Title
The Mere Belief of Social Interaction Improves Learning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 29(29)

ISSN
1069-7977

Authors
Okita, Sandra Y.
Bailenson, Jeremy
Schwartz, Daniel L.

Publication Date
2007

Peer reviewed

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 University
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 the mere belief in having a social interaction with someone improves learning and understanding. Participants studied a passage on the mechanism that causes a 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 exhibited better learning at posttest, even though all interactions within VR were held constant. Skin conductance indicated that the avatar condition caused more arousal and higher arousal was correlated with learning on a problem-by-problem basis. Further results suggest 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 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. In this research, we examined whether simply believing a virtual representation was an agent (computer) or an avatar (person) affected learning.

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 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). For example, people will respect the virtual "space" of a human representation if they believe it is an avatar.

VR also provides a unique way to examine the effects of social interaction on learning. Social interaction is a natural and powerful way to learn. Social interaction can generate well-tuned feedback, as in the case of a tutor. Moreover, social actors can provide models that learners might imitate. Meltzoff (2007), for example, demonstrated that infants learn to solve a puzzle box if they see a human move the parts, but not if the parts move by themselves. In the current work, we explore whether the mere belief that an interaction is with another person influences 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.

In the study, participants engaged in a scripted Q & A session with a computer agent on the mechanisms that maintain a fever. Participants were told that they were either interacting with an avatar or agent. Afterwards, they took a posttest outside of VR on the mechanisms of fever to see who had learned better. We also collected measures of skin conductance. Skin conductance levels (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 increases arousal and influences learning.

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

Figure 1 provides a schematic of the procedural flow for the comparison of avatar versus agent.

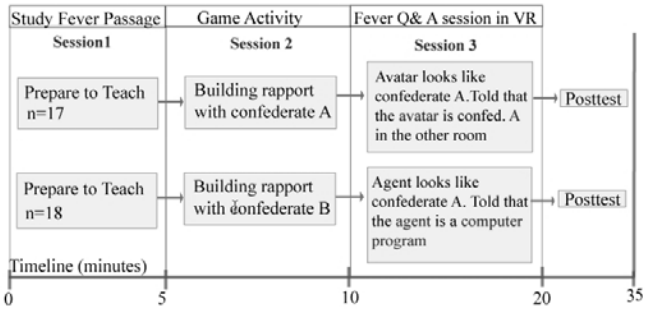

Figure 1: Procedural Flow and Design

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 so they would read more carefully. We did not tell the participants when they would be tutoring, so we could complete the experiment. (They never did tutor anybody.) The passage was removed and they were then introduced to a confederate posing as another participant named "Alyssa." They played the game of Operation with Alyssa for 5 minutes so they could get to know one another (in the physical world). The participants then moved to a separate room (Figure 2). They were told they would be interacting with a computer agent or with Alyssa's avatar. In reality, there was no avatar; a computer program played identical pre-recorded verbal and nonverbal responses in both conditions. To match interactions in the agent and avatar conditions, participants only read fever questions that appeared on the monitor beside the avatar/agent (Figure 3). The participant 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 with the pre-recorded answer. 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 for each of the questions. The question order was randomized across participants.

A within-subject factor manipulated the manner of the pre-recorded response. For each question, we pre-recorded three manners of response, each using the exact same wording, but a different tone of voice. An exhilarated voice was confident and excited. A shameful voice was unsure and apologetic. A neutral voice was flat. Participants heard each manner three times. Response manner was randomly assigned across question and order for each participant. The different response types were intended to increase variability in SCL that we could then correlate with learning (an exhilarated response may induce more arousal). Moreover, changing the manners of response "livened up" the interaction, which we thought was be important for maintaining the social belief in the avatar condition.

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 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 trigger the fever response (e.g., macrophages), the mechanisms that introduce more heat into the body (e.g., shivering), and the mechanisms that prevents heat release (e.g. blocking sweat).

Apparatus

Figure 2 shows a participant wearing a 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 equipment used to measure SCL was a BioGraph Infiniti 3.1 from Thought Technology Ltd.

