プロジェクト番号： **1-8**
研究課題名： **Learning from Avatars and Agents in Virtual Reality Environments: Mere Belief in Social Action Improves Complex Learning**
代表者：徳田英幸
報告書作成者：大喜多優 **Sandra Y. Okita (Teachers College, Columbia University)**
研究者（海外）
Sandra Y. Okita, Teachers College, Columbia University, okita@tc.columbia.edu
Jeremy Bailenson, Stanford University, bailenson@stanford.edu
Daniel L. Schwartz, Stanford University, danls@stanford.edu
研究者（国内）
徳田英幸 慶応義塾大学 政策メディア研究科教授
間博人 慶応義塾大学 政策メディア研究科 博士課程

**Abstract**: Three studies tested the hypothesis that the mere belief in having a social interaction with someone improves learning, more attention and higher arousal. Participants studied a passage on fever mechanisms. They entered a virtual reality (VR) environment and met an embodied agent. The participant either read aloud or silently, scripted questions on the fever passage. In the avatar-aloud and avatar-silent conditions, participants were told that the virtual representation was controlled by a person. The agent condition was told that the virtual representation was a computer program. All interactions within VR were held constant, but the avatar conditions exhibited better learning, more attention, and higher arousal. Further results suggest that this was not due to social belief per se, but rather in the belief of taking a socially relevant action.

## Introduction

Virtual reality (VR) permits novel investigations of what it means to be social, and provides a unique way to examine the effects of social interaction on learning with its well-tuned feedback. For example, it is possible to tell people that they are interacting with an embodied ***agent*** that is controlled by a computer. Alternatively, people can hear they are interacting with an embodied ***avatar*** that is controlled by a person. This research explores what makes an interaction socially alive, and what the implications are for people's learning, attention and arousal.

Research on VR 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; Schroeder, 2002). A different question asks if differences arise when people believe they are interacting with a person or 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). This environment provides a unique way to examine the effects of social interaction on learning. Neurological evidence indicates that attributions of humanness recruit different brain circuitry (Blakemore, Boyer, Meltzoff, Segebarth & Decety, 2003), but the effect of social attributions on learning is unknown, particularly if visual features and interactive opportunities are held constant.

A more theoretical question asks whether one element of what it means to be social is to learn. Social engagement has a number of known benefits for learning, but they are not strictly attributable to socialness per se. These include the opportunity to observe a mature performance, to receive questions and generate explanations, and to engage in the social motivation and institutions that sustain learning interactions. Researchers, for example, have found that babies learn more from face-to-face interaction than videotapes. However, in all instances, the effects of social on learning are readily attributed to the timing and quality of information delivery, which computers can largely mimic. Sociable computer programs can model physical human behaviors, increasingly engage in contingent social dialog, and can sustain engagement for long periods of time. Is there anything special left between social and learning once we equate the information in social and non-social interaction?

## Method

### Study Design and Procedure

The current studies take advantage of the research affordances of virtual reality (VR) to test the hypothesis that the mere belief of social interaction can improve conceptual learning and influence basic physiological responses associated with learning. Figure 1 shows a subject engaged with an embodied graphical character in a head mounted VR display. In the avatar conditions, people thought the character was driven by a person whom they had just met. In the agent conditions, people thought they were interacting with a computer agent. In reality, it was always a computer agent. Each study had roughly the same format. Adult subjects read a passage on the mechanisms that sustain a fever. They then interacted in physical reality with a confederate named “Alyssa” by playing the child’s game, Operation. Afterwards, subjects in the agent condition were told that they and Alyssa were going into separate experiments to interact with a computer agent in VR. The agent subjects were further told to say, “Computer program” before each statement to the agent. In the avatar condition, subjects were told that they and Alyssa were going to meet in VR. The avatar subjects were further told to say, “Alyssa” before each statement to the computer character. Otherwise, the conditions were identical (See Figure 2).


Figure 1. 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).

