The Mere Belief of Social Interaction Improves Learning

Sandra Y. Okita (yuudra@stanford.edu)
School of Education , Stanford University
485 Lasuen Mall Stanford CA 94305-3096 USA

Jeremy Bailenson (bailenson@stanford.edu)
Department of Communication, Stanford Universiy
450 Serra Mall Stanford, CA 94305-2050

Daniel L. Schwartz (danls@stanford.edu)
School of Education , Stanford University
485 Lasuen Mall Stanford CA 94305-3096 USA

Abstract

Thirty-five adult participants tested the hypothesis that one's mere belief in having a social interaction with someone improves learning and understanding. Participants studied a passage on the body's mechanism for causing fever. They then entered a virtual reality environment with an embodied agent on the other side of a table. The participant read scripted questions relevant to the fever passage, and the agent gave scripted responses. In the Avatar condition, participants heard that the virtual representation was controlled by a person whom they had just met. In the Agent condition, participants heard that the virtual representation was computer controlled. The Avatar condition yielded better learning and inference at posttest, even though all interactions within VR were held constant across conditions. Skin conductance measures also indicated that the Avatar condition exhibited more arousal and that higher arousal was correlated with learning on a problem-by-problem basis. Further results suggest the hypothesis that the learning effect was not due to social belief per se, but rather in the belief of taking a socially relevant action.

Keywords: Learning; virtual environments; agents; avatars.

Introduction

Virtual reality (VR) permits novel investigations of what it means to be social. For example, it is possible to tell participants that they are interacting with an embodied agent that is fully controlled by a computer. Alternatively, participants can be told that they are interacting with an embodied avatar that is being controlled by a person. In this research, we examined whether simply believing a virtual representation was an agent (computer) or an avatar (person) affected learning.

Recent research on virtual reality and other new media has examined what features cause people to treat a computer representation as a social being (e.g., Bailenson et. al 2005; Reeves & Nass, 1996; Schroeder, 2002). A different sort of question asks if differences arise when people believe they are interacting with a person or whether they are interacting with a machine, when all features are otherwise held constant. Research indicates that people's interaction patterns differ depending on whether they believe they are interacting with an agent or an avatar (Bailenson, Blascovich, Beal & Loomis, 2003; Blascovich et. al., 2002; Hoyt, Blascovich & Swinth, 2003). For example, people will respect the virtual "space" of a human representation if they believe it is an avatar.

Our particular interest in virtual reality is that it provides a unique way to examine the effects of social interaction on learning. Social interaction is a natural and powerful way to learn. One important aspect of social interaction is that it can generate well-tuned feedback, as in the case of a tutor. Another important aspect is that social actors can provide models that learners might imitate. In the current work, we explore whether the mere belief that an interaction is with another person, and therefore social, influences learning. Neurological evidence indicates that attributions of humanness to a displayed image recruit different brain circuitry (Blakemore, Boyer, Meltzoff, Segebarth & Decety, 2003), but the effect of human attributions on learning is unknown, particularly if all other visual features and interactive opportunities are held constant.

In the study, participants engaged in a scripted Q & A session with a computer agent on the mechanisms behind maintaining a fever. Participants were told that they were either interacting with an avatar or agent. Afterwards, we gave a posttest on the mechanisms of fever to see which condition led to deeper understanding. In addition to learning measures, we also collected measures of participant arousal (skin conductance levels). These measures, which were taken every 1/60th of a second in VR, can help reveal the time course of how a belief in a social interaction influences learning. Moreover, they can show whether the belief that one is interacting with another person increases arousal.

Prior research indicates that moderate levels of arousal at encoding are correlated with better "factual" memory, but to our knowledge, no research has demonstrated that arousal at encoding is correlated with deeper understanding.

Method

Participants

Thirty-five (17 female, 18 male) college students were paid to participate in the study. They were randomly assigned to one of two conditions.

Design and Procedure

The study had a between subject condition with two levels: Avatar and Agent. It also had a within-subject factor that had three levels: Manner of computer response. Figure 1 provides a schematic of the design and procedural flow for the between-subjects factor.

