Comfort with Robots Influences Rapport with a Social, Entraining Teachable Robot

Nichola Lubold[1](✉), Erin Walker[2], Heather Pon-Barry[3], and Amy Ogan[4]

[1] Arizona State University, Tempe, AZ 85281, USA
nlubold@asu.edu
[2] University of Pittsburgh, Pittsburgh, PA 15260, USA
eawalker@pitt.edu
[3] Mount Holyoke College, South Hadley, MA 01075, USA
ponbarry@mtholyoke.edu
[4] Carnegie Mellon University, Pittsburgh, PA 15213, USA
aeo@cs.cmu.edu

Abstract. Teachable agents are pedagogical agents that employ the 'learning-by-teaching' strategy, which facilitates learning by encouraging students to construct explanations, reflect on misconceptions, and elaborate on what they know. Teachable agents present unique opportunities to maximize the benefits of a 'learning-by-teaching' experience. For example, teachable agents can provide socio-emotional support to learners, influencing learner self-efficacy and motivation, and increasing learning. Prior work has found that a teachable agent which engages learners socially through social dialogue and paraverbal adaptation on pitch can have positive effects on rapport and learning. In this work, we introduce Emma, a teachable robotic agent that can speak socially and adapt on both pitch and loudness. Based on the phenomenon of entrainment, multi-feature adaptation on tone and loudness has been found in human-human interactions to be highly correlated to learning and social engagement. In a study with 48 middle school participants, we performed a novel exploration of how multi-feature adaptation can influence learner rapport and learning as an independent social behavior and combined with social dialogue. We found significantly more rapport for Emma when the robot both adapted and spoke socially than when Emma only adapted and indications of a similar trend for learning. Additionally, it appears that an individual's initial comfort level with robots may influence how they respond to such behavior, suggesting that for individuals who are more comfortable interacting with robots, social behavior may have a more positive influence.

Keywords: Teachable agent · Entrainment · Pitch · Loudness · Rapport · Learning

S. Isotani et al. (Eds.): AIED 2019, LNAI 11625, pp. 231–243, 2019.
https://doi.org/10.1007/978-3-030-23204-7_20

1 Introduction

When teaching others, learners attend more to the problem, reflect on misconceptions when correcting their peers' errors, and elaborate on their knowledge to construct explanations, leading to enhanced learning [1]. We are interested in exploring how a pedagogical agent can be used to help learners have successful "learning-by-teaching" experiences. Some research has shown that when learners feel more rapport for their agent, they are more likely to benefit [2]. We focus on how an agent's social behavior can promote rapport and potentially influence engagement and learning.

Social behaviors that can enhance rapport include facial expressions, movement, and social dialogue. Social dialogue in particular has been found to influence engagement, motivation, and learning [3, 4]. In this work, we are interested in a relatively novel area of social behavior which is complementary to social dialogue: paraverbal behavior (i.e. loudness and tone of voice). Some early work on paraverbal behavior has shown that learners respond more positively to pedagogical agents which utilize dynamic paraverbal expressions [5, 6]. We explore paraverbal behavior based on the conversational phenomenon of entrainment. Entrainment occurs when speakers adapt their behavior, including paraverbal features such as tone and loudness, to one another, becoming more similar over time. In human-human interactions, entrainment has been found to be related to rapport, agreement, engagement, and communicative effectiveness [7–10]. In human-computer interactions, we found that a teachable robot that entrained on pitch and utilized social dialogue increased learning significantly [11].

It is an open question whether entrainment can have a positive effect on rapport and learning on its own or if it is more powerful in the presence of other social behavior. Implementing paraverbal entrainment in agents and robots is still in the early stages, and explorations of entrainment as an independent social behavior are limited. On the one hand, the Communication Accommodation Theory (CAT) suggests that individuals entrain to achieve social approval [12]; an individual on the receiving end of a high level of entrainment is likely to feel more rapport for their partner than if they were a receiver of low entrainment. This would suggest that entrainment as an independent social behavior (i.e., in the absence of social dialogue) might enhance rapport. On the other hand, fine-grained analyses of human-human entrainment suggest that people entrain differently depending on dialogue content, such as entraining more on pitch when speaking socially [13, 14]. Entrainment might play a stronger role in building rapport when it is accompanied by other social behavior like social dialogue.

