"Is It Just Me?" : Evaluating Attribution of Negative Feedback as a Function of Virtual Instructor's Gender and Proxemics

Dan Feng
Northeastern University
Boston, MA, USA
danfeng@ccs.neu.edu

David C. Jeong
University of Southern California
Los Angeles, CA, USA
davidjeo@usc.edu

Nicole C. Krämer
University Duisburg-Essen
Duisburg, Germany
nicole.kraemer@uni-due.de

Lynn C. Miller
University of Southern California
Los Angeles, CA, USA
lmiller@usc.edu

Stacy Marsella
Northeastern University
Boston, MA, USA
marsella@ccs.neu.edu

ABSTRACT

Virtual agents are used in a number of different outcome-based contexts such as physical and mental health, skill-based training, as well as classroom learning and pedagogy. Virtual agents in such applications are largely designed so that they project positive attitude and feedback towards the human participant. Human-human interactions, however, are certainly not exclusively positive in valence. For example, teachers and educators engage in both positive and negative feedback strategies for pedagogical outcomes. While the distinct effects of positive and negative feedback on learning are well established, few studies have attempted to examine the effects of negative feedback across different combinations of instructor's gender and proxemics-based physical behavior. This study explores this very question with a 2 (instructor gender)*2 (proxemic behavior) between subject design. In this experiment, participants (N=63) actively engage in a learning task with a male/female virtual instructor that provides negative feedback while either standing stationary or while physically approaching the participant. Based on the different deliveries of the negative feedback, the study aimed to identify the sources of variations in participant reactions to the negative feedback, namely patterns of attribution and both behavioral and physiological measurements of emotions. The results indicate that participants attribute greater self-blame (internal attribution) for their purported poor performance when interacting with the female virtual instructor than when interacting with the male virtual instructor. Participants also generally exhibited greater positive affect in response to female virtual professors than male virtual professors. These results are highly relevant both to the design of virtual agents as well as to adding to our understanding of the role of gender and behavior in human-human, non-peer interaction.

Appears in: *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), S. Das, E. Durfee, K. Larson, M. Winikoff (eds.), May 8–12, 2017, São Paulo, Brazil.*

Keywords

Verbal and non-verbal behaviour, Human(s)-virtual agent(s) interaction, Virtual agents in education, Attribution Theory

1. INTRODUCTION

Applications of virtual agents (VA) today have become ubiquitous, and the ability of pedagogical agents to facilitate learning has been well-documented [19, 37, 22, 23]. While some research focuses on the agent's appearance [2, 23], others focus on the nature of the agent's communication. With the rapid growth of immersive virtual environment (IVE) technologies, more researchers have drawn their attention to studying psychological mechanisms of social interactions, namely agents' social and affective affordances to support students [44]. The combination of IVE and VA now make it feasible to simulate such social interactions in an immersive, yet controlled environment.

That being said, current research on virtual agents focuses mainly on the interactions with a positive valence, such as politeness [46] and rapport [45, 28]. Real world social interactions, however, certainly involve a wider spectrum of contexts that are by no means exclusively positive. For instance, an educational situation can be wrought by an instructor who is negative, even impatient and hostile. The current study seeks to explore the consequences of such a negatively valenced educational context by examine student's reactions and response to a virtual teacher's negative feedback.

A fundamental determinant to how people respond to feedback from virtual agents is the manner in which the feedback is delivered. Research on politeness within [46] and outside [7] the virtual agent research community demonstrate that forms of communication, namely phrasing, are critical. Prior research has shown that both positive phrases [46] and non-verbal behaviors [28] may positively impact the learning outcome.

In contrast to prior work, the current study is not about the role of phrasing, but rather the separate and combined effects of the agent's a) gender and b) nonverbal proxemics (body language) on participants' attributional and emotional

reactions. We used a 2 * 2 design (Gender * Approach) in a virtual environment, where a male or female virtual agent (depending on condition) provides an identical series of negative feedback messages while either standing stationary or approaching the participant, so as to invade the participant's personal space. In manipulating who delivered the negative feedback (i.e., a man or a woman) how (i.e., with what accompanying proxemics), our research goal for this study was to identify resultant differential patterns across conditions in participants' patterns of attributions and affective responses. The paper is organized as follows. In Section 2, we present the theoretical background and related work. Our methods and results will be discussed in section 3, 4 and 5. We conclude and discuss the implications and future directions of this work in Section 6 and 7.

2. THEORETICAL BACKGROUND AND RELATED WORK

2.1 Negative Feedback

The disparate effects of positive and negative feedback in real-world educational contexts has long been a topic of discussion in the field of education and psychology [13, 14] That said, the effects of negative feedback have proven to be inconsistent. While some have indicated that negative feedback can benefit learning [25], others have reported that it leads to a gradual deterioration of performance, known as "learned helplessness" [14].

According to politeness theory, in alignment with our social needs, one's perception of autonomy and control seem to be significant factors in students' engagement in learning [35]. Of course, negative feedback reduces this very sense of autonomy and control.

In order to mitigate negative feedback, students employ strategies such as increasing motivation and effort [9] as well as downward regulation of goals [25]. This feedback to goal regulation link has been shown to be consistent in the face of either accurate or manipulated feedback [20].

Commitment to goals has also been shown to mediate the effects of positive and negative feedback [43]. Specifically, positive feedback is a more effective motivator for tasks that people want to do whereas negative feedback is a more effective motivator for tasks that people have to do, suggesting that people are more prone to learn tasks that they are uncommitted to through negative feedback [18].