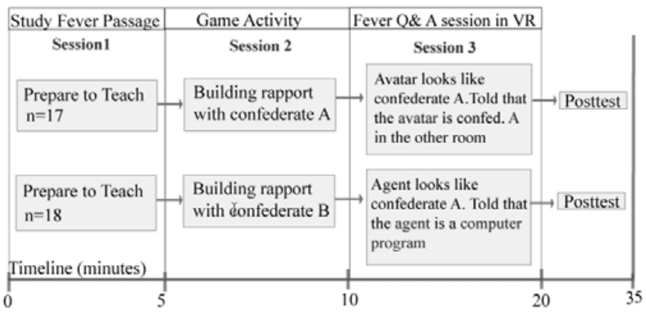

Figure 1: Procedural Flow and Design

Participants completed the study individually. They first studied a one-page passage on fever for 5 minutes and were told to "prepare to tutor" a student about the fever mechanism. Participants then had the passage taken away. They were then introduced to a confederate posing as another participant named "Alyssa." They played the child's game of Operation with Alyssa for 5 minutes so they could get to know one another (in the physical world). The participant then entered their experimental condition and moved to a separate room (See Figure 2). There the participant was told they would be interacting with a computer agent or Alyssa's avatar. The interaction was held constant in the virtual reality setting, and all interactions were made identical. There was no avatar; it was always a computer program that simply played identical pre-recorded verbal and nonverbal responses in both conditions. The way the interactions were held constant was as follows: participants saw a question about the fever passage that showed up on the monitor beside the avatar/agent (See Figure 3). The participant then called out to the avatar/agent and asked the question shown on the monitor. For example, for one question in the avatar condition, participants said, "*Alyssa*, why do your hands and feet get cold when you have a fever?" In the agent condition, participants said, "*Computer*, why do your hands and feet get cold when you have a fever?" After reading the question, the participants pressed a button, and the computer character responded to the question using a pre-recorded answer (the participants were not told it was pre-recorded). The pre-recorded response was always only partially correct and never included incorrect information. Participants read 9 questions from the screen and heard 9 answers. The question order was randomized across participants.

The within-subject factor manipulated the manner of the pre-recorded response: exhilarated, neutral, or shameful. For each question, we pre-recorded three manners of response, each using the exact same wording. Participants heard 3 exhilarated responses, 3 neutral responses, and 3 shameful responses. Response manner was randomly assigned across question and order for each participant. The different response types were given to see if they had an influence on learning and arousal. For example, a shameful response, which indicated insecurity about a response, might causes participants to think more deeply about the correct answer.

In review, the study used a 2 x 3 x 3 design with the between-subject factor of Condition (avatar v. agent), the within-subject factor of Response manner (exhilarating, neutral, shameful), and the crossed within-subject factor of Exposure Order to a particular Response manner (1^{st} , 2^{nd}, 3^{rd} exposure).

Material and Measures

Material

The fever passage explained how the human body gets and maintains a fever. It explained the mechanisms that triggers the fever response (e.g., macrophages), the mechanisms that introduce more heat into the body (e.g., shivering), and the mechanisms that prevents the body from releasing heat (e.g. blocking sweat).

Figure 2: Participant in VR learning environment: 1) Head-Mounted Display (HMD) and orientation tracker, 2)monitor showing the experimenter what participant is seeing in the HMD, 3) Game-pad used to notify agent/avatar, 4) rendering computer, 5) equipment recording skin conductance level (SCL).