In this work, we explore **how paraverbal entrainment influences rapport and learning** with a pedagogical agent by comparing three versions of the agent: a non-social version, a version which introduces paraverbal entrainment, and a version which combines paraverbal entrainment with social dialogue. To implement entrainment, we adapt paraverbal features over time. Prior work automating entrainment has generally focused on static approaches, where adaptation is relatively constant. Our previous implementation was one of the first to explore adaptation over time, called convergence [11]. In that prior work, we implemented convergence on one feature, pitch. However, entrainment on both pitch and loudness is more common in human interactions and is highly correlated with task-success and learning [15, 16]. For this work, we explore convergence on both pitch and loudness. This approach more closely mirrors observations of human conversation and may have stronger effects on rapport.

Another open question regarding the social effects of paraverbal entrainment is the role of individual differences. Prior work has indicated individual differences influence responses to social behavior [17–19], and these differences have largely been defined by gender. However, positioning different responses as broad gender differences might not appropriately represent the characteristics, experiences, and expectations that create these distinctions [20]. In human-human interaction, the dynamic judgements people make about their partners are based on their behavioral expectations of their partner, and these judgements form the basis for how rapport is built [21, 22]. In human-computer interactions, comfort level might better reflect the expectations and prior experiences that influence social responses to social behavior. Liete and colleagues found that as children got more comfortable with the iCat robot over multiple interactions, they began talking to the robot more off-task [23], and Huttenrauch and colleagues found higher engagement when individuals interacted with a robot that they were more comfortable with [24]. Individuals who are more comfortable interacting with a pedagogical agent might have higher expectations of the agent's ability to be social. Individuals with low comfort might be more cautious, more prone to anxiety and stress, with no expectations regarding an agent's social behavior. Depending on how expectations are met, low-comfort and high-comfort individuals might have different responses to social behavior. We therefore include an analysis on **how comfort level influences feelings of rapport and learning** in our exploration of a social, entraining pedagogical agent.

To examine paraverbal entrainment, social dialogue, and the effects of comfort-level, we utilize a type of pedagogical agent known as a teachable robot. Teachable robots have demonstrated potential in learning scenarios, including the ability to promote motivation, self-confidence, social engagement, and learning [25–27]. Teaching an agent can be beneficial due to the protégé effect, where learners can both feel more responsible for their agent and believe the onus of failure belongs to the agent, easing the negative repercussions of failure [28]. Influencing social responses may help enhance this protégé effect. Our teachable robot is a Nao robot named Emma. Emma engages learners using spoken dialogue, and learners teach Emma how to solve math problems. With Emma, we conducted a study with 48 middle school participants where learners taught Emma in one of three conditions: (1) an entraining condition where Emma converged on pitch and loudness, (2) a social + entraining condition where Emma spoke socially and converged (3) a non-social control. In the next section, we describe Emma, the implementation of paraverbal entrainment, and social dialogue. We then describe the study and the results of the three conditions in Sects. 3 and 4, and we discuss these results in Sect. 5.

2 Teachable Robot System

Emma is a Nao robot that 7th and 8th grade learners teach how to solve proportions, equations, and ratios; an example problem is given in Fig. 1. We describe the system in the next section, followed by the entrainment and social dialogue design.

Problem 1

Emma's friends have been arguing over who can make s'mores faster. Emma has an equation for how fast Tasha can make s'mores. Help her figure out an equation for how fast Zach can make s'mores.

Step	S'more Maker	Minutes (y)	S'mores (x)	Setup (b)	Slope (m)
0	Tasha	8	2	4	2
1	**Zach**	**9**	**2**	**1**	**???**

Fig. 1. Example of a problem, and an image of a learner interacting with Emma.

2.1 System

Learners taught Emma how to solve the math problems using spoken language and a touch-screen interface on a tablet computer (Microsoft Surface Pro). For each problem, Emma and the learner were given partial information, such as the second row in the table in Fig. 1. Emma initiated dialogue requesting the learner's guidance on how to solve for the missing information. To speak to Emma, the learner pressed and held a button on the interface while they spoke. The speech interaction was real-time. After the learner spoke, an image would appear on the screen to indicate that Emma was 'thinking' during which time a response was generated. The dialogue system consisted of an automatic speech recognizer (ASR), a dialogue manager, a paraverbal feature extractor, and a module for paraverbal manipulation and text-to-speech (TTS).