Further, negative feedback which decreased one's own perception of ability reduces both intrinsic and extrinsic motivation, leaving people generally unmotivated [42].

2.2 Attribution Theory

A crucial response to negative feedback in an educational context is one's attribution of blame or responsibility. That is, does the student attribute blame to their own poor abilities or do they attribute blame to the instructor's poor teaching abilities? While students' strengths and positive performance are attributed by the learner to the self, weaknesses and negative performance are typically attributed to others [24]. In fact, students even attempt to ignore negative feedback that contrasts with their own conceptions of their performance [8].

Attribution theory has long been discussed in the study of the educational process [49]. Cognitive evaluation theory suggests that external factors such as rewards, surveillance [29], and evaluations [38] tend to diminish sense of autonomy, which induces a shift of perceived locus of causality from internal to external [12], thus reducing intrinsic motivation to achieve the goal.

2.3 Attribution and Goals

Attribution in education involves two categories of learning goals. A mastery goal involves a belief that effort is linked with outcome, and this belief directs achievement behavior [48]. In contrast, performance goal orientation is linked to avoidance of challenging tasks [15], negative affect in response to failure, and a subsequent judgment that one lacks ability [21]. When a person adopts a performance goal, a perception of one's ability to complete the task determines one's self-worth [11]. When trying hard does not lead to success, the expenditure of effort can become a threat to one's self-concept of ability [10].

2.4 Spatial Interaction

Another significant factor contributing to the effect of negative feedback on students is the impact of nonverbal communication, specifically body movements and spatial distance, otherwise known as proxemics [17].

In human communication, invasion of space has been linked with discomfort and a rise in Galvanic Skin Response (GSR) [39]. When such an invasion of personal space occurs, people react by moving away to re-calibrate to an acceptable interpersonal distance between communicators. Further, the approaching behavior of those individuals of higher status, such as teachers, has been shown to elicit more discomfort [1].

2.5 Embodied Pedagogical Agent

Virtual environments have been shown to impact user behavior by conveying important social information [33]. Further, these environments are capable of impacting physiological responses through interactions with others in a manner that is congruent with everyday interaction [4]. Taken together, this use of technology appears to be a valid, and fruitful way to study negative feedback in a controlled laboratory setting.

Previous studies have focused on the effects of positive feedback from virtual agents in a virtual learning environment [41, 27, 28, 16, 45]. For instace, Wang et al. [46] found that an agent who uses polite requests had a more positive impact on learning that a more direct agent. Further, Krämer et al. [28] found a significant improvement on participant's performance when interacting with same-gender virtual agents that rapidly respond to the participants with positive non-verbal behavior. Departing from these previous works, the current study focuses on students' direct response to purely negative feedbackfrom virtual instructors.

Prior research on the effects of a virtual agent's gender on learning are generally inconsistent. For instance, students may feel higher positive affect and self-efficacy after interacting with a female virtual human [36], but may perceive female instructors to be less competent and intelligent than male instructors [3]. The goal of the researchers in the current study is to further explore the specific nature of the effects of virtual human gender on students' reactions to negative feedback.

In addition to virtual human gender, the current study

Figure 1: Screen-shot of the virtual environment. The character stands on the left is the virtual professor and the one on the right is another fellow student. The participants' lines will show up on the 3D interface in front of the participants.

examined nonverbal proxemics (body language) as a factor that might impact participants' reactions. For instance, people tend to attempt to calibrate an equilibrium of interpersonal distance when interacting with a virtual agent [26]. For a review of spatial construction of interactive virtual environments, see Blascovich [5].

The current study aims to add to our understanding of social perception in virtual environments by examining the individual and combined effects of virtual human gender and proxemics.

2.6 The Hypotheses of Current Study

2.6.1 Gender-based Hypotheses

We expect the gender bias introduced in the previous section to carry over to current study. That is to say, we expect the female virtual instructor to result in a greater tendency among participants to externally attribute their failure. Of course, "failure" is a controlled element in the current study as every participant will technically "fail" and receive negative feedback.

On the other hand, we expect critiques from the male (versus female) virtual instructor will lead to greater Skin Conductance Level (SCL) arousal and negative emotional reactions, especially under the approach condition.

2.6.2 Proxemics-based Hypothesis

We anticipate several proxemics-based results to occur. First, when personal space is invaded, we hypothesize that people will tend to move away in order to maintain basic social distance. Second, compared to the baseline (non-approaching condition), participants' arousal (measured by SCL) and negative affect will increase when the interpersonal distance between virtual instructor and the participant (approaching condition) is decreased. Finally, due to the invasion of personal space, participants should attribute their failure more to external factors.

3. VIRTUAL ENVIRONMENT DESIGN

A mismatch between resources and demands can create threat states that can potentially account for performance outcomes [6]. The inability to satisfy the person of power in this interaction is suggested to lead to feelings of helplessness and lack of motivation, key components of negative feedback. Using this scenario as a template, we created a virtual environment in which individuals are assigned to a task that they are incapable of completing to the satisfaction of an instructor. This system enables the researchers to develop a keener understanding of how participants respond, both verbally and physiologically, to the experience of negative feedback. This information can then be used to create intervention strategies to "buffer" participants against the instructor's negative feedback: These student tactics could be taught through successive iterations of the virtual scene. Thus, the first step before constructing such an intervention is to create a specific learning context where the instructor provides negative feedback and within which participants' responses can be assessed. Specifically, the interactive virtual environment here simulated an acting class scenario. One of the virtual characters was designed to be the instructor in the scene. Each participant and other non-player characters (NPC) were students who were asked to rehearse 'Romeo and Juliet: Act 3, Scene 3'. The researcher told the participants that their goal is to finish their rehearsal in a limited amount of time.