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

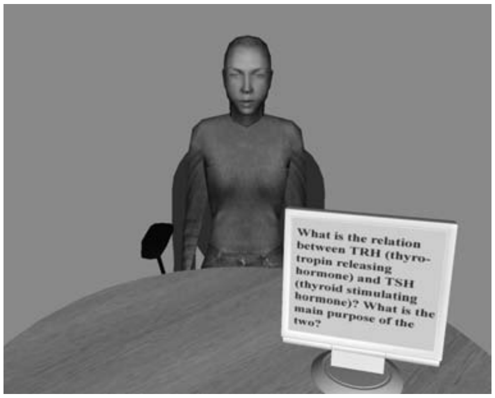

Figure 3: The Viewpoint of a Participant during the Q & A Session in VR

Measures

There were two measures. One was the posttest that measured learning, and the second was skin conductance level (SCL), which indexed arousal. SCL was measured while the participant was in the VR environment interacting with the agent/avatar. The posttest contained the original 9 questions given in VR during the Q & A session and an additional 6 new questions. The questions comprised three categories: Factual, Inferential, and Application. Factual questions could be answered directly from the passage (e.g., “What processes cause the body to increase temperature?”). Inferential questions depended on an integrated model of fever mechanisms (e.g., “Why is shivering not enough to cause a fever?”). Application questions gave participants commonly known facts and asked for an explanation (e.g. “Why does a dry nose mean a dog might have a fever?”). Each question was scored on a 0 to 2 point scale (0: incorrect/no answer, 1: partially correct but incomplete, 2: precise and detailed). Thus, for the posttest, the maximum possible score was 30. Table 1 provides a sample scoring. A secondary coder scored 20% of the transcripts with 97% agreement.

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.”	

Results

Learning Results

The leading question was whether the avatar condition would cause better learning than the agent condition. The posttest given after the VR experience provides the relevant data. 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 advantage occurred for the 9 questions heard in VR as well as for the 6 new questions. We conducted a repeated measures analysis using average score on old and new questions crossed by condition. There was a main effect of condition, $F(1, 33) = 4.14$, MSE = .18, $p < .05$. Old and new questions did not exhibit any differences or interaction with condition, F’s < 0.6. We repartitioned the questions by category, instead of old and new, and found similar results. There was a significant effect of condition, $p < .05$, and no condition by question type interaction, $p > .9$. However, as expected by the design of the questions, there was a main effect of question type; factual questions were the easiest and application questions were the hardest, $F(2, 32) = 9.6$, $p < .01$. Finally, we coded the computer’s responses with the same coding scheme that we applied to the participants’ scores. The average score of the responses we scripted for the computer was 0.7. This is significantly below the average score of 1.1 that the participants achieved for the exact same questions, $t(34) = 7.2$, $p < .001$. Combined, these results indicate that the avatar condition led to superior learning. This learning went beyond simply memorizing the responses of the computer, because participants outperformed the computer on the same questions and they did well on novel problems.

Table 2: Posttest Score

	Avatar M (SD)	**Agent** M (SD)
9 Old Questions from VR	1.2 (0.26)	0.98 (0.35)
6 New Questions	1.2 (0.33)	0.98 (0.36)

.

The second question was whether the different response manners had an effect on learning for the 9 questions heard in VR. Figure 4 shows the average question score for each manner (exhilarate, neutral, shameful) broken out by how many times the participants had been exposed to that manner. A repeated measures analysis crossed manner of response by order (how many times people had heard the manner) by condition. In addition to the condition effect found earlier, there was a condition by order interaction, $F(1,33) = 8.3$, $p < .01$. There was no effect of manner, $p > .6$. The relative novelty of each manner rather than the manner per se influenced learning, and this novelty only affected the avatar condition. The order by condition interaction alleviates concerns about hidden confounds, such as a higher level of prior knowledge in the avatar condition. If the avatar condition had some hidden advantage, then we would expect the participants to do better across the board, not in interaction with order.

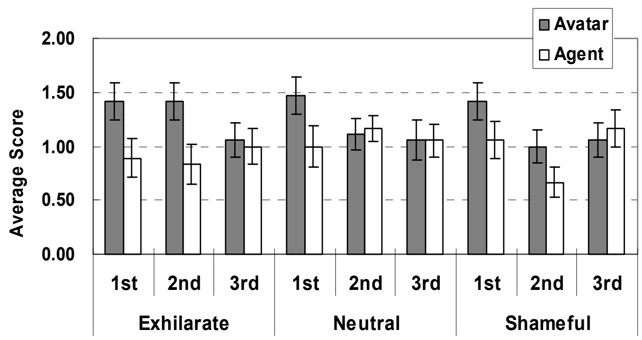

Figure 4: Score by Response Type & Exposure Order (error bars equal SE of the mean)

The effect of order of exposure to a manner 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 systematic 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 people's belief that they were interacting with another person influence their physiological response? If so, was the physiological response correlated with learning outcomes? To answer these questions, SCL was recorded while the participants interacted with the agent/avatar in VR. SCL gauges arousal by passing small electric charges between two points on the body (e.g. one finger to another). When a person's arousal level increases, there is an increase in skin moisture, which increases conductivity across the skin.