For the ASR, we utilized the HTML5 Speech API available in Chrome. For the paraverbal feature extraction, we utilized Praat [29]. For the paraverbal manipulation and TTS generation, we utilized the Nao robot's TTS system. For the dialogue manager we utilized a rule-based chatbot system with the AIML framework, making use of the PandoraBots tool for AIML [30]. The AIML framework implements a rule-based process of linking keywords to pattern/transform rules and has shown promise as a means of dialogue management [3]. We utilized this process to develop responses suited to the domain content of Emma by identifying potential keywords in the learners' utterances and designing the rules and transforms to create Emma's responses.

To facilitate the dialogue flow and reduce the effects of ASR errors, we incorporated several additional pieces of functionality. Keywords were mapped to potential explanation paths for each problem which could kick off short dialogue trees when matched, helping to provide context for identifying appropriate responses. State information such as the current problem and step was also used to provide additional context. If a learner's speech could not be matched, a response was selected from a set of 'generic' utterances which included requests for clarification (i.e. "can you please repeat that?"). Finally, we enabled "autonomous life". This is a default capability that comes with the Nao robot and introduces a slight swaying and listening behavior indicating awareness.

2.2 Multi-feature Paraverbal Entrainment as Convergence

We implemented a form of multi-feature entrainment based on the single feature approach in our previous work. Previously, we implemented a form of convergence in which pitch was adapted over a series of turns. We mirrored this approach to adapt pitch and loudness over time. Using the learner's mean pitch and mean loudness, we adjusted the robot's pitch first and then the robot's loudness such that both features converged or grew closer to the learner's features. We used the text-to-speech (TTS) system which accompanies the Nao robot to generate and modify the responses.

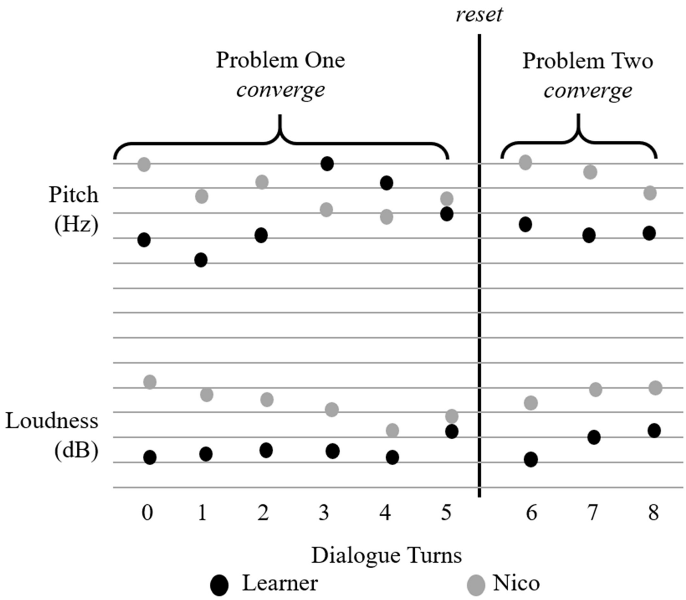

Fig. 2. Pitch and loudness both converge to the learner over time.

The manipulation of Emma's prosody was designed to incrementally converge toward the learner over the course of five dialogue turns as shown in Fig. 2. The degree to which a single utterance was adjusted was calculated as a percentage of the difference between the learner's mean and Emma's prior mean, modulated by the number of exchanges that had passed (one exchange = learner speaks, Emma speaks) and five, the number of exchanges allowed to pass before maximum convergence. Five was chosen as the number of exchanges prior to maximum convergence based on the average number of exchanges per step found to occur in four pilot evaluations. At the beginning of the conversation before the dialogue has begun, Emma's speech is generated at the specified baseline values for Emma (i.e. 230 Hz, 68 db). With the maximum number of exchanges prior to convergence set to five, Emma's speech is adapted in the direction of the user by 20% after one exchange, 40% after two exchanges by, 60% after three, and so on. At five exchanges and beyond until a reset is indicated, Emma's TTS will be generated at a mean approximated based on the estimated mean of the learner. When manipulating multiple features, it is necessary to consider how the features interact [31]. With the Nao TTS system, we found the effects of manipulating pitch on loudness were lower, with an average 0.5 dB change. We concluded that manipulating pitch in this way should have minimal effects on loudness.