Each time the participant finished reading a line, the virtual instructor provided negative feedback in a number of ways including harsh language, negative non-verbals, encroaching on personal space, and ridiculing the participants' performance. Although the negative feedback from virtual instructor were scripted and identical for all participants, participants were told the feedback was tailored based on their performance and they should follow the instructor's directions to the best of their ability. At the conclusion of the experiment, participants were debriefed about the scripted and non-authentic nature of the "feedback".

To invoke negative affect, the system utilized social interaction and an impossible task framework as mechanisms. All feedback given by the virtual instructor, regardless of actual performance, were designed to be negative and variable in nature. For example, "Woah woah woah, stop. You are sounding way too excited. Let us do it again, and tone it down. Casual-like.", "Ugh, stop. You sound like a dead fish... Let's do it again, and put a little more energy into it.", "Hang on. You are giving it too much energy. Try bringing it down a notch, okay?". In order to get the participants engaged in the experiment, the participants were given a time limit for each line they read. They would be interrupted by the virtual instructor if they could not finish it in time.

The nonverbal behavior such as gesture, facial expression, gaze and posture of the virtual agents are generated by Cerebella [30, 31] to convey negative affect. Cerebella is an intelligent framework which can take a communicative intent as input and generates a multimodal nonverbal behavior commands using the Behavior Markup Language (BML). Taking advantage of the 3D environment, the proxemics between virtual characters could also be manipulated as shown in Figure 3.

4. EXPERIMENT METHOD

4.1 Participants and Design

63 students from two universities (31 men and 32 women), with an average age of 21.37 (SD = 3.06), were randomly assigned to one of 4 conditions in a 2 (virtual human gender) * 2 (approach/no-approach) between-subjects design. The gender balance and average age across conditions was similar. There were 15 participants in condition 1(Male Approach, $N_m = 8$, $N_f = 7$, $M_{age} = 21.33$, $SD_{age} = 3.52$),

Figure 2: The example of negation gesture when the virtual instructor said "No, that's not right. Honestly, how hard is it?"

Figure 3: The example of the approaching behavior. The virtual agent started from 1 meter away from the participants and stopped right in front of the participant's face. Note: image sequence captured from the participant's perspective

18 participants in condition 2(Female Approach. $N_m = 9$, $N_f = 9$, $M_{age} = 21.72$, $SD_{age} = 3.53$), 14 participants in condition 3(Male No-approach, $N_m = 6$, $N_f = 8$, $M_{age} = 21.29$, $SD_{age} = 3.05$), and 16 participants in condition 4(Female No-approach, $N_m = 9$, $N_f = 7$, $M_{age} = 21.06$, $SD_{age} = 3.07$).

4.2 Measures

4.2.1 Subjective Measurement

Rosenberg Self Esteem Scale. The Rosenberg Self-Esteen Scale (RSE) was used to measure pre-test self-esteem levels among participants[36]. The RSE consists of 10 items. People were asked to evaluate each item on a 4-point Likert scale from Strongly Agree (3) to Strongly Disagree (0).

Positive and Negative Affect Schedule-Expanded Form (PANAS-X). The general dimension PANAS-X scales of affect were included in this study scale [47]. Items were measured on a 5-point Likert scale ranging from (1) "*Very slightly*" to (5) "*Extremely*". PANAS-X includes positive affect items such as active, determined and strong as well as negative affect items such as afraid, nervous and distressed. Participant responses were collected pre and post the experiment.

Revised Causal Dimension Scale II. The Revised Causal Dimension Scale (CDSII) [32] was used to measure assignment of causal attributions after the conclusion of the experiment. The CDSII consists of four individual dimensions, "locus of causality" (internality), "stability", "personal control", and "external control". Responses are made on 9-point semantic differential scale with anchoring statements at either end of the scale. For example, one item examines the "stability" with '*changeable*' as the anchoring statement on the left side of the scale and '*unchangeable*' as the anchoring item on the right side of the scale.

In addition to the original items, items tailored for this experiment were included as slight modifications to the existing items. For instance, "*Professor can regulate : professor cannot regulate*", "*That reflects an aspect of yourself : reflects an aspect of the professor*" and "*Something about you : something about the professor*".[1].

[1]A complete list of items included in this modified scale can be found at www.northeastern.edu/cesar/project/VH

Ad-hoc Questionnaire. Upon completion of the experiment, participants were also asked to respond to an ad-hoc questionnaire on a 5-point Likert scale designed to comprehensively evaluate the system's performance. We first asked the participants to evaluate the virtual human with regard to valance (e.g. likable), dominance, activity, attributions of the professor's behavior (e.g. due to his/her personality, his/her temporal mode or participants' underperformance). Additionally, we asked whether people would report the incident to their family, friends or fellow students by asking the question "*will you report this situation to [fellow students/ the university/ family member/ friends]?* ". Each item was evaluated individually in the later analysis.