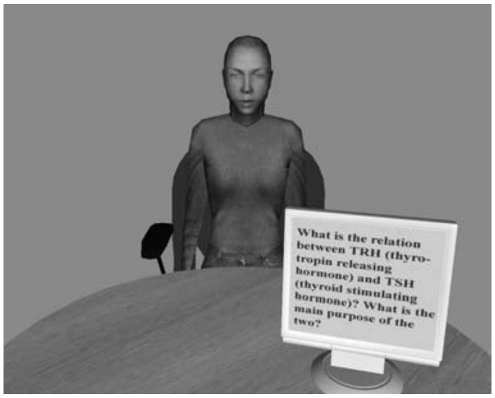

Figure 3: The Viewpoint of a Participant during the Q & A Session in VR

Apparatus

Figure 2 shows a participant wearing the Head Mounted Display (HMD), which allows participants to see and interact in the virtual world. The HMD contains a separate display monitor for each eye (50 degrees horizontal by 38 degrees vertical field-of-view with 100% binocular overlap). The graphics system renders the virtual image separately for each eye for stereoscopic depth at approximately 60 Hz. The software used to assimilate the rendering and tracking was Vizard 2.53. Participants wore a Virtual Research 8 HMD that featured dual 640 horizontal by 480 vertical pixel resolution panels. The biofeedback equipment used to measure the participant's Skin Conductance Level (SCL) is BioGraph Infiniti 3.1 from Thought Technology Ltd.

Measures

There were two measures. One was the posttest that measured learning, and the second was a physiological measure of arousal measured by Skin Conductance Level (SCL). The SCL was measured while the participant was in the VR environment interacting with the agent/avatar.

Table 1: Scoring Method.

Scoring Method (0-2 point scale)		
0: incorrect or no answer	**1**: partially correct but incomplete	**2**:precise and detailed
Why is shivering not enough to create a fever?		
0 point:	"Because its not enough, you need more"	
1 point:	"Because shivering alone creates heat, but the brain is not involved so it doesn't set the temperature set point."	
2 points:	"You can create heat with shivering, but you also need a mechanism that doesn't let that heat escape, so you need the hypothalamus to raise the set point."	

The posttest contained the original 9 questions given in VR during the Q & A session and an additional 6 new questions. The 9 questions given in the VR environment were largely factual (e.g. "What processes cause the body to increase temperature?"), and the 6 additional questions were inferential (e.g. "Why does a dry nose mean a dog might have a fever?") which required a deeper understanding of the fever passage to answer. Each question was scored on a 0 to 2 point scale (1: incorrect/no answer, 2: partially correct but incomplete, 3: precise and detailed). Thus, for the posttest, the maximum possible score was 30. Table 1 provides a sample scoring.

Results

Learning Results

All reported results were tested at the $p < .05$ threshold for significance using repeated measures analyses of variance. The prediction was that the belief of interacting with an avatar would lead to better learning and understanding than a computer program. The results of the posttest given after the VR experience supported this hypothesis. Table 2 shows the percent accuracy for participants in each condition on each portion of the posttest. The results are notable because they show that the avatar effect occurred for the 9 questions heard in VR as well as for the 6 new questions with no appreciable differences in the avatar advantage. This means that avatar participants developed a better model of the domain, which enabled them to draw novel inferences. Moreover, both groups did better on average than the answers supplied in VR (average of 0.7 points worth of information was given by the avatar/agent for each response in VR), so they were not only remembering what they heard.

Table 2: Posttest Score

	Avatar (std error)	**Agent (std error)**
All 15 questions	60% (.032)	49% (.039)
9 Questions in VR	61% (.031)	49% (.041)
6 New questions	58% (.040)	49% (.043)

Table 3: Mean Accuracy Across All Response Manners

	Avatar	**Agent**
Exhilarate Responses	65%	45%
Neutral Responses	61%	54%
Shameful Responses	58%	48%

The second hypothesis was whether the different response manners had an effect on learning outcomes for the 9 questions heard in VR. We examined the accuracy score by three response manners (exhilarate, neutral, shameful) as seen in Table 3. The avatar advantage only appeared for each manner of presentation early on (1^{st} and 2^{nd} trial), and there were no significant differences due to the manner of response per se. Therefore, the effect was not general, it seemed specific to something about hearing a new response type.