Table 1. Example of social and non-social dialogue between Emma and learner

	Non-social dialogue	Social dialogue
Emma:	Before I go to the store, can you help me figure out how much of my battery I will use over time?	*Sweet*! Before I go to the store, can *we* figure out how much of my battery I will use over time?
Learner:	And we need-we need to figure out how much for one hour. So, let's multiply…	Yes, I can!
Emma:	So, I know I use 1/20th of the battery in 2/3rds of an hour. I need to figure out how much I use in one hour. So, I multiply?	*Yay*! *Thank you*! *You are awesome*. Where do *we* start?
Learner:	Yeah we multiply two thirds by three over two.	Do you see that the time is one hour? So, we must try and s-either times or divide it by the, um, battery usage
Emma:	I think I get it. But can you explain your thoughts about why I multiply?	*Oh right! You're so smart*! So, *we* know I use 1/20th of the battery in 2/3rds of an hour. To figure out how much I use in one hour, *we* multiply?

2.3 Social Dialogue

In addition to multi-feature entrainment, Emma can speak socially. Emma's social dialogue consisted of the same exact social strategies as designed and explored in our own prior work [11, 32] and includes behaviors such as name usage, inclusive language, praise, and off-task dialogue. Emma introduces one to two of these behaviors in every utterance. These behaviors are based on observations of social dialogue during peer tutoring [33] and theories of rapport, which suggest that individuals manage rapport through positivity, attention, and linguistic politeness [21, 22, 34]. Prior work has shown that these social dialogue strategies can moderately enhance learning when introduced on their own, and that combining these social dialogue behaviors with entrainment on pitch significantly enhanced learning [11]. Examples of the social dialogue from this study can be found in Table 1.

3 Study

We conducted a between-subjects experiment in which learners taught Emma in one of three conditions: (1) **non-social**: Emma exhibited dialogue to foster learning (no social dialogue, no entrainment), (2) **entraining**: Emma entrained to the learner on pitch and loudness, and (3) **social + entraining**: Emma entrained on pitch and loudness and spoke socially. Across all conditions, the instructions and the content were held constant.

Participants were 48 middle-school students from a public middle school in the United States with a mean age of 13.1 (SD = 0.75) (see Table 2). Sessions lasted 60 min and took place at the school. As in Fig. 1, participants sat a desk with the tablet in front of them. Emma stood on the desk to the right of the participant. Two participants experienced technical issues and were excluded. Thus, 15 participants were in the non-social condition, 15 in the entraining, and 16 in the social-entraining.

Table 2. Gender breakdown and dialogue statistics per session

	Females	Males	Turns	Words per turn
Non-social	8	8	116 (24)	7.1 (2.5)
Entraining	9	7	125 (26)	9.2 (3.2)
Social-entraining	9	7	119 (21)	8.9 (3.3)

Participants began with a short pre-survey and then completed a 10-min pretest. After completing the pretest, they were given a few minutes to review the worked-out solutions to the problems pertaining to Emma. They watched a short video on how to interact with Emma and then taught her for 30 min. Afterwards, they completed a 10-min posttest and a short survey on self-efficacy, rapport, and their goals. For this analysis, we were interested in the effects of rapport, learning, and comfort. We did not explore effects of self-efficacy or goals here.

To measure rapport, we asked 12 questions on attention, positivity, and coordination [34] averaged to create a single construct (Cronbach's α = 0.81). To assess learning, we utilized a pretest-posttest design with two isomorphic tests counterbalanced within condition. The tests contained conceptual and procedural questions on ratios, proportions, and word problems, and were iterated on with four pilot studies. The scores were used in statistical analyses to assess learning. We measured comfort level towards robots with two questions on a Likert scale of 1 to 5: "I feel comfortable interacting with human-looking robots" and "I feel comfortable interacting with robots." We designed these questions based on work on comfort level in other domains [35–37]. We averaged the two questions (Cronbach's α = 0.79) and then split the result into a high/low comfort categorical variable where scores less than three were low comfort (n = 23) and scores greater than three were high (n = 25).