4.2.2 Objective Measurements

Head Mounted Display (HMD) Movements. To reiterate, this study involved the use and application of a three dimensional virtual environment, which was essentially a model of a superordinate space. In this particular three dimensional space, the environment was not designed to appear bounded to a specific space, such as an enclosed room. Rather, the characters in the virtual environment appeared and interacted with the participants in a seemingly limitless environment, with every point in the space identified by three coordinates (x, y, and z), with each point representing a specific data point for analysis. The movement of each participant's head along the x, y, and z planes in this study was tracked by a three-axis sensing system integrated within the head mounted display.

Skin Conductance Level (SCL). Emotion is a set of interactions with subjective and objective factors, which can give rise to increases in arousal and activate extensive physiological change. Therefore, the emotional state of the participant is evaluated by objective methods (skin conductance levels) in addition to subjective methods (PANAS-X). Skin conductance level is the electrical characteristics of human skin that varies with the state of sweat glands in the skin. Since sweating is controlled by sympathetic nervous system, the SCL could be used as a measure of arousal. Increasing SCL means the arousal state of autonomic nervous system is increased and the human subject has a higher level of agitation. The SCL measurement devices used in this project

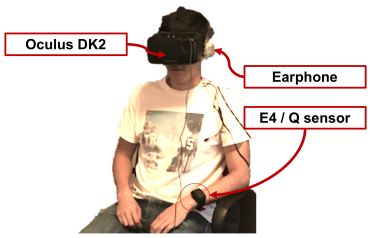

Figure 4: System Apparatus

were the Empatica E4 and the Affetiva Q sensor. SCL was measured over the course of the experiment, including at baseline (base), before approaching interaction(start), after approaching interaction(approach), and at post-task periods(post).

4.3 Apparatus

The 3D virtual environment was developed using Unity3D. The framework of virtual human and character animation is developed based on the Virtual Human Toolkit. The head mounted display (HMD) is Oculus Rift Development Kit 2. The experiment apparatus is shown in Figure 5.

4.4 Procedure

Prior to arrival, participants were randomly assigned to one of four conditions (male approach, female approach, male no approach, female no approach). Participants were informed about the experiment and their role in the study before being instructed to read and, if they agreed to participate in the study, to sign the informed consent form. As they began to read the informed consent form, participants were fitted with the E4/ skin conductance measure bracelet.

Participants were asked to sit and relax for at least 3 minutes to allow the various skin conductance measures to reach an appropriate baseline level prior to the start of the actual experiment in the virtual environment. Participants were then asked to read an instructions sheet that broke down the specific nature of the experiment in greater detail. Participants were informed that they would be evaluated by a virtual professor based on his/her acting performance and that they should react and adjust their performance according to the feedback being received. Here, the participants were implicitly informed of the gender of the virtual professor, as separate instruction sheets included pictures of a virtual male and female professor, depending on the condition. After completing this briefing session, participants were asked to fill out the PANAS-X (pre-test) and the Rosenberg Self-Esteem Scale. After completing those two questionnaires, participants were fitted with the HMD and headphones at an appropriate distance of about 5 feet from the HMD sensor. When the participant reached a comfortable state and indicated readiness, the virtual acting rehearsal began.

Upon completing the experiment, the participants were asked to fill in the PANAS-X (post-test), CDS II, ad-hoc questionnaire, and basic demographics form. A secondary function of the post-test was to allow the participants to rest for at least 3 minutes to collect the post-experiment physiological data. Finally, participants were debriefed on the goals and nature of the experiment and were informed that the negative feedback they received was not authentic. Participants were asked about their reactions and were encouraged to ask questions and/or provide feedback. Each session for a given participant lasted no more than 30 minutes.

5. RESULTS

5.1 Data Preparation

PANAS-X. We calculated the change in positive and negative affect (PANAS-X) between pre and post to assess if there was a significant change in participant responses to the PANAS-X items after being exposed to the experiment. As the reliability of difference scores has been questioned [34, 40], change score variables were calculated by generating the unstandardized residual values (difference between observed and predicted values, $e = y - \hat{y}$) between pre-test (as independent) and post-test (as dependent) with linear regression in SPSS 24. The newly computed variables representing the residualized change of positive affect items as well as negative affect were high in reliability (Cronbach's alphas were .81 for residualized positive affect; .88 for residualized negative affect).

CDSII. Factor analysis was conducted on individual subscales that make up the Causal Dimension Scale II. Five items under the "Locus of causality" dimension of the CDSII were examined via principal components analysis as the primary purpose was to establish and compute composite variables for each subscale of the CDSII. All five items loaded onto one factor and were retained under a "locus of causality" composite measure. The five items under this measure had a Cronbach's alpha of .89. Three items under the "personal" dimension of the CDSII were examined via principal components analysis and were all found to load on one factor. The three items under this measure had a Cronbach's alpha of .83. Three items under the "stability" dimension of the CDSII were examined via principal components analysis. Though all three items loaded onto one factor, one item ("Permanent : Temporary") had a relatively low loading of .59. Indeed, the three items had low reliability with a Cronbach's alpha of .48. Six items under the "external" dimension of the CDSII were examined via principal components analysis using varimax rotation. Three items ("Over which others have control : Over which others have no control", "Professor can regulate : Professor cannot regulate", "Other people can regulate : Other people cannot regulate" did not load on the first factor and were dropped from the composite "external" measure. To note, these three items were also dropped from the overall composite causality measure above. The three remaining "external" dimension items had a Cronbach's alpha of .71.

SCL. Mean Skin Conductance Level (SCL) was measured over 4 phases (base, start, approach, post) of the experiment. This data was processed in one second windows and then averaged with each of the 4 phases. These means were subjected to repeated measures analysis of variance (ANOVA) with block as the repeated measure.