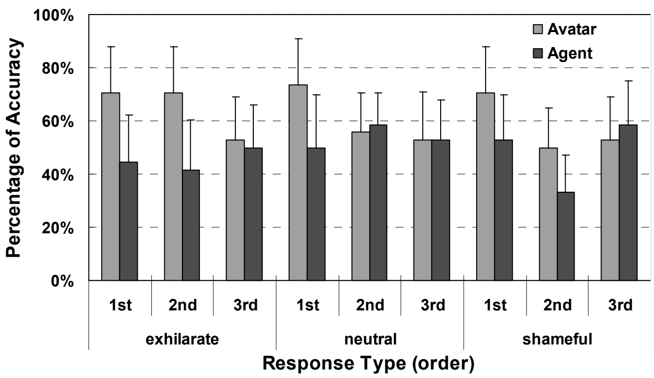

Figure 4: Percent Accuracy by Response Type & Order (error bars equal SE of the mean)

Figure 4 provides closer look at the order of the different response manners. The first two exhilarating, first neutral, and first shameful questions all received high scores for the Avatar condition, while there is no systematic variation in the agent condition. The order x response manner effect is conflated with the absolute order that people saw each question; for example, the first exposure to an exhilarated response is likely to occur early on in the 9 questions. Therefore, it is difficult to separate the effects of absolute order from the effects of response manner exposure (1^{st}, 2^{nd}, or 3^{rd}). In regression analyses, response manner exposure provided a better fit to the posttest accuracy than absolute order, but the two predictors are highly co-linear. In the subsequent analyses, we continue to use Response Manner by Exposure as factors because it was the a priori design of the study.

In summary, the mere belief of interacting with a human led to superior learning of complex material. The learning carried over to new questions not heard in VR. A second analysis examined only the questions heard in VR. Participants in the agent condition showed minimal variability in their posttest scores for each problem, regardless of when they heard the question in VR. In contrast, participants in the avatar condition showed the strongest learning for those questions that were answered with a relatively novel manner of response.

Arousal

Did the belief that they were interacting with another person influence participants' physiological response? And if so, was the physiological data correlated with learning outcomes? To answer these questions, Skin Conductance Level (SCL) was taken while the participant interacted with the agent/avatar in the VR environment.

In preparation for analyzing the SCL data, two steps were taken, 1) normalize and aggregate data and 2) eliminate participants with corrupted data.

Normalizing and aggregating data. The SCL recording equipment took 60 samples per second. This level of precision is favorable for sudden events where the length of the study is a few seconds. However, a coarser aggregation was more appropriate for the current study, which spanned over tens of minutes. In addition, activities varied in length, for example, some questions had longer pre-recorded answers than others. To aggregate across questions it was necessary to normalize the SCL data into relative intervals within an event (i.e., average SCL for the first third of an answer). A third consideration with the SCL data is that different people have different baseline levels of hand perspiration. Therefore, we found people's average baseline before each read-answer event (10 sec blank screen period), and we subtracted this average from their SCL during the read-answer event.

All told, we took the average SCL for each third of the reading and listening phases; Reading (1^{st}-Beginning, 2^{nd}-Middle, 3^{rd}-End), Listening (1^{st}-Beginning, 2^{nd}-Middle, 3^{rd}-End), and took average of the baseline phase, normalizing the value to 6 data points. From each of the six values, the baseline skin conductance value was subtracted to get a measure of change. The process was repeated for each of the 9 question events in VR.

Eliminating corrupted data First, we eliminated 6 participants due to corrupted data, possibly to equipment failure. Second, we identified subjects who showed high variability that averaged .025 micro Siemens (uS) across problems. Higher variability increased the chances of detecting correlations between arousal and learning. As a result, we had 10 participants in avatar condition and 9 participants in agent condition. The remaining participants exhibited less variability. In the following analysis we describe the 19 participants with high variability. Statistical analyses that included the remaining participants yielded highly similar results as the ones described below. However, at the time of this submission, we did not have a chance to re-create the graphs including all participants.

To overcome the concern that the 19 sampled participants were not representative of the learning results of the complete data set, we plotted the posttest score for the 19 subjects broken out by response type and exposure order. The results for the 19 participants and the total of 35 participants have the same pattern and effects.

Figure 5 shows the average SCL when aggregating across response type. The results showed a consistent pattern of peaking during reading, and then declining while listening to the response. The results also showed a significant effect of condition, particularly for the first exposure to a given response type. As a reminder, the response type was delivered from the computer after people showed an increase in arousal during reading. So the effect of response type seemed to diffuse over time. However, early on many of the participants in the avatar condition still thought they were dealing with a real person. One implication may be that it is necessary to have a variety of response types to keep people engaged with technology.