4 Results

We were interested in two open questions regarding paraverbal entrainment: (1) how paraverbal entrainment influences rapport and learning and (2) the role of comfort level in influencing rapport responses to entrainment. In particular, we were interested in how entrainment performs as an independent social behavior. We explored these questions with the teachable robot Emma where learners taught Emma in one of three conditions: a social-entraining condition, an entraining only condition, and a non-social condition. The descriptive statistics for comfort level, rapport, and learning are given in Table 3. Despite random assignment to conditions, the pretest scores for the non-social condition were significantly higher than the social-entraining ($p = .02$) and the entrainment-only conditions ($p = .02$). Therefore, in all the analyses reported, we controlled for pre-test. We also evaluated whether comfort level interacting with robots differed across conditions prior to analyzing how this factor influenced responses; we did not observe any significant differences, $\chi^2(2, 46) = .61$, $p = .74$.

Table 3. Descriptive statistics for rapport, learning, comfort, and speech recognition errors.

	Non-social	Entraining	Social-entraining
Pretest	.48 (.18)	.28 (.19)	.29 (.19)
Posttest	.63 (.16)	.36 (.25)	.53 (.22)
Rapport	4.1 (.47)	3.9 (.54)	4.4 (.47)
Comfort level	4.1 (.24)	4.0 (.20)	4.2 (.17)
Speech errors	17.5 (6)	16.4 (9)	18.7 (9)

A power analysis conducted beforehand using the effect size for rapport ($d = .41$) from our previous work would suggest a sample size of 222 to obtain statistical power at the recommended .80 level [38]. However, it was infeasible to collect that amount of data. Therefore, we interpret significance at $p < .005$, which has been suggested as a method for handling underpowered studies [39]. In addition, we report the raw Bayes factor which has been suggested as an alternative to assessing statistical significance in data [40–42]. With the Bayes Factor, we have additional insight into whether the data favors the null hypothesis over the alternative. We calculate the Bayes Factor using the approach suggested by Rouder and colleagues [43].

4.1 Rapport

We utilized an ANCOVA to explore how rapport responses differed by condition and how comfort level influenced these responses. We treated rapport as the dependent variable, condition and comfort level as independent variables, and pre-test as a covariate. Condition was significant, $F(2, 40) = 6.6$, $p = 0.003$, $\eta^2 = 0.20$, as was comfort level, $F(1, 40) = 11.5$, $p < 0.002$, $\eta^2 = 0.20$. We found a slight interaction between comfort level and condition $F(2, 40) = 3.2$, $p = .05$, $\eta^2 = 0.07$ though not significant at $p < 0.005$. We explored differences in rapport for individuals with high-comfort versus low-comfort. Individuals with low-comfort did not differ in their rapport between the social-entraining ($M = 4.1$, $SD = .51$), entraining ($M = 3.8$, .73), and non-social ($M = 3.8$, $SD = .41$) conditions, $F(2, 19) = .86$, $p = .4$, $\eta^2 = 0.12$. However, individuals who expressed high comfort interacting with robots were significantly influenced by the robot's social behavior, $F(2, 21) = 6.65$, $p = .005$, $\eta^2 = 0.31$, with individuals in the social-entraining condition feeling significantly more rapport ($M = 4.64$, $SD = .3$) than individuals in the entraining-only condition ($M = 4.0$, $SD = .31$). The estimated Bayes factor suggested that the data were 6.1 to 1 in favor of the alternative hypothesis, supporting the significant difference between the social-entraining and entraining conditions. The difference between the social-entraining and the non-social condition was not significant.

4.2 Learning

We then explored whether learning differed across conditions with a repeated measures ANOVA. We treated pretest and posttest as the dependent variables and condition as the independent variable. Overall, learning was significant, $F(1, 43) = 47.9$, $p < .001$, $\eta^2 = 0.53$, and there was a suggestion of an effect of condition, $F(2, 43) = 3.91$, $p = .03$, $\eta^2 = 0.12$. Tukey post-hoc analyses suggest that the difference is due to the social-entraining condition compared to the entraining-only ($p = .03$). The nonsocial condition did not have significantly higher gain than the entraining-only ($p = .08$), nor did the social-entraining condition over the non-social condition, ($p = .8$). Potentially, learners who felt more rapport for Emma may have been more willing to teach her, address misconceptions, and learn. We analyzed this with a partial correlation between rapport and post-test, controlling for pretest. The correlation was not significant, $r(41) = .29$, $p = .05$ and the Bayes factor was 1.0 with the data equally likely under either hypothesis.