We calculated the change in SCL between the start of the experiment and the approach phase of the experiment to assess if there were significant differences between the changes of SCL during the experiment. As such, we computed residualized change scores between the baseline and approach values of SCL.

Head Movement. We calculated the amount of head movement at each time point when any virtual human/player was speaking to assess the differences in head movements

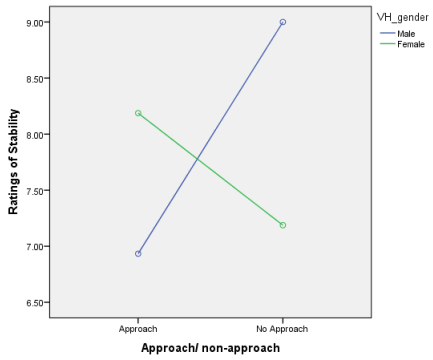

Figure 5: Interaction between Approach Conditions and VH Gender Conditions on Stability dimension of CDSII

across the various conditions– particularly between the approach and non-approach conditions to examine if there were significant differences between the changes of head movement during the experiment. As such, we computed residualized change scores between the start and end values of head movement for the x, y, and z axes.

5.2 Statistical Analyses

To understand if there is an interaction between the two independent variables (IVs), i.e. VH Gender (Male, Female) and Approach (Approach, Non-approach) on the dependent variables(DVs) in question (CDSII, PANAS-X, and Ad-hoc questionnaire measures), we firstly conducted a series of two-way MANOVAs (Multivariate Analysis of Variance) in order to examine effects at both the multivariate and univariate levels of CDS II dimensions and PANAS-X. The Ad-hoc questionnaire items were examined only at the univariate level because its items did not constitute a scale.

5.2.1 Multivariate Analyses

CDSII While no multivariate main effects for either independent variable (IV) were found here, a marginal multivariate interaction effect was found for Approach and VH Gender when it came to dimensions of the factor analyzed CDSII ($F(4,54) = 2.558$, $p = .049$). Univariate analyses revealed that the stability dimension of the CDS accounted for much of the observed multivariate interaction effect. In other words, subjects who interacted with a Male VH professor that Approached reported lower levels of stability than those who interacted with Male VH professor that did not Approach. On the contrary, subjects who interacted with a Female VH professor that Approached reported higher levels of stability than those who interacted with a Female VH that did not approach (See Figure 5).

PANAS-X No main nor interaction effects were found for either IV when it came to composite measures of residualized change for positive and negative affect.

Ad-hoc Specific items of the questionnaire that accounted for this main effect included perception of the helpfulness of the feedback ($F(1,59) = 5.989$, $p = .017$), perception of the accuracy of the feedback ($F(1,59) = 5.505$, $p = .022$), the perceived likability of the virtual instructor ($F(1,59) = 4.850$, $p = .032$) and attributions of the instructor's negative feedback to one's own underperformance ($F(1,59) = 5.649$, $p = .021$). We did not find a main effect for Approach conditions nor an interaction between the two IV's on the Ad-hoc Questionnaire.

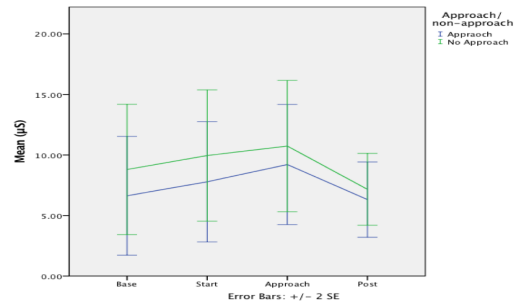

Figure 6: Mean SCL by Block. There is no significant difference between approach and no-approach conditions. But in the approach condition, the SCL changes more from start phase to approach phase than no-approach condition.

SCL A mixed-model ANOVA was conducted for SCL with 4 phases (base, start, approach, post) as a within-subject variable and approach conditions as a between-subject variable. The result revealed a significant block effect, $F(3, 48) = 11.08$, $p < .001$, but no significant effect of approach conditions on the SCL values.

5.2.2 Proxemics-based Results

When the virtual human approached the participants, participants were more concerned about their performance relative to when the virtual human did not approach (n) than approach (a) ($M_a = 2.88$, $SD_a = .99$, $M_n = 3.43$,$SD_n = 1.25$, $t(61) = -1.958$, $p = .055$). Participants in these approach conditions also tended to externally attribute their poor performance to the virtual instructor's personality ($M_a = 3.55, SD_a = 1.06, M_n = 4.17, SD_n = .83, t(61) = -2.562, p = .013$)

While participants in the approach conditions did have greater SCL level changes than those in non-approach conditions, no significant differences in SCL were observed between the approach conditions and the non-approach conditions, counter to our hypotheses (See Figure 6). We did however, find significant differences for participants' head movements in response to the experiment.

