2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)

Toward Affect-Sensitive Virtual Human Tutors: The Influence of Facial Expressions on Learning and Emotion

Nicholas V. Mudrick, Michelle Taub, Roger Azevedo
Department of Psychology
North Carolina State University
Raleigh, NC, USA
{nvmudric, mtaub, razeved} @ ncsu.edu

Jonathan Rowe, James Lester
Department of Computer Science
North Carolina State University
Raleigh, NC, USA
{jprowe, lester} @ ncsu.edu

***Abstract*— Affective support can play a central role in adaptive learning environments. Although virtual human tutors hold significant promise for providing affective support, a key open question is how a tutor's facial expressions can influence learners' performance. In this paper, we report on a study to examine the influence of a human tutor agent's facial expressions on learners' performance and emotions during learning. Results from the study suggest that learners' performance is significantly better when a human tutor agent facially expresses emotions that are congruent with the content relevancy. Results also suggest that learners facially express significantly more confusion when the human tutor agent provides incongruent facial expressions. These results can inform the design of virtual humans as pedagogical agents can inform the design of virtual humans as pedagogical agents and designing intelligent learner-agent interactions.**

1. Introduction

Research on learning with intelligent tutoring systems (ITSs) has consistently emphasized the importance of emotion, which has been shown to be critical during learning [1]. One component of ITSs that has been found to significantly influence learners' emotions are virtual human tutors [2]. Virtual human tutors are embodied pedagogical agents (PAs) that emulate the realistic appearance and behavior of humans. They leverage a broad range of verbal and non-verbal behaviors to scaffold student learning during naturalistic one-on-one tutoring interactions. Facial expressions have a key role in virtual human tutors' repertoire of pedagogical behaviors. For example, a furrowed eyebrow could signal that a virtual human tutor is confused by a student's recent problem-solving step, or a half smile could provide encouraging feedback to a student who has just overcome an impasse. However, there is a gap within the literature regarding how these facial expressions can influence learners' performance and emotions during learning. As such, there exists a need to systematically examine specific components within human-PA interactions to assess how these expressions can influence learners' performance and emotions during learning.

In this paper, we examine the influence of a human tutor agent's facial expressions on performance and emotions during learning with biology content with varying levels of relevancy with the MetaTutor Learning Environment. Within the MetaTutor Learning Environment, we manipulated the relevancy of the content to the question learners needed to answer so they would look towards the human tutor agent for information regarding their own assessments and performance. The human tutor agent then provided facial expressions congruent (i.e., joy) or incongruent (i.e., confusion) with the content relevancy or a neutral facial expression (included as a comparison). For the human tutor agent, we used video recordings of a human's expressions to investigate the effects of realistic facial displays of emotion on student learning. As there exists little research to draw on, using video recordings allowed us to assess the specific influence of realistic facial expressions on learners' performance and emotions to inform the design of a virtual human tutor's behavior. Investigating these effects is a formative step in the design of an intelligently interactive, autonomous virtual human tutor with naturalistic verbal and non-verbal behaviors. As such, our study contributes to an evidence base informing the design of intelligent virtual human tutors and their behaviors to facilitate more intelligent learner-agent interactions and influence learning.

1.1. Related Work

The behavioral realism of realistic virtual humans has prompted growing interest in their use as PAs. Leveraging their realism can help us understand students' behaviors during learning, as virtual human behaviors can be finely controlled, repeated, and closely resemble actual humans [3-4]. Additionally, with their behavioral realism, more naturalistic virtual agents have been argued to facilitate increased perceptions of their believability [4]. Research examining the emotional expressions of virtual agents has identified that facial expressions and gestures achieve high recognition rates [5] and higher ratings of authenticity in their expressions than human actors [6]. Furthermore, research has demonstrated the timing of these expressions to significantly moderate individuals' perceptions of these expressions during interactions [7]. As such, research has identified that facial expressions of a virtual human alone are

978-1-5386-0563-9/17/$31.00 ©2017 IEEE

enough to influence participants' perceptions of their emotions [3]. However, there has been limited research on the impact of virtual humans' non-verbal expressions of emotion within ITSs, including their effects on student learning and emotions. We address this gap in the literature by examining the influence of a human tutor agent's facial expressions on learner's emotions and performance.