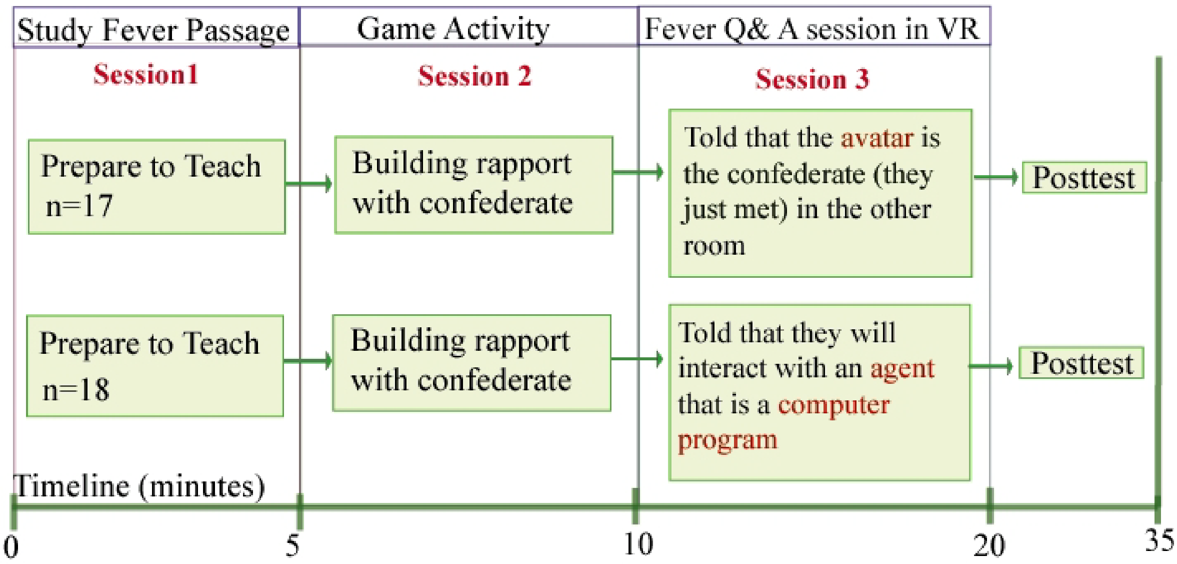



Figure 2. Procedural Flow and Design

Within VR, subjects saw a simulated desk and computer screen, with a computer character on the other side (See Figure 3). They heard that there would be questions that appear on the computer screen, and their task was to read them to the computer-program/Alyssa. There were two warm up questions (e.g., "Computer-Program/Alyssa, can you hear me?" and "Computer-Program/Alyssa, do you know what we are doing?"). Then there were nine questions about fever mechanisms with order randomized across subjects. Before each question, subjects saw a 10second blank blue screen. The scene then reappeared with a new question displayed on the virtual computer screen. Subjects read the question aloud and pressed a button to indicate it was time for the computer/Alyssa to answer, which it did. In both conditions, participants spoke identical words asking questions, and the virtual human provided identical pre-recorded verbal and nonverbal responses. In this way, all interactions and information were held constant across conditions. Therefore, the experiment isolates "social belief" from other important aspects of social interaction for learning.

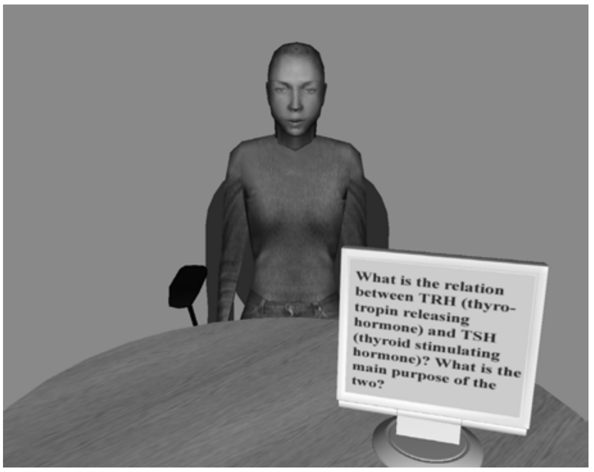



Figure 3. Viewpoint of Participant

The VR character always gave partial answers that were incomplete, but neither wrong nor misleading. The character also had three manners of response created by recording the confederate reading each of the nine answers in a confident, neutral, and doubtful tone. Subjects heard each manner of response for three of the questions, with manner randomly distributed across different questions for each subject.

## Materials and Measures

Figure 1 shows the 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.