In the approaching conditions, participants tended to move their heads further away from the virtual human, which can be interpreted as an attempt to maintain a certain degree of personal space as the virtual human invaded it. Participants in approach conditions demonstrated a marginally significant difference when the virtual instructor started to walk towards the participants($M_a = 14.58mm$, $SD_a = 65.09$, $M_n = -11.5mm$, $SD_n = 26.47$, $t(48) = 1.90$, $p = .060$) as well as a statistically significant change in head position after the virtual human starts in front of the participants face ($M_a = 18.76mm$, $SD_a = 64.43$, $M_n = 14.74mm$, $SD_n = 24.49$, $t(48) = 2.438$, $p = .019$). Participants' head height (y axis) in approach conditions demonstrated a marginally significant difference when the virtual instructor started to walk towards the participants ($M_a = 2.93mm$,$SD_a = 7.36$, $M_n = 2.30mm$, $SD_n = 11.04$, $t(48) = 1.91$, $p = 0.06$) as well as a statistically significant change in head height after the virtual human starts in front of the participants face ($M_a = 3.83mm$, $SD_a = 12.60$, $M_n = -3.00mm$, $SD_n = 11.18$, $t(48) = 2.032$, $p = 0.048$).

5.2.3 VH Gender-based Results

When interacting with female instructor, participants at-

tributed more to themselves (Locus of causality) compared to those participants interacting with a male instructor ($M_m = 22.55, SD_m = 9.67, M_f = 27.09, SD_f = 8.08, t(59) = -2.00, p = .050$). In line with internal attribution, when interacting with a female instructor, participants felt they had more personal control($M_m = 15.10, SD_m = 6.04, M_f = 18.15, SD_f = 4.60, t(60) = -2.25, p = .028$). The result from Ad-hoc scale further confirmed this attribution result.

Participants who interacted with a female virtual instructor reported greater positive affect than those who interacted with a male virtual professor. Specifically, these participants felt more inspired when they interacted with the female virtual instructor than those who interacted with the male instructor($M_m = 15.10, SD_m = 6.04, M_f = 18.15, SD_f = 4.60, t(60) = -2.91, p = .005$).

Given the consistent nature of negative feedback, the ad-hoc measurements demonstrated that participants in general preferred to receive the negative feedback from the female virtual instructor as opposed to the male virtual instructor. Specifically, the female virtual instructor was found to be more helpful ($M_m = 15.10$, $SD_m = 6.04$, $M_f = 18.15$, $SD_f = 4.60$, $t(60) = -2.48$, $p = .016$), more accurate ($M_m = 15.10$, $SD_m = 6.04$, $M_f = 18.15$, $SD_f = 4.60$, $t(60) = -2.41$, $p = .019$), and more likable ($M_m = 15.10$, $SD_m = 6.04$, $M_f = 18.15$, $SD_f = 4.60$, $t(60) = -2.25$, $p = .026$) than the male virtual instructor.

6. DISCUSSION AND FUTURE WORK

Recall that one of our hypotheses was that subjects would exhibit a greater tendency to externally attribute failure to the female professor in comparison to the male professor. Interestingly, what we found was the opposite: subjects tended to attribute blame externally more to the male professor than the female professor. One possible explanation here follows from factoring in the emotional and coping responses of the subjects. The female professor may be seen as less threatening and as a consequence less blameworthy. Indeed, the fact that subjects reported having a more positive affect in response to the female professor supports that explanation. Teasing apart such an interplay will require further analysis in future studies. Regardless, the results suggest a difference in how subjects attribute causality when interacting with a female versus male virtual human that is useful in the design of applications.

Another hypothesis was that the approach would impact arousal and lead to a negative impact on the cognitive and other affective measures. This was not seen in our results. A possible explanation here concerns the current design of the experience. The approach behavior occurs in the last five seconds of the experiment, potentially leaving insufficient time for the approach to have an impact on our measurements of arousal. By the time the approach behavior occurs at the end of the experiment, participants would have engaged coping strategies to normalize the negative feedback from the instructor. Applying the approach behavior towards the beginning or the middle of the experiment in future studies should elevate the impact and intensity of the approach on the participants.

The present study presented a number of limitations. The first notable limitation involves the limited nature of the stimuli for the experiment. One virtual male instructor character (Brad) as shown in Figure 3 and one virtual female instructor character (Olivia) as shown in Figure 2 were used in this study, with each having a defined appearance and a defined voice/tone. The issue with having a single defined appearance and speech is that we are making a generalized claim about virtual human gender when in fact the respective virtual genders are each being represented by a single character. This presents multiple issues. First, the negative feedback being administered by each character is not necessarily controlled in terms of their voice and tone, and participants may be responding to a male virtual character that objectively looks and sounds more hostile than the female virtual character. Second, by using only one virtual human per gender, we are unable to sample from a wider scope of appearances and speech patterns that would afford greater generalizability claims needed for a claim about "gender differences" per se.

Another notable limitation of this study stems from the integration of two distinct sample populations. While one group of participants were recruited from a Computer Science department in the northeast region of the US, another group of participants were recruited from a Psychology department in the west coast of the US. Aside from obvious differences in region and discipline, we acknowledge that there may have been distinct gender effects occurring that were unique to each respective sample.

This issue on samples touches on another limitation of the present study: the lack of exploration of subject gender effects. We recommend that future studies further explore the distinct responses of individual subject gender to each respective condition of this study, particularly the gender of the virtual instructor. Indeed, it would be interesting to monitor the patterns of attributions that emerge across separate combinations of subject gender to virtual instructor gender interactions.

Another limitation of the study was reflected in the lack of significant reportable results in the Skin Conductance Levels of the respective participants. One plausible cause of this result – or lack thereof – may be attributed to inconsistency in the hardware as well as inconsistency in the administering of the measurement. Indeed, establishing consistency and testing the complete functionality of the hardware across multiple researchers should be a point of emphasis for future studies. Another potential reason of the lack of effect is the late introduction of the approach behavior in the approach conditions of the experiment. A lack of arousal in non-approach conditions was generally expected, but the lack of noticeable differences between approach and non-approach conditions indicated a potential weakness or lack of prominence in the nature of the approach stimulus.