Research on the use of PAs within ITSs stems from the assumption that PAs can facilitate increased learning and performance by supporting and instructing learners through the provision of scaffolding and feedback [8]. Within the literature on PAs, their design, purpose, and feedback varies substantially, and they have been shown to influence learners' emotions and learning differently. For example, research has shown confusion induced through written trialogues between the learner and the PAs resulted in students achieving higher levels of comprehension [9]. Other research has revealed that the frequency of spoken prompts and feedback provided by PAs can lead to learner frustration and boredom, which negatively influenced learning [10]. Alternatively, research has identified that the content of feedback can influence learning differently [11]. Lastly, research has also suggested that PA feedback in the form of facial expression and gesture can lead learners to perform better on transfer tests (i.e., embodiment effect; [12]).

The majority of this research has used cartoonlike PAs and may not be representative of the influence of realistic intelligent virtual agents that can leverage real-time 3D rendering and naturalistic animations to deliver non-verbal feedback [13]. Furthermore, none of these studies specify the content, facial expressions, or gestures by the PAs. Given the variety of the different populations, roles agents played, learning contexts, and content, it is difficult to generalize their effects. As such, systematically examining the influence of specific components of the human-agent interaction (e.g., the influence of specific facial expressions of emotions) on learners' performance is needed. However, given the lack of research regarding the specific influence of facial expressions on learners' performance and emotions, more evidence is needed to inform the design of future virtual human tutors. Therefore, the study reported in this paper uses realistic, natural facial expressions provided by a human tutor as a first step toward developing virtual human tutors for more intelligent interactions during learning with ITSs.

1.2. Theoretical Frameworks

To assess the influence of content relevancy and human tutor facial expression congruency on learners' performance and emotions during learning, multiple theoretical frameworks are needed. Specifically, we use the Emotions as Social Information (EASI) model [14] and Dynamics of Affective States Model [15] to explain the influence of the human tutor agent's facial expressions on the emotions and learning outcomes examined in this study. The EASI model suggests that facial expressions of emotion serve to provide information about a situation (e.g., congruence between the relevancy of a text or diagram to a question), but this information is dependent on how accurately the learner processes the expression (i.e., allocating effort to understand why the expresser is facially expressing an emotion) [14].

The Dynamics of Affective States Model argues that when learning is interrupted by an impasse, contradiction, or unexpected event, students experience confusion [15]. If their confusion is left unresolved, students will transition into frustration and then boredom. However, if students resolve their confusion through effortful cognitive processing, they will transition back into a state of deep learning.

Based on these models, we hypothesize that learners monitor and perceive the congruent information provided by a human tutor agent's facial expressions, and learners use that information as affirmation of their own content judgments, thereby influencing their performance on the multiple-choice questions. Additionally, we hypothesize that when learners are presented with incongruent or neutral facial expressions, they experience confusion based on the contradictory information to the relevancy of the content.

2. Methods

2.1. The MetaTutor Learning Environment

We conducted a study that involved students interacting with the MetaTutor Learning Environment to answer a series of multiple-choice questions regarding biology concepts. The MetaTutor Learning Environment was designed to examine the influence of a virtual human tutor agent on students' cognitive learning strategies, metacognitive judgments, and emotional responses during learning. The virtual human tutor agent resided in the upper-right section of MetaTutor's user interface while students read text passages and diagrams about human body systems (Figure 1). In this study, the virtual human tutor agent was realized with recorded video segments of a human tutor agent. These video recordings were produced with a professional actress who facially expressed joy, confusion, and neutral based on the strategic expression of specific action units (AUs) outlined in the Facial Action Coding System (FACS; [16]). The tutor's facial expressions were also validated by a trained and certified FACS coder. The facial expressions of joy, confusion, and neutral were selected in collaboration with a biology professor who indicated these emotions arise during one-on-one tutoring sessions with college students, as well as the Dynamics of Affective States Model that argues these emotions play a critical role in student learning [15]. In this study, the tutor's behavior consisted of facial expressions of specific emotions, which were dependent on the relevancy of biology content that was presented on-screen to students. For example, the tutor displayed a joyful facial expression when the on-screen content was fully relevant to the question that the student was attempting to answer, or a confused facial expression when the on-screen content was weakly relevant.