There were three dependent measures: learning, arousal, and attention. The learning data were collected in a posttest outside of VR. Subjects answered the nine questions heard in VR (old), plus they received 6 questions they had not heard (new). The 15 questions comprised equal numbers of factual, inference, and application problems. Factual questions could be answered directly from a portion of the passage ("Why do your hands and feet get cold during a fever?"); inference questions required integrating information from across the passage (e.g., "Why is shivering not enough to cause a fever?"); and, application questions required explaining familiar, real world facts about fever (e.g., "Why does a dry nose mean a dog might have a fever?"). Two coders independently scored the subjects' answers: 0 points for a wrong or no answer, 0.5 points for a partial answer (e.g., described one of two mechanisms), and 1 point for a complete answer (See Table 1). The coders had 97% agreement. Using the same coding scheme, the average answer given by the VR character was 0.35.

Table 1: Scoring Method

| Scoring Method (0-2 point scale) | | |
|---|---|---|
| **0:** incorrect/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." | |

Skin conductance measures (SCL) were use to measure arousal. SCL reflect changes in the arousal of the autonomic nervous system. Arousal comprises multiple biological systems, and it is involved in emotion and alertness. Prior research indicates that moderate levels of arousal at encoding correlate with better "factual" memory (Lang, 2000). SCL measures within VR may help reveal whether and when the belief in social or socially relevant action increases arousal and influences learning. We describe arousal and attention after reviewing the learning results.

## Study 1

In study 1, thirty-five college students were randomly assigned to the agent or avatar treatments with the constraint of roughly equal gender. To maximize the visual and auditory cues so the avatar subjects would believe it was Alyssa, the look and voice of the VR character were taken from the confederate in the avatar condition. The subjects in the agent condition interacted with a different confederate, so there was less perceptual similarity with the VR character.

The subjects in the avatar condition learned more, scoring an average .61 and .58 points for old and new questions, respectively. These were greater than the scores of .49 and .49 in the agent condition; $F(1, 33) = 4.14$, $MSE = 0.04$, $p < .05$. The mere belief that the character was a "real" person improved learning, and this learning carried to new problems outside of the virtual world. Also, there were higher arousal measures (physiological sensors that measure skin conductance level-SCL) relative to the agent condition.

Greater arousal correlated with better learning on a problem-by-problem analysis. We found that the peak SCL was the highest when the participant was reading the last portion of a question (See Figure 4). The SCL measures provide some indication of the time course of processing during each Q and A event. Moreover the SCL scores during reading were correlated with learning at the problem-by-problem level. This suggests a possibility that the locus of the learning effect occurs when people took the socially-relevant action of reading, which may have in turn, prepared them to learn more deeply when listening to the response. The SCL data suggest the interesting hypothesis that the learning effect is not due to a general belief that they are listening to a human. Rather, the effect may be the belief that they are taking a socially relevant action, and that the arousal during this action is what prepares them to learn from the response.