Relation between learning and arousal

So far, we have shown that the avatar participants showed greater learning when they heard a response type for the first time. We also found that avatar participants showed greater arousal when they heard a manner for the first time. This implies that arousal is related to learning. To make a closer examination, we switched to problem as unit of analysis. We back-sorted the arousal measure by score (posttest) by taking the average arousal for all questions where participants scored a 2, scored 1, or a 0.

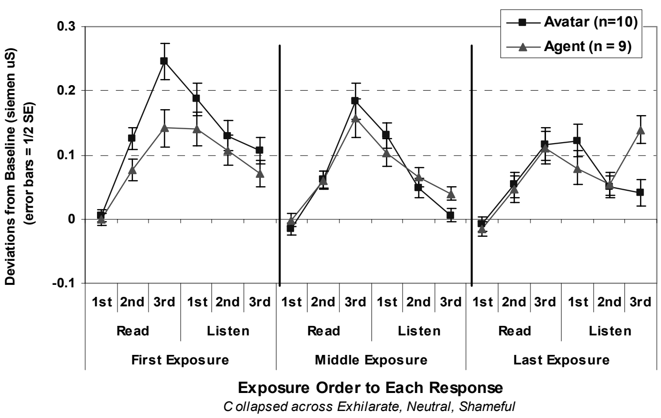

Figure 5: Changes in Skin Conductance Level

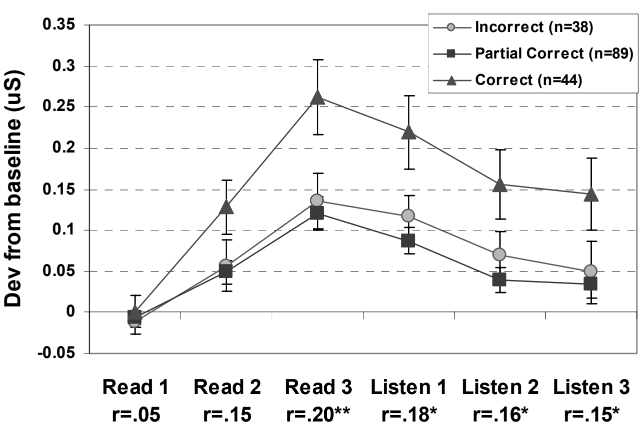

Figure 6: Skin Conductance Back Sorted by Score on Learning

Figure 6 shows the SCL measure in relation to the posttest scores and the read/listen samples. In this problem level analysis, the ***n*** in the right hand corner refers to the number of problems that received full, partial or no credit. For example, out of 171 problems (19subjects x 9 questions), 44 problems scored a 2 (full correct credit), 89 problems received a score of 1, and 38 received a 0. The results indicate that higher scores on the posttest had significantly higher SCL levels during the read-listen event. We also took the correlation of each arousal sample (e.g., 1st-third of reading, 2nd-third of reading, etc.) with the posttest score for that session. The correlations are shown as the bottom row of the figure. The strongest correlation between SCL with the posttest score was at the Read 3 sample (final third of reading phase); r = .20**. The relatively high correlations during the listening phase are most likely due to auto-correlations with the peak level of arousal during reading (SCL dropped off slowly from the peak arousal). This suggests that the arousal period that was most correlated with learning occurred <u>before</u> the participants heard the response. Quite interestingly the highest correlation was in the action, and not in the response. This made the last third of the reading phase the best predictor for learning, meaning arousal occurs when participants were reading not listening.

Discussion

The current study explored whether the mere belief that you interacted with another person makes a difference in how much you learn. Results showed that this "belief" in the avatar condition resulted in a significant learning gain. Although there were learning gains in the agent condition, it was smaller compared to the avatar condition. Again, in both conditions, participants spoke identical words asking questions, and the virtual human provided identical pre-recorded verbal and nonverbal responses.

Even though the agent/avatar did not contribute much information in their response for the Q & A session, a significant learning effect was found for the questions given in the virtual environment and for the additional 6 questions that appeared on the posttest. These new inferences indicated that participants in the avatar condition had made a better model of fever mechanisms.