Students' interactions with the MetaTutor Learning Environment consisted of 9 counter-balanced, linearly structured, self-paced trials that interleaved science questions, multimedia science content, and metacognitive judgment prompts. The 9 trials had an identical format. In each trial, learners were first presented with a science question, and they were asked to submit an ease of learning judgment, "*How easy do you think it will be to learn the information needed to answer this question?*" on a 0 to 100% scale. Learners were next presented with a text passage, diagram, and the science question, as well as an idle video of the human tutor agent expressing a neutral facial expression. After 30 seconds learners were prompted to assess the relevancy of the text and the diagram: "*Do you feel the text/diagram on this page is relevant to the question being asked?*" Learners made two content evaluation judgments (CEs; [17]) on a Likert scale (ranging from 1-3) to the following statements: "*The text/diagram is relevant*", "*The text/diagram is somewhat relevant*", and "*The text/diagram is not relevant.*" Upon making their text and diagram CEs, the on-screen tutor expressed either a congruent, incongruent, or neutral facial expression based on the relevancy of the content. The tutor's facial expression lasted 10 seconds from start to finish. Following the facial expression, the human tutor agent video returned to a looping neutral expression.

Learners were permitted to read the text and inspect the diagram at their own pace. After they were finished examining the content, learners were prompted to answer the science question by choosing the correct response from 4 options. After submitting their answer, learners were prompted to make a retrospective confidence judgment (RCJ) by answering, "*How confident are you that the answer you provided is correct?*". Learners made their judgment on a 50 to 100% scale (where a score of 50% would indicate learners believed they had a 50/50 chance of submitting the correct answer). After submitting their response, learners were prompted to justify their answer by typing their response into a text box to ensure that learners had not simply skimmed the material and guessed at their answer. Learners were then asked to make another RCJ based on their justification.

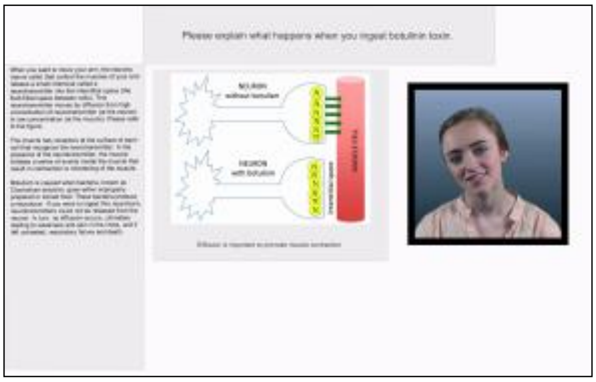

Figure. 1. Example screenshot of the question, text, diagram, and human tutor agent with a congruent facial expression of joy.

2.2. Participants

Forty-four (n = 44) undergraduate students enrolled at a large mid-Atlantic North American university participated in this study. Participants' ages ranged from 18-24 (M = 20.10, SD = 1.61). Scores from the 18-item science pre-test indicated that participants had some prior knowledge of the science domains presented in the experiment (M = 11.56 [64.22%], SD = 2.04). Participants were monetarily compensated $30 for their participation.

2.3. Design

The study used a 3x3 within-subjects design resulting in 9 trials. Our study investigated the impact of two factors. The first factor was content relevancy, which refers to the relationship between a question and the text and/or diagram that accompanied the question. The second factor was human tutor agent's facial expression congruence, which refers to the relationship between the human tutor agent's facial expression and the relevancy of the content. Specifically, learners interacted with 3 levels of *content relevancy* to the question they needed to answer: *high relevance* (i.e., the text and diagram were completely relevant), *low text relevance* (i.e., the text contained information that was not required for answering the question, but the diagram was still fully relevant to the question asked), and *low diagram relevance* (i.e., the diagram contained information that was not required for answering the question, but the text was still fully relevant to the question asked). For the three levels of *congruency*, the human tutor agent facially expressed different emotions depending on the relevancy of the content: *congruent* (i.e., the emotion matches the relevancy of the content), *incongruent* (i.e., the emotion does not match the relevancy of the content), and *neutral* (i.e., the lack of an emotional expression). Based on these manipulations, each learner completed 9 counter-balanced, randomized trials, with different combinations of *content relevancy* and human tutor agent facial expression *congruency* (see Table 1).