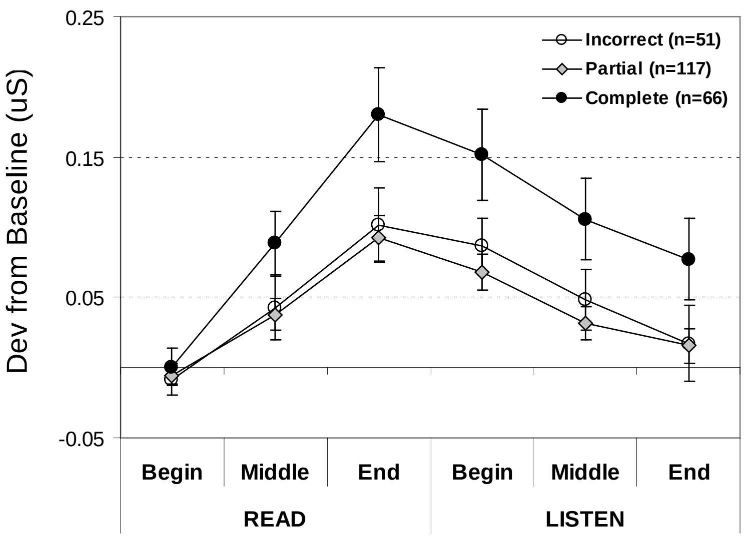



Figure 4: Skin Conductance Back-Sorted by Score on Posttest

## Study 2: Replication Study

To replicate these findings, an identical study implemented a nearly identical design (N=37 after losing one subject). The sole difference was that the same confederate, who matched the look and voice of the VR character, served for both conditions. The results for old and new questions amplified the first study; Avatar > Agent, $F(1,35) = 22.6$, $MSE = 0.04$, $p < .01$. Thus, it is the belief in social interaction, and not only perceptual similarity to the confederate, that drives the learning effect. The results showed successful replication of the avatar and agent condition from the previous study, where the avatar conditions (avatar, avatar-silent) led to superior learning than the agent condition. However, the avatar condition started to show moderate advantage over avatar-silent condition as the problem progressed to harder inferential questions that required the development of a fuller model of temperature regulation. The SCL scores showed a similar trend for the avatar and agent condition as in the first study, but the SCL scores for the new avatar-silent condition was below that of the agent.

Given the close replication, the data from the two studies were consolidated to analyze the more variable arousal and attention data. The arousal data were collected during VR by measuring skin conductance levels (SCL). Increases in hand moisture indicate higher arousal, which is a broad but short-term response of the sympathetic nervous system. For instance, arousal from watching erotica can spill over to affect subsequent but unrelated retaliatory behaviors. At the same time, mild levels of arousal have been correlated with memory for specific experiences. Whether it also correlates with conceptual learning is addressed by the following analyses.

To normalize the arousal data across subjects and questions, each question was partitioned into read and listen phases, and each phase was divided into three equal periods (See Figure 5). The average SCL for each period, minus the baseline SCL from preceding the blue screen, yielded six measures of arousal for each of the nine questions for each subject.