Finally, a potential issue with the Skin Conductance measure is the method of measurement. As mentioned above, the method of measurement in the present study was a measurement of the change scores between the initial SCL and the SCL at the end of the experiment. As the negative feedback stimulus involved a complete interaction that lasted several minutes, fluctuations and variations between the start and end of the experiment should not only be possible but expected. An alternative method of measurement would be to focus on event-related measurements of SCL. That is, a future study should examine SCL's at every utterance (and proxemic movement) of negative feedback from the virtual instructor to the participant. The present study made use of a very strategically designed negative feedback script that employed the use of communication theory as

well as professional voice actors. An alternative would be to examine the differential effects of the various negative feedback attempts made by the virtual instructor on the event-related SCL's of the participants throughout the experiment – rather than operationalizing the totality of the experiment as a single event. Additionally, future work should include more participants to be sufficiently powered to examine the nuances of these limitations.

7. CONCLUSION

The results of the present study have numerous implications for the design of virtual agents for learning outcomes as well as the methodological design of studies utilizing virtual agents in virtual environments. The results of the attribution analyses reveal a tendency to attribute greater blame to male virtual characters than female virtual characters in a negative feedback situation. As mentioned earlier, social situations in human-human communication naturally fluctuate in valence and nature. The results of the present study demonstrate evidence that female characters should be considered in favor of male characters when designing games and interventions that simulate negatively-valenced and/or emotionally-charged social situations such as a conflict-resolution.

This study is also very relevant and applicable in education and pedagogy. Since the effects of negative feedback on learning have been well-established, teachers and educators tend to avoid negative feedback in the learning process. That said, this does not preclude all educators from employing negative feedback in their teaching. Some teachers have bad days and some teachers are quite simply ineffective teachers. By simulating a negative feedback teaching situation in a virtual environment, we present the potential for a more precise understanding of the effects of negative feedback on students' learning, emotional state, attribution patterns, and even their nonverbal reactions to the negative feedback. For obvious reasons, it simply would not be practical or feasible to test for these effects of negative feedback in an actual teaching environment.

Finally, the study presents a number of implications for the design of virtual agents as well as the design of studies using virtual agents. Future studies should integrate a greater variety of virtual agents to be used as stimuli in order to obtain generalizable results about virtual agent gender differences. Further, the lack of expected effects and differences between approach and non-approach conditions suggest that in studying nonverbal movements within social situations, the movement patterns should be made more prominent and distinct throughout the entirety of the stimulus. In sum, while there are many implications for the analytical results of the present study, the present study also presents a number of critical diagnostic findings that will aid in the future design of studies utilizing virtual agents.

Acknowledgments

We thank Stephen J. Read for his help in selecting psychological measures for this study and his suggestions on modifications of the CDSII scale. This research was supported by the Air Force Office of Scientific Research under grant FA9550-14-1-0364. Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

REFERENCES

[1] I. Altman and A. M. Vinsel. Personal space. In *Human behavior and environment*, pages 181–259. Springer, 1977.

[2] A. L. Baylor. Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1535):3559–3565, 2009.

[3] A. L. Baylor and Y. Kim. Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. In *International Conference on Intelligent Tutoring Systems*, pages 592–603. Springer, 2004.

[4] J. Blascovich. 25 challenge and threat. *Handbook of approach and avoidance motivation*, page 431, 2013.

[5] J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson. Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, 13(2):103–124, 2002.

[6] J. Blascovich, W. B. Mendes, S. B. Hunter, and K. Salomon. Social" facilitation" as challenge and threat. *Journal of personality and social psychology*, 77(1):68, 1999.

[7] P. Brown and S. C. Levinson. Politeness: Some universals in language use. 1999.

[8] J. D. Campbell and A. Tesser. Motivational interpretations of hindsight bias: An individual difference analysis. *Journal of Personality*, 51(4):605–620, 1983.

[9] C. S. Carver. Resilience and thriving: Issues, models, and linkages. *Journal of social issues*, 54(2):245–266, 1998.

[10] M. V. Covington and C. L. Omelich. Effort: The double-edged sword in school achievement. *Journal of Educational Psychology*, 71(2):169, 1979.

[11] M. V. Covington and C. L. Omelich. Task-oriented versus competitive learning structures: Motivational and performance consequences. *Journal of Educational psychology*, 76(6):1038, 1984.

[12] R. De Charms. *Personal causation: The internal affective determinants of behavior*. Routledge, 2013.

[13] C. S. Dweck. The role of expectations and attributions in the alleviation of learned helplessness. *Journal of personality and social psychology*, 31(4):674, 1975.

[14] C. S. Dweck, W. Davidson, S. Nelson, and B. Enna. Sex differences in learned helplessness: Ii. the contingencies of evaluative feedback in the classroom and iii. an experimental analysis. *Developmental psychology*, 14(3):268, 1978.

[15] C. S. Dweck and E. L. Leggett. A social-cognitive approach to motivation and personality. *Psychological review*, 95(2):256, 1988.

[16] J. Gratch, A. Okhmatovskaia, F. Lamothe, S. Marsella, M. Morales, R. J. van der Werf, and L.-P. Morency. Virtual rapport. In *International Workshop on Intelligent Virtual Agents*, pages 14–27. Springer, 2006.