Figure. 2. Human tutor agent facial expressions of emotion: joy (left), confusion (middle), and neutral (right).

TABLE I. HUMAN TUTOR AGENT FACIAL EXPRESSIONS BY STUDY CONDITION.

	Content Relevancy		
	Fully Relevant	***Text Less Relevant***	***Diagram Less Relevant***
Congruent	Joy	Confusion	Confusion

Incongruent	Confusion	Joy	Joy
Neutral	Neutral	Neutral	Neutral

2.4. Measures and Multimodal Setup

The materials and equipment used in this study consisted of the following: an SMI RED 250 eye tracker, a Logitech HD Pro webcam 920, and a Shimmer 3+ wireless bracelet to capture participants' electrodermal activity [EDA]. The SMI RED 250 eye tracker collected learners' eye movements at 120 Hz, while the webcam recorded learners' facial expressions of emotions at a rate of 30 Hz at 1080p. Participants' EDA was logged with a Shimmer 3 that collected at a rate of 128 Hz.[1] Lastly, learners' performance was assessed with 4-item multiple-choice questions administered near the end of each trial.

2.5. Procedure

After participants entered the lab and completed an informed consent form, the eye tracker was calibrated by the researcher. Participants were asked to face a blank monitor screen and hold a neutral facial expression for 6 seconds to establish a baseline. Afterward, participants were asked to sit still for 5 minutes to establish a baseline for their EDA. After establishing these baselines, participants were asked to complete a computerized demographic questionnaire and an 18-item pretest that assessed their basic biology content knowledge. After participants completed the pretest, participants completed the 9 previously described trials in MetaTutor. The average interaction with the MetaTutor Learning Environment lasted approximately 60 minutes (M = 56.53 m, SD = 15.23).

2.6. Data Sources

1) *Multiple-choice response accuracy.*

Learner responses to multiple-choice questions were coded for accuracy. The coding scheme was created in collaboration with a biology professor. Fully correct responses were given a score of 1, partially correct responses were coded as .5 for containing relevant but not correct information, and incorrect responses were coded as 0. Each multiple-choice question had one correct answer, two partially correct answers, and one incorrect answer.

2) *Facial expressions of emotion.*

The videos of each learner's facial expressions captured during the experimental session were analyzed using Attention Tool FACET developed by iMotions [18]. FACET analyzed each video frame and classified 19 action units (AUs) and nine emotions (i.e., joy, anger, disgust, fear, surprise, sadness, contempt, confusion, frustration, and neutral) and calculated evidence scores for each AU and emotion at 30 Hz. The evidence scores were formatted as probabilities, representing the log odds of the presence of a facial expression as coded by an expert human coder. Though each AU and emotion were analyzed, we focused primarily on learners' facial expressions of joy, confusion, and neutral.

An absolute threshold of zero was used during analyses of the evidence score data; in other words, any negative evidence scores (i.e., the evidence of a facial expression not being present) were identified and set to zero. This threshold was used because we were primarily interested in the presence of specific emotions during the experiment, as opposed to the absence of specific emotions. The positive evidence score values were then smoothed using a symmetrical moving average filter configured with a fixed window of 11 data-points. This filter is useful for analyzing temporal data (e.g., facial expressions over time) as it smooths out random outliers while retaining the step response of the signal. As such, each smoothed data point was calculated by averaging the raw data point with the previous five and next five raw data points in the original time-ordered data stream.

To investigate the influence of the human tutor agent's facial expressions on learners' emotions, we distinguished three time segments of learners' FACET data: 1) the period before the human tutor agent's facial expression (M = 84.08 s, SD = 43.31), 2) the period during the human tutor agent's facial expression (10 s), and 3) the period following the tutor's facial expression until the learner responded to the multiple-choice question (M = 6.61 s, SD = 11.45). Three mean scores were created for learners' facial expressions of confusion, frustration, and joy for each level of relevancy and congruency. This produced 27 mean values per emotion (i.e., 3 relevancy conditions x 3 congruency conditions x 3 time segments).