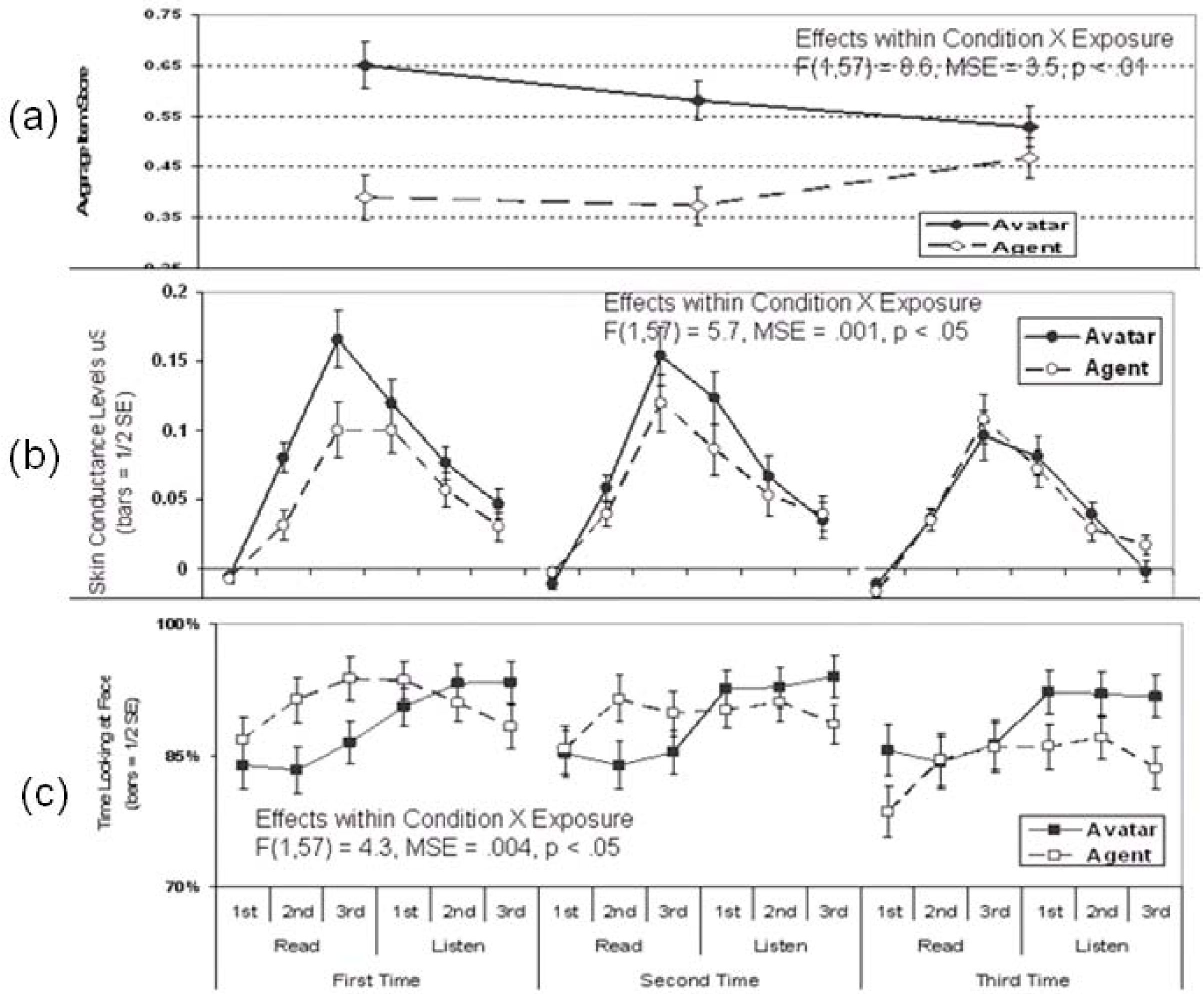



Figure 5. Condition by Exposure (1st, 2nd, 3rd exposure to the same type of manner response: confident, neutral, and doubtful). Figure (a) shows learning: condition x exposure, (b)SCL arousal measure: condition x exposure , (c)attention which is percentage of looking at avatar/agent: condition x exposure

The attention data (Figure 5 (c)) captured whether the subject's head was positioned so the character's face appeared within 10$^o$ of the center of the screen. The percentage of time looking at the character's face was partitioned into the scheme of three periods within the read and listen phases. These data were a proxy for eye position, which has worked well as an index of attention, though people can always pay attention without looking and vice versa.

The posttest learning scores for the old questions were mapped to when the subject heard the question in VR. Figure 5 shows the main effects of the two treatments on learning (a), arousal (b), and attention (c). (Due to equipment and movement artifacts, the figure and following analyses include 29 avatar and 30 agent subjects.) The manner of response (confident, neutral, and doubtful) had no distinguishable effects. However, the number of times a person had been exposed to a given manner was important. Therefore, the figure aggregates results across the manners of response but maintains how many times subjects had been exposed to the manners. Figure 5 (a)(b) indicates that earlier in the VR session, the avatar subjects learned more than the agent subjects and showed higher arousal. The differences between conditions presumably attenuated because the response manners became repetitive, and the belief in social waned. For attention, (Figure 5(c)) the avatar subjects steadily attended to the character's face as it answered, whereas the agent subjects looked less over time.

The correlations of learning with arousal and attention were the same across conditions, despite the overall condition differences. Figure 6 collapses across conditions to show the associations in more detail. Arousal and attention measures were backsorted by the posttest learning scores for each old question. Figure 6 shows the higher levels of arousal and attention associated with better answers.

Within subjects, arousal and attention rose and maintained higher levels on questions for which subjects subsequently scored above their own average; $F(4, 244) = 3.1$, $T = 0.1$, $p < .05$. Arousal and attention were also separable predictors of the learning; respectively, $F(1,57) = 4.2$, $MSE = .014$, $p < .05$, and $F(1,57) = 5.8$, $MSE = .002$, $p < .05$. For instance, when using all 12 arousal and attention measures in a stepwise regression to predict learning for each question, the last arousal and attention measures both enter the equation; $p$'s < .05. In sum, the belief in social increases learning, arousal, and attention; and, arousal and looking are differentially predictive of subsequent learning outcomes.

The relations between learning, arousal, and attention are only correlations; for example, people may have been more aroused or attentive simply because they felt they knew the answer. Nevertheless, given the predominance of information processing models of learning and the fact that content-relevant information was constant, it is useful to posit a causal model explain the effects.