[17] E. T. Hall. *Proxemics: The study of man's spatial relations*. 1962.

[18] J. Hattie and H. Timperley. The power of feedback. *Review of educational research*, 77(1):81–112, 2007.

[19] S. Heidig and G. Clarebout. Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review*, 6(1):27–54, 2011.

[20] R. Ilies and T. A. Judge. Goal regulation across time: the effects of feedback and affect. *Journal of applied psychology*, 90(3):453, 2005.

[21] C. M. Jagacinski and J. G. Nicholls. Competence and affect in task involvement and ego involvement: The impact of social comparison information. *Journal of Educational psychology*, 79(2):107, 1987.

[22] W. L. Johnson and J. C. Lester. Face-to-face interaction with pedagogical agents, twenty years later. *International Journal of Artificial Intelligence in Education*, 26(1):25–36, 2016.

[23] 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*, 26(1):160–169, 2016.

[24] O. Klein and L. Licata. Explaining differences between social groups: The impact of group identification on attribution. *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie*, 60(4):244, 2001.

[25] A. N. Kluger and A. DeNisi. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin*, 119(2):254, 1996.

[26] J. Kolkmeier, J. Vroon, and D. Heylen. Interacting with virtual agents in shared space: Single and joint effects of gaze and proxemics. In *International Conference on Intelligent Virtual Agents*, pages 1–14. Springer, 2016.

[27] N. C. Krämer and G. Bente. Personalizing e-learning. the social effects of pedagogical agents. *Educational Psychology Review*, 22(1):71–87, 2010.

[28] N. C. Krämer, B. Karacora, G. Lucas, M. Dehghani, G. Rüther, and J. Gratch. Closing the gender gap in stem with friendly male instructors? on the effects of rapport behavior and gender of a virtual agent in an instructional interaction. *Computers & Education*, 99:1–13, 2016.

[29] M. R. Lepper and D. Greene. Turning play into work: Effects of adult surveillance and extrinsic rewards on children's intrinsic motivation. *Journal of personality and social psychology*, 31(3):479, 1975.

[30] M. Lhommet and S. C. Marsella. Gesture with meaning. In *International Workshop on Intelligent Virtual Agents*, pages 303–312. Springer, 2013.

[31] S. Marsella, Y. Xu, M. Lhommet, A. Feng, S. Scherer, and A. Shapiro. Virtual character performance from speech. In *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 25–35. ACM, 2013.

[32] E. McAuley, T. E. Duncan, and D. W. Russell. Measuring causal attributions: The revised causal dimension scale (cdsii). *Personality and Social Psychology Bulletin*, 18(5):566–573, 1992.

[33] C. McCall, D. P. Bunyan, J. N. Bailenson, J. Blascovich, and A. C. Beall. Leveraging collaborative virtual environment technology for inter-population research on persuasion in a classroom setting. *PRESENCE: Teleoperators and Virtual Environments*, 18(5):361–369, 2009.

[34] L. A. McFarland and A. M. Ryan. Variance in faking across noncognitive measures. *Journal of Applied Psychology*, 85(5):812, 2000.

[35] S. B. Nolen and T. M. Haladyna. Motivation and studying in high school science. *Journal of Research in science teaching*, 27(2):115–126, 1990.

[36] E. A. Plant, A. L. Baylor, C. E. Doerr, and R. B. Rosenberg-Kima. Changing middle-school students' attitudes and performance regarding engineering with computer-based social models. *Computers & Education*, 53(2):209–215, 2009.

[37] N. L. Schroeder, O. O. Adesope, and R. B. Gilbert. How effective are pedagogical agents for learning? a meta-analytic review. *Journal of Educational Computing Research*, 49(1):1–39, 2013.

[38] R. Smith and J. Leang. Evaluations of correlations for two-phase flowmeters three current–one new. *Journal of Engineering for Power*, 97(4):589–593, 1975.

[39] R. Sommer. Personal space. the behavioral basis of design. 1969.

[40] J. Tisak and C. S. Smith. Defending and extending difference score methods. *Journal of Management*, 20(3):675–682, 1994.

[41] J. C. L. S. G. Towns, C. B. Callaway, and J. L. V. P. J. FitzGerald. Deictic and emotive communication in animated pedagogical agents.

[42] R. J. Vallerand and G. Reid. On the causal effects of perceived competence on intrinsic motivation: A test of cognitive evaluation theory. *Journal of Sport Psychology*, 6(1):94–102, 1984.

[43] D. Van-Dijk and A. Kluger. Positive (negative) feedback: Encouragement or discouragement. In *th Annual Convention of the Society for Industrial and Organizational Psychology, New Orleans, Louisiana*, 2000.

[44] G. Veletsianos and G. S. Russell. Pedagogical agents. In *Handbook of research on educational communications and technology*, pages 759–769. Springer, 2014.

[45] N. Wang and J. Gratch. Rapport and facial expression. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–6. IEEE, 2009.

[46] N. Wang, W. L. Johnson, R. E. Mayer, P. Rizzo, E. Shaw, and H. Collins. The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies*, 66(2):98–112, 2008.

[47] D. Watson and L. A. Clark. The panas-x: Manual for the positive and negative affect schedule-expanded form. 1999.

[48] B. Weiner. A theory of motivation for some classroom experiences. *Journal of educational psychology*, 71(1):3, 1979.

[49] B. Weiner, D. Russell, and D. Lerman. Affective consequences of causal ascriptions. *New directions in attribution research*, 2:59–90, 1978.