3. Results

3.1. Do human tutor agent facial expression congruency and content relevancy interact to affect learners' multiple choice question accuracy?

A two-way RM-ANOVA was conducted to determine the influence of congruency type and relevancy type on multiple choice response accuracy. There was a significant interaction between congruency and relevancy on response accuracy, $F(4, 172) = 17.68$, $p < .001$, $\eta_p^2 = 1.00$. Therefore, simple main effects were conducted. Response accuracy was not significantly different for congruent (M = .72, SD = .38), incongruent (M = .56, SD = .16), or neutral (M = .69, SD = .38) facial expressions for content with high relevance. Congruency type significantly influenced response accuracy for low text relevant content ($F(2, 86) = 25.41$, $p < .001$, η_p^2 = .37). Specifically, question responses were more accurate for congruent (M = .97, SD = .17) than incongruent (M = .71, SD = .10) and neutral facial expressions (M = .48, SD = .05).

[1] Although eye-tracking and EDA data were collected, they were not analyzed for this opaper.

Additionally, incongruent facial expressions coincided with higher question response accuracy than neutral facial expressions. Response accuracy was significantly different for low diagram relevance content. Mauchly's test indicated that the assumption of sphericity had been violated $\chi^2(2) = 11.17$, $p = .004$, as such Greenhouse-Geisser values are reported. Results revealed that congruency significantly influenced response accuracy, $F(1.62, 69.72) = 10.82$, $p < .001$, $\eta_p^2 = .20$, such that responses were more accurate for incongruent facial expressions ($M = .86$, $SD = .25$) and neutral facial expressions ($M = .78$, $SD = .05$) than congruent facial expressions ($M = .55$, $SD = .48$). There were no significant differences in response accuracy for incongruent and neutral expressions for low diagram relevance content.

3.2. Do congruency type, relevancy type, and timing interact to influence learners' facial expressions of confusion, frustration, and joy?

A three-way RM-ANOVA was conducted to determine if congruency type, relevancy type, and time interacted to influence student's facial expressions of confusion. Results revealed a significant three-way interaction between all the independent variables. Mauchly's test revealed that the assumption of sphericity had been violated ($\chi2(35) = 85.06$, $p < .001$), as such Greenhouse-Geisser values are reported, $F(3.35, 229.92) = 2.23$, $p < .05$, $\eta_p^2 = 74$. We did not find evidence of a simple two-way interaction between congruency and content relevancy prior to the human tutor agent's facial expression ($F(3.06, 131.48) = .55$, $p > .05$), nor did we find a two-way interaction after the human tutor agent's facial expression $F(2, 86) = 2.82$, $p > .05$. There was a significant two-way interaction between congruency and content relevancy during the human tutor agent's facial expression $F(3.38, 145.29) = 3{,}68$, $p < .05$, $\eta_p^2 = .83$. There was a simple main effect of congruency for low text relevance content $F(2, 86) = 4.60$, $p < .05$, $\eta_p^2 = .76$, but not for highly relevant or low diagram relevance content (p's $> .05$). Pairwise comparisons revealed that learner confusion was highest during neutral expressions by the tutor ($M = .68$, $SD = .08$) and significantly different from learner confusion during congruent expressions by the tutor ($M = .49$, $SD = .08$). We did not find evidence of a significant difference in learner confusion between neutral tutor facial expressions and incongruent tutor facial expressions ($M = .61$, $SD = .09$). There was no significant difference in learner confusion between incongruent and congruent tutor facial expressions for low diagram relevant content.

Two additional three-way RM-ANOVAs were conducted to determine the influence of congruency type, relevancy type, and time on learners' facial expressions of frustration and joy. Results revealed Mauchly's test was significant and the assumption of sphericity was violated for both facial expressions of frustration, (χ^2 (35) = 130.29, $p = .000$), and joy (χ^2 (35) = 111.26, $p < .001$). Greenhouse-Geisser statistics revealed there was no significant interaction between the three independent variables on frustration ($F(4.70, 202.13) = 1.86$, $p = .108$, $\eta_p^2 = .61$) or joy ($F(4.61, 197.94) = 0.64$, $p > .05$, $\eta_p^2 = .22$).