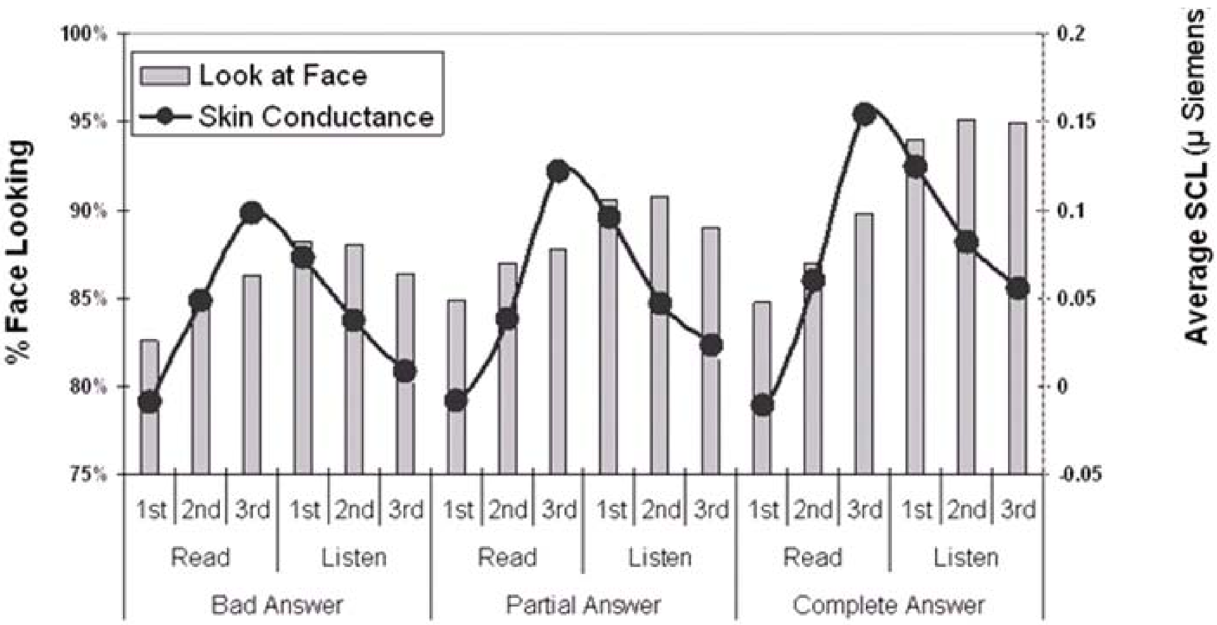



Figure 6. Arousal and attention associated with learning measure (bad, partial and complete answers)

## Study 3

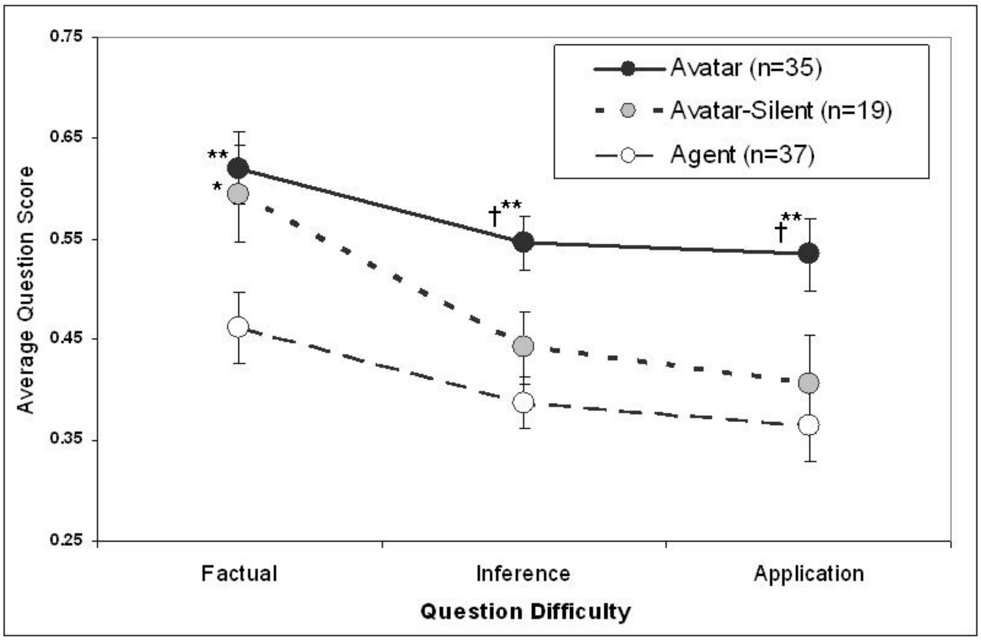
As with the combined SCL data from study 1 and 2, suggest the interesting hypothesis that the learning effect may not be due to a general belief that they are listening to a human. Rather, the effect may be that people believe they are taking socially relevant action, and that the engagement/arousal during this action is what prepares them to learn from the response. This led to a second follow up study where participants read the questions silently rather than aloud to the avatar. 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 may stay low. If so, this might help explain some of the common wisdom that listening is not always as good as interacting. This study explores whether it is the social action, or potential for social action, that prepares one to listen to the response.

