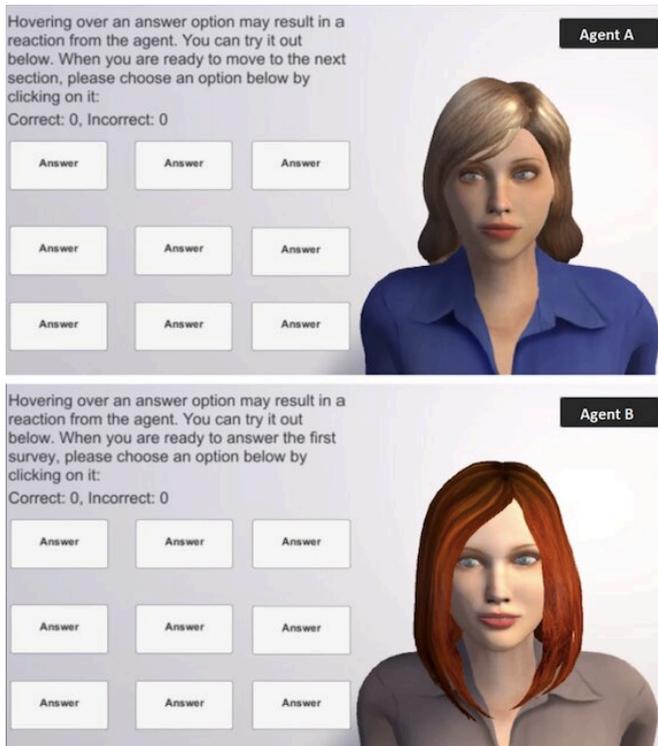


# Examining how humans develop trust towards embodied virtual agents

3 May 2021, by Ingrid Fadelli



Participants familiarize themselves with both agents in the introduction, before beginning the experiment. Credit: Moradinezhad & Solovey.

Embodied virtual agents (EVAs), graphically represented 3D virtual characters that display human-like behavior, could have valuable applications in a variety of settings. For instance, they could be used to help people practice their language skills or could serve as companions for the elderly and people with psychological or behavioral disorders.

Researchers at Drexel University and Worcester Polytechnic Institute have recently carried out a study investigating the impact and importance of trust in interactions between humans and EVAs. Their paper, published in Springer's *International*

*Journal of Social Robotics*, could inform the development of EVAs that are more agreeable and easier for humans to accept.

"Our experiment was conducted in the form of two Q&A sessions with the help of two virtual agents (one agent for each session)," Reza Moradinezhad, one of the researchers who carried out the study, told TechXplore.

In the experiment carried out by Moradinezhad and his supervisor Dr. Erin T. Solovey, a group of participants were presented with two sets of multiple-choice questions, which they were asked to answer in collaboration with an EVA. The researchers used two EVAs, dubbed agent A and agent B, and the participants were assigned a different agent for each set of questions.

The agents used in the experiment behaved differently; one was cooperative and the other uncooperative. However, while some participants interacted with a cooperative agent while answering one set of questions and an uncooperative agent when answering the other, others were assigned a cooperative agent in both conditions or an uncooperative agent in both conditions.

"Before our participants picked an answer, and while their cursor was on each of the answers, the agent showed a specific facial expression ranging from a big smile with nodding their head in agreement to a big frown and shaking their head in disapproval," Moradinezhad explained. "The participants noticed that the highly positive facial expression isn't always an indicator of the correct answer, especially in the 'uncooperative' condition."

The main objective of the study carried out by Moradinezhad and Dr. Solovey was to gain a better understanding of the process through which humans develop trust in EVAs. Past studies suggest that a user's trust in computer systems can

vary based on how much they trust other humans. insight about the experiment."

"For example, trust for computer systems is usually high right at the beginning because they are seen as a tool, and when a tool is out there, you expect it to work the way it's supposed to, but hesitation is higher for trusting a human since there is more uncertainty," Moradinezhad said. "However, if a computer system makes a mistake, the trust for it drops rapidly as it is seen as a defect and is expected to persist. In case of humans, on the other hand, if there already is established trust, a few examples of violations do not significantly damage the trust."

As EVAs share similar characteristics with both humans and conventional computer systems, Moradinezhad and Dr. Solovey wanted to find out how humans developed trust towards them. To do this, they closely observed how their participants' trust in EVAs evolved over time, from before they took part in the experiment to when they completed it.

"This was done using three identical trust surveys, asking the participants to rate both agents (i.e., agent A and B)," Moradinezhad said. "The first, baseline, survey was after the introduction session in which participants saw the interface and both agents and facial expressions but didn't answer any questions. The second one was after they answered the first set of questions in collaboration with one of the agents."

In the second survey, the researchers also asked participants to rate their trust in the second agent, although they had not yet interacted with it. This allowed them to explore whether the participants' interaction with the first agent had affected their trust in the second agent, before they interacted with it.