4. Discussion and Future Directions

The study's objective was to systematically examine the influence of tutor facial expression congruency, content relevancy, and timing on learner performance and emotion. The set of examined variables in this study can significantly augment current computational models of affect dynamics and intelligent human-agent interactions in complex learning situations with ITSs [6, 8, 9]. Additionally, our findings suggest that contextual congruence is important for models of emotion in virtual human tutors.

Results from our first research question indicated that learners' performance was highest for low text relevant content when the human tutor agent expressed confusion, congruent with the relevancy of the content. These results were consistent with our hypothesis, which suggested that learners would use the information provided by the human tutor agent's facial expression as affirmation of their own evaluation of the content, thereby influencing their performance [20]. Results from our second research question indicated that learner confusion was highest when the learner interacted with low text relevant content while the tutor expressed a neutral facial expression. We did not find evidence of a significant relationship between content relevancy and tutor facial expression congruency on learner joy or frustration. This may have been due to the nature of the task: although there is evidence suggesting that joy is occasionally present during learning with ITSs, research also suggests that joy varies considerably between learning tasks [1]. The Dynamics of Affective States Model suggests that frustration occurs after periods of unresolved confusion. The trial-by-trial nature of the learning task with MetaTutor, may not have afforded the opportunity to experience lengthy periods of confusion [15]. The results on student confusion were consistent with our hypotheses, which suggested that the contradiction between the information-neutral facial expression and content relevancy would increase learners' feelings of confusion. Lastly, as results indicated that learners expressed confusion *during* the human tutor agent's facial expressions, these findings support prior research suggesting that the timing of affective expressions significantly influences human-virtual agent interactions [7].

This study examined a human tutor agent's facial expressions of emotion congruent with multimedia biology content. However, the results have important implications for contextual congruency of virtual tutor emotion expressions in other contexts, such as with the learners' affective states or the tutor's own internal state. For example, the tutor's facial expression could mimic a learner's expressions of emotion (i.e., facially expresses confusion as the learner does) to facilitate an increase in the learner's awareness of their own emotions. Alternatively, a virtual human tutor could express emotions representative of their own appraisal of the learner's actions (e.g., confusion when a learner repeatedly makes poor notes about irrelevant content). Additionally, as our results indicated the importance of timing of the facial expressions, designing future affect-sensitive virtual human tutors will need to be sensitive to the timing of their non-verbal feedback by accounting for how these expressions influence learners' own emotions. For example, designers will need to identify appropriate instances during which facial expressions may have deleterious influences on learning, such as causing a learner excessive confusion. Alternatively, identifying instances where leveraging the emotional influence of a specific facial expression will be needed, but may be challenging given current research in affective computing. For example, if the learner facially expresses confusion, the virtual human tutor can intervene with a facial expression of concern to alleviate or reaffirm the learner. The timely modeling and computational demands required for such precise intervention is likely to challenge yet advance current systems. Lastly, our results indicating information-neutral facial expressions also increased learners' confusion emphasize the importance of providing learners' informational feedback in the form of facial expressions. Promising future directions include investigating affect-sensitive virtual tutor agents that facially express emotions that are dependent on different contexts and time and are self-modifying based on a human's affective reactions to the agents' expressions thus supporting intelligent interactions and leading to optimal learning and performance.

Investigating the impact of emotive facial expressions by virtual human tutors raises several challenges: it requires creating believable expressions of emotion that are relevant to learning, it calls for granular sensor-rich measurement of learner affect (i.e., facial expressions), and it demands consideration of the temporal dynamics of emotion among virtual humans as well as learners. In this study, we have leveraged videos of a trained human actor to systematically examine the influence of a human tutor agent's facial expressions on learner performance and emotions in the MetaTutor Learning Environment. Our results imply that contextual and temporal congruence are important features for the development of future virtual human tutors.

Acknowledgments

This research is supported by the National Science Foundation (NSF DRL 1431532). The authors thank the members of the SMART Lab and the IntelliMedia Group.

References

[1] S. K. D'Mello, "A selective meta-analysis on the relative incidence of discrete affective states during learning with technology.", *Journal of Educational Psychology*, vol. 105, no. 4, pp. 1082-1099, Sep. 2013.