Conditioning the Arousal Data

In preparation for analyzing the SCL data, we removed 4 participants from the avatar condition and 5 from the agent condition due to corrupt SCL data. We re-analyzed the learning data for the remaining 26 participants. The results mirrored the results when all 35 participants were included in the sample; the loss of the 9 participants' data does not distort the learning results.

We also normalized the SCL data. The equipment took 60 samples per second. This level of precision is favorable for sudden events. A coarser aggregation is more appropriate for this study, because each question episode spanned tens of seconds. In addition, episodes varied in length; for example, some questions had longer pre-recorded answers than others did. To conduct analyses across questions of different length, we found the mean SCL scores for intervals within an event. We split each question into a reading and listening phase. We further split each phase into thirds; Reading (1^{st}-Beginning, 2^{nd}-Middle, 3^{rd}-End), Listening (1^{st}-Beginning, 2^{nd}-Middle, 3^{rd}-End). We found the mean SCL for each of the thirds.

Different people have different levels of skin moisture. We further normalized the SCL data by subtracting individual baselines. In between each Q & A episode, there was a 10 second blue screen that let SCL levels recover from the preceding event. We computed people's average baseline for the final 3 second of the blank screen period. We subtracted the pre-event baseline from participants' SCL for the subsequent Q & A. Thus, the baseline was updated between each question. The process was repeated for each of the 9 question events in VR.

Effects of Condition on Arousal

The learning analyses indicated that the effect of avatar condition was strongest when the manner of response was relatively novel. We analyzed the arousal data to capitalize on this localization of the learning effect. Figure 5 reflects the structure of the analysis and results.

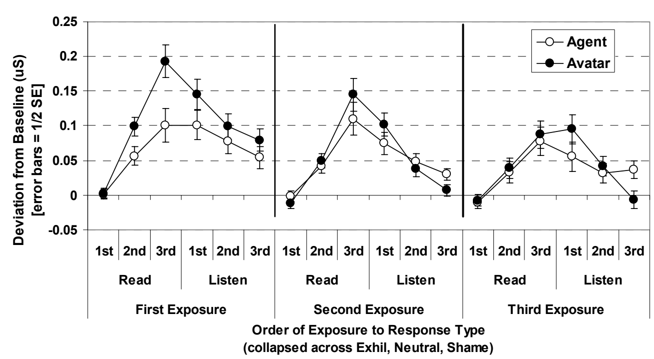

Figure 5: Changes in Skin Conductance Level

A repeated measures analysis included the between-subjects factor of Condition and the within-subject factors of Exposure Order (First time, Second time, Third time) by Read/Listen by Thirds (e.g., 1^{st} third of a read event). The dependent measure was the participants mean SCL aggregated across the three response manners. Figure 5 shows a stable pattern of peaking while reading the question, and then declining while listening to the answer; Read/Listen by Thirds interaction, $F(2, 48) = 13.7$, $p < .01$. There was also a significant interaction with condition. Avatar participants showed a steeper rise in SCL for the early exposures to the computer's response manner than the agent condition; Condition by Exposure Order by Thirds, $F(2.16, 96) = 3.2$, $p < .05$ (applying the Huhn-Feldt correction).

The computer responses appeared after people read the question. Thus, the higher reading arousals for early exposures cannot be attributed to the subsequent response manner. Perhaps the avatar participants were more readily aroused when hearing a variety of response manners, and they became generally less aroused as they began to experience the same types of response repeatedly. An implication may be that it is useful to have a variety of response manners to keep people engaged. This implication, however, does not extend to non-social technologies; arousal levels were unaffected by response types in the agent condition, which parallels the learning results.

Relation between learning and arousal

The avatar participants showed greater learning when the response manners were still relatively novel. They also showed greater arousal when the response manner were still relatively novel. This implies that arousal and learning were correlated. To conduct a closer examination, we switched to problem as the unit of analysis. We back-sorted the arousal measures by posttest score, so we could find the arousal profile for questions where participants scored a 2, 1, or 0.