"Similarly, in the third trust survey (which was after the second set, working with the second agent), we included the first agent as well, to see whether the participants' interaction with the second agent changed their opinion about the first one," Moradinezhad said. "We also had a more open-ended interview with the participants at the end of the experiment to give them a chance to share their



Moradinezhad (left) preparing to do a task on the computer while Dr. Solovey (right) is adjusting the fNIRS sensors on his forehead. The sensor data is read and stored by the fNIRS computer (in the background) for further analysis. Credit: Moradinezhad & Solovey.

Overall, the researchers found that participants performed better in sets of questions they answered with cooperative agents and expressed greater trust in these agents. They also observed interesting patterns in how the trust of participants shifted when they interacted with a cooperative agent first, followed by an uncooperative agent.

"In the 'cooperative-uncooperative' condition, the first agent was cooperative, meaning it helped the participants 80% of the time," Morandinezhad said. "Right after the first session, the participants took a survey about the trustworthiness of the agents and their ratings for the first agent were considerably low, even at times comparable to ratings other participants gave the uncooperative agent. This is in line with the results of other studies that say humans have high expectations from automation and even 80% cooperativeness can be perceived as untrustworthy."

While participants rated cooperative agents poorly after they collaborated with them in the first Q&A session, their perception of these agents seemed to shift if they worked with an uncooperative agent in

the second session. In other words, experiencing agents that exhibited both cooperative and uncooperative behavior seemed to elicit greater appreciation for cooperative agents.

"In the open-ended interview, we found that participants expected agents to help them all the time and when for some questions the agents' help led to the wrong answer, they thought they could not trust the agent," Morandinezhad explained. "However, after working with the second agent and realizing that an agent can be way worse than the first agent, they, as one of the participants put it, 'much preferred' to work with the first agent. This shows that trust is relative, and that it is crucial to educate users about the capabilities and shortcomings of these agents. Otherwise, they might end up completely ignoring the agent and performing the task themselves (as did one of our participants who performed significantly worse than the rest of the group)."

Another interesting pattern observed by the researchers was that when participants interacted with a cooperative agent in both Q&A sessions, their ratings for the first agent were significantly higher than those for the second. This finding could in part be explained by a psychological process known as 'primacy bias.'

"Primacy bias is a cognitive bias to recall and favor items introduced earliest in a series," Morandinezhad said. "Another possible explanation for our observations could be that as on average, participants had a lower performance on the second set of questions, they might have assumed that the agent was doing a worse job in assisting them. This is an indicator that similar agents, even with the exact same performance rate, can be seen differently in terms of trustworthiness under certain conditions (e.g., based on their order of appearance or the difficulty of the task at hand)."

Overall, the findings suggest that a human user's trust in EVAs is relative and can change based on a variety of factors. Therefore, roboticists should not assume that users can accurately estimate an agent's level of reliability.

"In light of our findings, we feel that it is important to

communicate the limitations of an agent to users to give them an indication of how much they can be trusted," Morandinezhad said. "In addition, our study proves that it is possible to calibrate users' [trust](#) for one agent through their interaction with another agent."

In the future, the findings collected by Morandinezhad and Dr. Solovey could inform practices in social robotics and pave the way toward the development of virtual agents that human users perceive as more reliable. The researchers are now conducting new studies exploring other aspects of interactions between humans and EVAs.

"We are building machine learning algorithms that can predict whether a user will choose an answer suggested by an agent for any given question," Morandinezhad said. "Ideally, we would like to develop an algorithm that can predict this in real-time. That would be the first step toward adaptive, emotionally aware intelligent agents that can learn from user' past behaviors, accurately predict their next behavior and calibrate their own behavior based on the user."

In their previous studies, the researchers showed that a participant's level of attention can be measured using functional near-infrared spectroscopy (fNIRS), a non-invasive brain-computer interface (BCI). Other teams also developed agents that can give feedback based on brain activity measured by fNIRS. In their future work, Morandinezhad and Dr. Solovey plan to further examine the potential of fNIRS techniques for enhancing interactions with virtual agents.

"Integrating brain data to the current system not only provides additional information about the user to improve the accuracy of the machine learning model, but also helps the agent to detect changes in users' level of attention and engagement and adjust its behavior based on that," Morandinezhad said. "An EVA that helps users in critical decision making would thus be able to adjust the extent of its suggestions and assistance based on the user's mental state. For example, it would come up with fewer suggestions with longer delays between each of them when it detects the user is in normal state,

but it would increase the number of and frequency of suggestions if it detects the user is stressed or tired."

**More information:** Investigating trust in interaction with inconsistent embodied virtual agents. *International Journal of Social Robotics*(2021). DOI: [10.1007/s12369-021-00747-z](https://doi.org/10.1007/s12369-021-00747-z)

© 2021 Science X Network

APA citation: Examining how humans develop trust towards embodied virtual agents (2021, May 3) retrieved 13 May 2021 from <https://techxplore.com/news/2021-05-humans-embodied-virtual-agents.html>

*This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.*