[2] E. Romero-Hall. "Animated pedagogical agents and emotion," in *Emotions, Technology, Design, and Learning*, S.Y. Tettegah, and M. Gartmeier, Eds. Cambridge, MA: Academic, 2015, pp. 225-237.

[3] B. Liebold, and P. Ohler, "Multimodal emotion expressions of virtual agents, mimic and vocal emotion expressions and their effects on emotion recognition," In *Affective Computing and Intelligent Interaction (ACII)*, 2013, pp. 405-410.

[4] F. D. Schönbrodt and J. B. Asendorpf, "The challenge of constructing psychologically believable agents," *J. Media Psychology*, vol. 23, pp. 100-107, Apr. 2011.

[5] J. M. Beer, A. D. Fisk, and W. A. Rogers, "Emotion recognition of virtual agents facial expressions: the effects of age and emotion intensity," in *Proc. of the Human Factors and Ergonomics Society Annu. Meeting*, San Antonio, TX, 2009, pp. 131-135.

[6] C. M. de Melo, P. Carnevale, and J. Gratch, "The effect of virtual agents' emotion displays and appraisals on people's decision making in negotiation," in *Intelligent Virtual Agents, 12th Int. Conf., Santa Cruz, CA*, Y. Nakano, M. Neff, A. Paiva, and M. Walker, Eds., New York, NY: Springer, 2012, pp. 53-66.

[7] E. G. Krumhuber, L. Tamarit, E. B. Roesch, and K. R. Scherer, "FACSGen 2.0 animation software: generating three-dimensional FACS-valid facial expressions for emotion research," *Emotion*, vol. 12, pp. 351-363, Apr. 2012.

[8] R. Azevedo, S. A. Martin, M. Taub, N Mudrick, G. Millar, and J. Grafsgaard, "Are pedagogical agents' external regulation effective in fostering learning with intelligent tutoring systems?," in *Proc. 13th Int. Nat. Conf. on Intelligent Tutoring Systems.*, Zagreb, Croatia, 2016, pp. 197-207.

[9] S.K. D'Mello, and A.C. Graesser, "Feeling, thinking, and computing with affect-aware learning," in *The Oxford handbook of affective computing*, Oxford: Oxford Press, 2014, pp. 419-434.

[10] N. Mudrick, R. Azevedo, M. Taub, and F. Bouchet, "Does the Frequency of Pedagogical Agent Intervention Relate to Learners' Self-Reported Boredom while using Multiagent Intelligent Tutoring Systems?" In *Proc. 37th Ann. Meeting Cog. Sci. Soc.*, Pasadena, CA: Cog. Sci. Soc. pp.136-141, July 2015.

[11] Y. Kim and A.L. Baylor, "Research-Based Design of Pedagogical Agent Roles: a Review, Progress, and Recommendations", *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 160-169, 2016.

[12] R. Mayer and C. DaPra, "An embodiment effect in computer-based learning with animated pedagogical agents", *Journal of Experimental Psychology: Applied*, vol. 18, no. 3, pp. 239-252, 2012.

[13] Swartout, W., Hill, R., Gratch, J., Johnson, W. L., Kyriakakis, C., LaBore, C., ... & Moore, B. (2006). *Toward the holodeck: Integrating graphics, sound, character and story*.

[14] G. Van Kleef, "How Emotions Regulate Social Life", *Current Directions in Psychological Science*, vol. 18, no. 3, pp. 184-188, 2009.

[15] S.K. D'Mello, and A.C. Graesser, "Dynamics of affective states during complex learning", *Learning and Instruction*, vol. 22, no. 2, pp. 145-157, 2012.

[16] P. Ekman and W. Friesen, "Measuring facial movement", *Environmental Psychology and Nonverbal Behavior*, vol. 1, no. 1, pp. 56-75, 1976.

[17] J. Greene and R. Azevedo, "A macro-level analysis of SRL processes and their relations to the acquisition of a sophisticated mental model of a complex system", *Contemporary Educational Psychology*, vol. 34, no. 1, pp. 18-29, 2009.

[18] *Attention Tool*, 6.2, iMotions, Boston, MA, 2016.