When we took a closer look at the order of the different response types in Figure 4 we saw that the first two exhilarating, first neutral, and first two shameful questions all received relatively higher scores in the Avatar condition than the Agent condition. Several reasons can be thought of. One was that each time the participants heard a new manner of presentation they paid more attention. The fact that there was little difference in response for the Agent condition may imply that it was not the novelty, but the social belief that drove the participant's interest.

The SCL (arousal) measure showed that the belief in social led to greater arousal, especially on first exposures to a response manner. Greater arousal correlated with better learning on a problem-by-problem analysis. We found that the strongest correlation between SCL (arousal) and the posttest score was most significant at the 3rd sample for reading. What was interesting was that these arousal results suggest a different role for response manner. The learning analysis showed that Avatar condition learned best when they heard a new type of response. The arousal analysis showed that effect of arousal on learning seemed to occur during reading, which was before participants heard the manner of response. Possibly the effect of the manner of the response on learning is somewhat indirect. The different manners of response kept people socially engaged for a while, which in turn led to more generally high arousal.

SCL is a measure that indexes some internal process (we are not claiming that hand sweat causes learning). Our main interest in the SCL measures is that they provide some indication of the time course of processing during each Q & A event. The SCL scores that were most correlated with learning occurred during the reading phase and not the listening phase. This suggests that the locus of the learning effect occurred when people took the socially-relevant action of reading, which in turn, prepared them to learn more deeply when listening to the response. Of course, the relation between the arousal data and learning is only correlational. For example, it is possible that people became more aroused when they read questions where they thought they knew the answer. Nevertheless, the SCL data suggest the interesting hypothesis that the learning effect is not due to a general belief that one is socially interacting with a human. Rather, the effect is that people believes they are taking socially relevant action. The engagement/arousal during this action is what prepares them to learn from the response. In on-going work, we are testing this hypothesis. For example, in a new avatar condition, participants read silently rather than aloud. This way, they cannot take any socially relevant action. If people listen passively to an avatar, they may not learn as well and their arousal signatures would stay low. If so, this might help explain some of the common wisdom that listening is not always as good as interacting. It is the social action, or potential for social action, that prepares one to listen to the response.

Conclusion

In this study, we found that the mere belief you were interacting with another person led to superior learning in the avatar condition, especially when there was novel variability in the response type that also carried over to new inference questions. This belief also led to greater skin conductance changes early on when there was novel variability in the response type. The SCL was correlated with better learning, and the highest correlation occurred during the action of reading, possibly implying that when people take a socially-relevant action, this influences what they learn when they hear a response.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants No. SLC-0354453. and No. 0527377. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

We would like to thank the members of the LIFE Media group, Ben Trombley-Shapiro, Michael Jin, Kathryn Rickertsen, Jessica Yuan, Liz Tricase, Nick Cooper, and Nick Yee for all their help and assistance in this study.

References

Bailenson, J.N., Beall., A.C., Blascovich, J., Loomis, J., & Turk, M. (2005). Transformed Social Interaction, Augmented Gaze, and Social Influence in Immersive Virtual Environments. Human Communication Research, 31, 511-537.

Bailenson, J.N., Blascovich, J.,Beall, A.C., & Loomis, J.M., (2003). Interpersonal distance in immersive virtual environments. Personality and Social Psychology Bulletin, 29. 1-15.

Blakemore, S.-J., Boyer, P., Meltzoff, A.N., Segebarth, C., & Decety, J. (2003). The detection of contingency and animacy in the human brain: An fMRI study. Cerebral Cortex, 13, 837-844.

Blascovich, J., Loomis, J., Beall, A., Swinth, K., Hoyt, C., & Bailenson, J. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry, 13*(2), 103-124.

Hoyt, C., Blascovich, J., & Swinth, K. (2003). Social inhibition in immersive virtual environments. Presence, 12, 183-195.

Reeves, B., Nass, C. (1996). The media equation: how people treat computers, television, and new media like real people and places, NY: Cambridge University Press.

Schroeder, R. (Ed.). (2002). The social life of avatars. London: Springer-Verlag.