Finally, we explored the role of comfort level with respect to learning. Adding comfort level to the repeated measures ANOVA, we did not observe significant differences on learning for individuals with high versus low comfort, $F(2, 40) = 2.5$, $p = .12$, $\eta^2 = 0.02$. Condition and comfort level suggested a potential interaction on learning, $F(2, 40) = 2.54$, $p = .09$, $\eta^2 = 0.02$. Exploring post hoc analyses, individuals with a high comfort around robots approached significantly less learning in the entraining-only condition compared to the social-entraining ($p = .006$). However, the estimated Bayes factor was 1.0. The entraining-only and non-social was not significant ($p = .04$).

5 Discussion

We were interested in the effects of paraverbal entrainment on feelings of rapport and learning, and the role of comfort level in understanding those effects. Exploring the responses of 48 middle school learners as they interacted with the teachable robot Emma, we found a significant difference in how much rapport learners felt when Emma entrained and spoke socially compared to when Emma only entrained. This difference appears to have been driven by the individuals who felt more comfortable interacting with robots. We also observed significant learning overall and a slight trend of increased learning in the social-entraining condition. We did not observe significant differences between the social-entraining condition and the non-social control.

Unlike prior work, we explored multi-feature paraverbal entrainment as an independent social behavior in its own condition. We found the social behavior of entrainment performed poorly on its own. One possible explanation is that automatic speech recognition errors (ASR) may have contributed to the dip in social responses. We calculated the number of ASR errors (Table 3). However, ASR errors did not differ across conditions, F (2, 43) = .31, p = .74, and did not appear to influence the results.

People build rapport with multiple social behaviors, and the combination of social behaviors in agents and robots has been found to be significantly more effective on some occasions than a single behavior [6, 44]. Our results indicate social behaviors may interact with one another, where the presence of one behavior can enhance the perception of the other. We utilized the same exact social dialogue here as presented in our prior work. In that prior work, social dialogue alone had a moderate effect on learning, but, when combined with pitch entrainment, social dialogue significantly enhanced learning over the non-social control. Here, when that same social dialogue is combined with an entrainment behavior that performed poorly on its own, rapport responses were enhanced. The social-entraining condition performed well, and even better than expected if we consider the prior mediocre performance of social dialogue as an independent social behavior and the poor performance of entrainment. This suggests that social behaviors can interact with one another in potentially positive yet complex relationships, while social behaviors when used alone may not have the desired effects.

We also found that individuals with a higher comfort level interacting with robots drove the difference in rapport responses. It is possible that individuals who were more comfortable had higher expectations regarding Emma's ability to be social. In human-human analyses of entrainment, higher entrainment can occur when individuals are speaking socially [14]. For high-comfort individuals, entrainment in the absence of social dialogue may have been less appealing than no social behavior at all. We did not observe significant differences across conditions for individuals who were less comfortable. This suggests that for individuals who were less comfortable, the robot's social behavior neither positively nor negatively violated their expectations of how the robot should behave. Low-comfort individuals may have been more stressed or anxious due to being less comfortable; for the robot's social behavior to have a positive effect, these factors may need to be addressed first. Interestingly, comfort level was not related to an individual's prior experience with robots or their gender, χ^2 (1, 46) = .49, p = .48.

6 Conclusion

We explored the potential of paraverbal entrainment for enhancing rapport and learning with the teachable robot Emma. We found that individuals felt more rapport for Emma when the robot both adapted and spoke socially than when Emma only adapted and indications of a similar trend for learning. This appeared to be driven by individuals who were more comfortable around robots. These findings suggest several directions for future work. First, in designing entrainment, there are alternative approaches based on how people adapt; exploring these additional patterns and combinations with social behavior is an important area of future work. Secondly, the social plus entraining condition was more appealing to individuals highly comfortable interacting with robots. Future work should explore whether increasing how comfortable individuals are around robots is needed before social behavior can have positive effects on rapport. Overall, paraverbal entrainment is a complex phenomenon and responses to it are influenced by individual differences; understanding these differences is vital for use of social behavior to enhance rapport and learning.

Acknowledgements. This work is supported by the National Robotics Initiative and the National Science Foundation, grant #CISE-IIS-1637809.