To explore this possibility, an avatar-silent treatment changed one feature of the preceding avatar conditions. Nineteen subjects (after losing one) read the questions silently and pressed the button to let Alyssa know it was time for her to answer the question, which she was allegedly reading. The preceding studies indicated that arousal rose during the reading phase. Therefore, if the reading were removed, the arousal might not occur and learning might decline. Equally important, the avatar-silent treatment also tested whether it is the belief in social or the belief in social action that drives the learning. For example, the social benefit may be that people apply a theory of mind to infer what Alyssa is thinking behind her answers. If so, then listening quietly should be as effective as reading aloud in the avatar condition. Alternatively, if the learning is driven by the belief in taking a socially relevant action, then the avatar-silent condition should do the same as the agent condition, because the avatar-silent subjects do not take much socially relevant action either.

Figure 7 compares the scores of the avatar-silent treatment with the previous avatar and agent subjects broken out by question type. The results indicate that sitting quietly listening to a virtual person you think is real works well for learning factual knowledge, but it does not support the integrated learning that can transfer to solve real-world application problems.

Arousal for the avatar-silent condition was flat with no changes relative to the blue-screen baselines. The attention data indicated that avatar-silent subjects looked at the computer character roughly four-fifths as much as the avatar or agent subjects. However, the same relative pattern of attention was found whereby avatar-silent subjects exhibited more looking on those questions that they answered more effectively outside of VR (*ns* difference from other conditions on looking and learning relations).

The belief in social exhibited better learning when avatar-silent subjects paid more attention, but the full learning benefit of social belief depends on social action that, by hypothesis, yields the dual-encoding benefits of arousal. Subjects in the avatar-silent condition exhibited no arousal, and did not develop the integrated understanding of subjects in the avatar condition.




Greater than Agent condition: * p < .05, ** p < .01

Greater than Avatar-Silent; † p < .05

Figure 7. Comparing the scores from avatar-silent treatment with the previous avatar and agent conditions results broken out by question type

## Discussion

In summary, the current research demonstrated that one element of what it means to believe one is in a real social exchange is to learn better. The current experiment isolated the element of social from a relevant action, to explore whether it is the mere belief of being social, or taking a socially relevant action that contributes to learning. As a result, the mere belief of "being social with a human", and taking a "socially relevant action" led to superior learning and deeper understanding.

The studies also demonstrated that action in a social context leads to arousal and attention. The correlations among these measures and with learning led to the hypothesis of a dual pathway that may explain why people can learn better in a social context, even though the information and interactions are the same as a non-social context. This hypothesis needs further testing, as do the effects across a broader class of social and interactive situations.

In the mean time, the results have some implications for the design of virtual worlds. Virtual worlds that gather people from afar to hear a live presentation or lecture will be useful for learning facts compared to a non-social experience. But virtual lectures still suffer the fate of their real world counterparts. People cannot take social action, and they have less opportunity to become aroused. To wit, they become bored, and they do not think as deeply about what is being said. Perhaps, with the help of clever programming, virtual environments can support more belief of social action, even if nobody else in the world can see its consequences. In a school setting, one motivation for lectures over books is that there is a person speaking. Our study suggests that evidently, making them believe it's a person (rather than a computer program) buys you factual knowledge, but not deep understanding unless there is a socially relevant action involved. A simpler explanation may be that people are not aroused when listening to a lecture, since there is no belief in the possibility of social action. Using SCL measures has helped reveal whether belief and action increases arousal/learning, and provide indication of time course in leaning process during social interaction.

## 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., 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.

Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of Communication*, 50(1), 46-70.

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


## Acknowledgments


We would like to thank Byron Reeves, Roy Pea, Anthony Wagner, Girija Mittagunta, and the members of the LIFE Media group, the VRITS lab members 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.