactions," in *International Conference on Social Robotics*, pp. 327–336, Springer, 2015.

[18] E. Wang, C. Lignos, A. Vatsal, and B. Scassellati, "Effects of head movement on perceptions of humanoid robot behavior," in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, pp. 180–185, 2006.

[19] 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*, pp. 247–252, IEEE, 2011.

[20] C. D. Kidd and C. Breazeal, "Effect of a robot on user perceptions," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, vol. 4, pp. 3559–3564, IEEE, 2004.

[21] H. Köse, P. Uluer, N. Akalın, R. Yorgancı, A. Özkul, and G. Ince, "The effect of embodiment in sign language tutoring with assistive humanoid robots," *International Journal of Social Robotics*, vol. 7, no. 4, pp. 537–548, 2015.

[22] J. Wainer, D. J. Feil-Seifer, D. A. Shell, and M. J. Mataric, "Embodiment and human-robot interaction: A task-based perspective," in *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 872–877, IEEE, 2007.

[23] N. Koenig and A. Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, vol. 3, pp. 2149–2154, IEEE, 2004.

[24] 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.

[25] R. Kaushik and R. Simmons, "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*, pp. 62–66, 2021.

[26] 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.