References

1. Roscoe, R.D., Chi, M.T.H.: Understanding tutor learning: knowledge-building and knowledge-telling in peer tutors' explanations and questions. Rev. Educ. Res. **77**(4), 534–574 (2007)
2. Ogan, A., Finkelstein, S., Mayfield, E., Adamo, C.D.: 'Oh, dear Stacy !' Social Interaction, Elaboration, and Learning with Teachable Agents (2012)
3. Gulz, A., Haake, M., Silvervarg, A., Sjödén, B., Veletsianos, G.: Building a social conversational pedagogical agent. In: Conversational Agents Natural Language Interaction: Techniques and Effective Practices, pp. 128–155 (2011)
4. Kumar, R., Ai, H., Beuth, J.L., Rosé, C.P.: Socially capable conversational tutors can be effective in collaborative learning situations. In: Aleven, V., Kay, J., Mostow, J. (eds.) ITS 2010. LNCS, vol. 6094, pp. 156–164. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13388-6_20
5. Kory Westlund, J.M., et al.: Flat vs. expressive storytelling: young children's learning and retention of a social robot's narrative. Front. Hum. Neurosci. **11**, 295 (2017)
6. Lubold, N., Walker, E., Pon-Barry, H.: Effects of voice-adaptation and social dialogue on perceptions of a robotic learning companion. In: ACM/IEEE International Conference on Human-Robot Interaction, vol. 2016–April, pp. 255–262 (2016)
7. Vaughan, B.: Prosodic synchrony in co-operative task-based dialogues: a measure of agreement and disagreement. In: Interspeech, pp. 1865–1868 (2011)
8. Friedberg, H., Litman, D., Paletz, S.B.F.: Lexical entrainment and success in student engineering groups. In: Spoken Language Technology Workshop, pp. 404–409 (2012)
9. Borrie, S., Barrett, T., Willi, M., Berisha, V.: Syncing up for a good conversation: a clinically-meaningful methodology for capturing conversational entrainment in the speech domain. J. Speech, Lang. Hear. Res. (2018)
10. Lubold, N., Pon-Barry, H.: Acoustic-prosodic entrainment and rapport in collaborative learning dialogues categories and subject descriptors. In: Proceedings ACM Workshop on Multimodal Learning Analytics and Grand Challenge (2014)
11. Lubold, N., Walker, E., Pon-Barry, H., Ogan, A.: Automated pitch convergence improves learning in a social, teachable robot for middle school mathematics. In: Penstein Rosé, C., et al. (eds.) AIED 2018. LNCS (LNAI), vol. 10947, pp. 282–296. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93843-1_21

12. Giles, H., Coupland, N., Coupland, J.: Accommodation theory: communication, context, and consequence. In: Contexts of Accomodation: Developments in Applied Sociolinguistics, pp. 1–68 (1991)
13. Levitan, R., Gravano, A., Willson, L., Benus, S., Hirschberg, J., Nenkova, A.: Acoustic-prosodic entrainment and social behavior. In: Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 11–19 (2012)
14. Lubold, N., Walker, E., Pon-Barry, H.: Relating entrainment, grounding, and topic of discussion in collaborative learning dialogues. In: 12th International Conference on Computer Supported Collaborative Learning, pp. 0–1 (2015)
15. Ward, A., Litman, D.: Dialog convergence and learning. In: Proceedings 2007 Conference Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work, pp. 262–269 (2007)
16. Borrie, S.A., Lubold, N., Pon-Barry, H.: Disordered speech disrupts conversational entrainment: a study of acoustic-prosodic entrainment and communicative success in populations with communication challenges. Front. Psychol. **6**, 1187 (2015)
17. Arroyo, I., Burleson, W., Tai, M., Muldner, K., Woolf, B.P.: Gender differences in the use and benefit of advanced learning technologies for mathematics. J. Educ. Psychol. **105**(4), 957–969 (2013)
18. Siegel, M., Breazeal, C., Norton, M.I.: Persuasive robotics: the influence of robot gender on human behavior. In: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2563–2568 (2009)
19. Schermerhorn, P., Scheutz, M., Crowell, C.R.: Robot social presence and gender: do females view robots differently than males? In: ACM/IEEE International Conference on Human Robot Interaction, pp. 263–270 (2008)
20. Schwalbe, M.L., Staples, C.L., Demo, D., Gecas, V., Kleinman, S., Risman, B.: Gender differences in sources of self-esteem*. Soc. Psychol. Q. **54**(2), 158–168 (1991)
21. Spencer-Oatey, H.: Managing rapport in talk: using rapport sensitive incidents to explore the motivational concerns underlying the management of relations. J. Pragmat. **34**(5), 529–545 (2002)
22. Spencer-Oatey, H.: (Im)Politeness, face and perceptions of rapport: unpackaging their bases and interrelationships. Politeness Res. **1**(1), 95–119 (2005)
23. Leite, I., Martinho, C., Pereira, A., Paiva, A.: As time goes by: long-term evaluation of social presence in robotic companions. In: Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, pp. 669–674 (2009)
24. Huttenrauch, H., Green, A., Norman, M., Oestreicher, L., Eklundh, K.S.: Involving users in the design of a mobile office robot. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **34** (2), 113–124 (2005)
25. Jacq, A., Lemaignan, S., Garcia, F., Dillenbourg, P., Paiva, A.: Building successful long child-robot interactions in a learning context. In: ACM/IEEE International Conference on Human-Robot Interaction, vol. 2016–April, pp. 239–246 (2016)
26. Tanaka, F., Matsuzoe, S.: A self-competitive method for the development of an educational robot for children. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2927–2933 (2016)
27. Walker, E., Girotto, V., Kim, Y., Muldner, K.: The effects of physical form and embodied action in a teachable robot for geometry learning. In: IEEE 16th International Conference on Advanced Learning Technologies, pp. 381–385 (2016)
28. Chase, C.C., Chin, D.B., Oppezzo, M.A., Schwartz, D.L.: Teachable agents and the protege effect: increasing the effort towards learning. J. Sci. Educ. Technol. **18**(4), 334–352 (2009)
29. Boersma, P.: Praat, a system for doing phonetics by computer. Glot Int. **5** (2002)