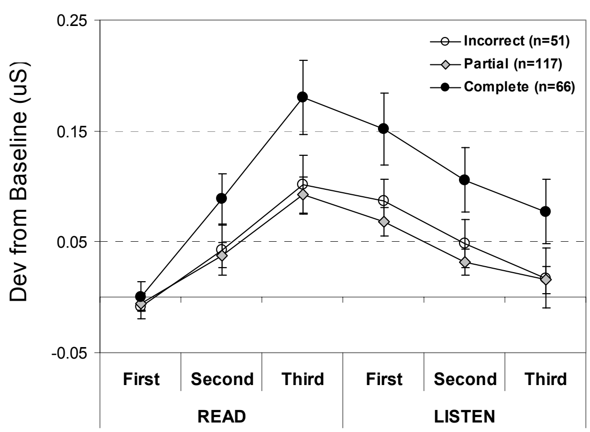

Figure 6: Skin Conductance Back-Sorted by Score on Posttest Measure of Learning

Figure 6 shows SCL means in relation to the posttest scores and the read/listen thirds. The legend indicates the sample size for each arousal profile. For example, out of 234 problems (26 subjects x 9 questions), 66 problems scored a 2 (full correct credit). Higher scores on the posttest had significantly higher SCL levels during the read-listen event. To analyze the relation of SCL to posttest score, we used the question as the unit of analysis. SCL levels for the six periods (Read/Listen by Thirds) were dependent measures in a multivariate analysis. Posttest score and condition were crossed-factors. Results indicated that SCL means for the Read3 and Listen 1-3 were higher for correct answers than for partial and incorrect answers; all F's > 3.0, p's < .05. Partial and incorrect answers did not exhibit different levels of arousal; all p's > .5. There was no interaction of condition by posttest score on the SCL measures; $F(12, 446) = 1.02$, Wilkes Lambda = .95, $p > .4$. Thus, the relation of arousal to learning was the same for both conditions; namely, higher arousal levels were correlated with better learning outcomes. This is a useful finding, because prior research has indicated that arousal is correlated with factual memory (Lang, 2000), but here we found a correlation with learning that went beyond what participants had heard in the VR setting.

The SCL data provide information about the time course of learning in VR. For problems on which participants did well, the highest level of arousal occurred during reading. For these problems, there was also an elevated level of arousal when listening, but it is less than the reading peak. The elevated arousal during listening may be due to events that occurred during listening (e.g., a surprising manner of response). But, they may also be due to a naturally slow descent of arousal levels from the peak at reading; a higher peak at reading causes higher levels during listening. This latter alternative is interesting because it suggests that the arousal period that was most significant for learning occurred <u>before</u> the participants listened to the response. We discuss one implication of this explanation below.

Discussion

This study explored whether the mere belief that one is interacting with another person makes a difference in how much one learns. The experiment was unique in that it held all interactions constant. In both conditions, participants spoke identical words asking questions, and the virtual human provided identical pre-recorded verbal and nonverbal responses. Therefore, the experiment isolates "social belief" from other important aspects of social interaction for learning, such as the tailored responses that often arise in social exchange. Results showed that the "belief" in the avatar condition resulted in a significant learning gain relative to the agent condition. This gain occurred even though we scripted the computer's responses so they were neither highly informative nor incorrect. Avatar participants constructed a model of the mechanisms of fever that went beyond the information provided by the computer; they gave better answers for the questions they heard in VR, and they were able to apply their knowledge to novel questions they had not heard in VR.

When we took a closer look at the order of the different response manners 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. There are multiple possible explanations. One is that when the computer's responses varied in tone and the participants had not yet noticed the repetitions in tone, they were more engaged. However, the effect is not simply due to novelty, because the agent participants experienced identical levels of novelty in tone and it had no effect on them. Rather, it was the interaction of novelty and social belief.

The SCL (arousal) measure showed that the belief in social led to greater arousal, especially on early exposures to a response manner. Greater arousal correlated with better learning on a problem-by-problem analysis. We found that the peak SCL occurred during the final third of the reading a question. Originally, we had thought that different response manners would affect learning differentially, but there was no effect of response manner per se. 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 was indirect. The different manners of response kept people socially engaged for a while, which in turn led to more arousal responses.

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 peak SCL scores that were 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 they are listening to a human. Rather, the effect is that people believe 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 may 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 one is interacting with another person leads to superior learning, especially when there is novel social variability in the tone of response. The learning goes beyond remembering what was said, because avatar participants maintained their advantage for novel questions. The "belief in social" 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 greatest SCL occurred during the action of reading, possibly implying that 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 SLC-0354453, 0527377, and REC-0634044. 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 thank Byron Reeves, Roy Pea, Anthony Wagner 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.

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.

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

Meltzoff, A. N. (2007). Infants' causal learning: Intervention, observation, imitation. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 37-47). Oxford: Oxford University Press.

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.