30. Wallace, R.S.: The Elements of AIML Style. ALICE AI Foundation INC (2013)
31. Levitan, R., et al.: Implementing acoustic-prosodic entrainment in a conversational avatar. In: Annual Conference of the International Speech Communication Association, Interspeech, 08–12 September, pp. 1166–1170 (2016)
32. Lubold, N., Walker, E., Pon-Barry, H., Flores, Y., Ogan, A.: Using iterative design to create efficacy-building social experiences with a teachable robot. In: 13th International Conference of the Learning Sciences, pp. 737–744 (2018)
33. Bell, D.C., Arnold, H., Haddock, R.: Linguistic politeness and peer tutoring. Learn. Assist. Rev. **14**(1), 37–54 (2009)
34. Tickle-degnen, L., Rosenthal, R.: The nature of rapport and its nonverbal correlates. Pyschological Inq. **1**(4), 285–293 (1990)
35. Inoue, K., Nonaka, S., Ujiie, Y., Takubo, T., Arai, T.: Comparison of human psychology for real and virtual mobile manipulators. In: IEEE International Workshop on Robot and Human Interactive Communication, ROMAN 2005, pp. 73–78 (2005)
36. Kuthy, R.A., Mcquistan, M.R., Riniker, K.J., Heller, K.E., Qian, F.: Students' comfort level in treating vulnerable populations and future willingness to treat: results prior to extramural participation. J. Dent. Educ. **69**(12), 1307–1314 (2005)
37. Hicks, C.M., Gonzales, R., Morton, M.T., Gibbons, R.V., Wigton, R.S., Anderson, R.J.: Procedural experience and comfort level in internal medicine trainees. J. Gen. Intern. Med. **15**(10), 716–722 (2000)
38. Cohen, J.: A power primer. Pyschol. Bull. **112**(1), 155–159 (1992)
39. Boeck, D., et al.: Redefine statistical significance. Nat. Hum. Behav. **2** (2018)
40. Wagenmakers, E.-J.: A practical solution to the pervasive problems of p values. Psychon. Bull. Rev. **14**(5), 779–804 (2007)
41. Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., Iverson, G.: Bayesian t tests for accepting and rejecting the null hypothesis. Psychon. Bull. Rev. **16**(2), 225–237 (2009)
42. Jarosz, A.F., Wiley, J.: What are the odds? A practical guide to computing and reporting bayes factors. J. Probl. Solving Spec. Issue **7** (2014)
43. Rouder, J.N., Morey, R.D., Speckman, P.L., Province, J.M.: Default Bayes factors for ANOVA designs. J. Math. Psychol. **56**, 356–374 (2012)
44. Kennedy, J., Baxter, P., Belpaeme, T.: The robot who tried too hard: social behaviour of a robot tutor can negatively affect child learning. In: Proceedings ACM/IEEE International Conference Human-Robot Interaction, no. 801, pp. 67–74 (2015)