Enhancing User Experience in Virtual Reality with Fallible Agents that Express Uncertainty

Hailey Graham

Abstract

Since the recent advancements of artificial intelligence, particularly with Large Language Models (LLMs), the notion of virtual reality (VR) and intelligent agents assisting with user learning and enhancing their interactive experiences has been of great interest. However, there is still substantial dislike of intelligent agents from the public. This research intends to engage users in a positive way with intelligent agents and investigate what attributes, in the agents, evoke a positive emotional response, increase user motivation and sense of presence. This study aims to investigate the impact that embodied intelligent agents who express uncertainty have on user experience in VR and whether users can identify this uncertainty. The research question that guided this investigation was:

How do embodied agents that express uncertainty impact users' experience in VR?

The study was exploratory. Ten participants completed surveys before and after the VR experience to measure their emotional response, intrinsic motivation and sense of presence. Then, they interacted with the virtual environment and the two embodied agents (one that appeared uncertain, and the other was sure about the information they provided). After, participants also engaged in an interview about their experience with the researcher. From the initial interviews, participants often preferred one agent over the other (and not necessarily the agent that expressed uncertainty). This preference came from the agent appearing assertive, friendly, human-like or approachable. It is hypothesised that users will have a more positive emotional response to agents, the more human-like they appear. However, users will be more motivated to engage with an activity when agents appear sure of themselves and give clear instructions. This study shows potential for intelligent agents to appear approachable and enhance users' sense of presence in VR.

Introduction

Rapid advancements of Large Language Models (LLMs) such as ChatGPT have been of great interest. Industries that use interactive environments have begun to use intelligent agents and virtual reality to enhance users' experience. Intelligent agents can be used as the artificial intelligence (AI) behind non-playable characters (NPCs) that inhabit a gaming environment. The idea is that they are more human-like and provide a more varied experience with each time a person engages with an interactive environment (games, simulations and otherwise). Particularly, when virtual environments and in-person experiences are not an option, the importance of VR and intelligent agents provide an enjoyable and engaging experience. However, there is still a substantial dislike of AI. Studies such as Demir & Guraksin (2022) suggest that the less human and more infallible agents appear, the more likely users are to view them negatively. This study aims to investigate the impact of embodied agents expressing uncertainty on user experience in VR. It is hypothesised that fallible agents appear more human and approachable, evoking a positive response and increasing user motivation to engage with the environment. Moreover, we

wanted to see if users could identify that the agents were expressing uncertainty. The following research question guided this study with three sub-questions:

1. How do **embodied agents** that express **uncertainty** impact users' experience in a **VR** setting?
    a. How is **user motivation** towards learning financial literacy in VR affected by embodied agents that express uncertainty?
    b. How is **user motivation** to engage with embodied agents affected by the agent expressing uncertainty?
    c. How do embodied agents that express uncertainty impact **users' emotional response** towards the agent?

Background

Embodied Agents

An embodied agent is an intelligent agent that has a physical form that can move around and interact with the virtual environment that it inhabits (Jiafei et al., 2022). In this study, the environment is in virtual reality.

Virtual Reality

Virtual reality (VR) is a simulated world using position tracking and 3D near-eye displays so that users feel like they are inside a virtual environment (Goode, 2019).

ChatGPT

ChatGPT is an open source chatbot developed by OpenAI, released on 30 November 2022. It is based on an LLM and allows users to have specific and detailed conversations about a wide range of topics (OpenAI, 2022). In the virtual world, two embodied agents (Isabelle – agent 1 – and Freya – agent 2) are backed by ChatGPT to enable them to have improvised conversations with the user.

Literature Review

In Cassel's chapter (2022) on socially interactive agents, she reviews the impacts of intelligent agents as peers to students in their learning. She informs the reader about virtual peers compared to human ones and the difference between asymmetric relationships and symmetric relationships in learning. Cassel's research focuses on early childhood learning. Cassel comments on the importance of peer learning for problem-solving skills, and how human peer-to-peer interactions are not always possible, meaning that virtual peers have a place in student learning. Cassel's research reveals when virtual peers are effective or ineffective. The main limitation of the chapter is that it only considers learning in early childhood; however, these conclusions that Cassel has made can be investigated to higher education through this study. Cassel highlights that there is a place for virtual peers in education and that artificial intelligence can in some ways be more effective when it is not seen as an all-knowing, powerful being.

Demir & Guraksin (2022) investigate the perception of artificial intelligence from middle school students and how this perception can be changed through education. The authors aimed to determine middle school students' views towards AI and categorise these opinions through metaphoric analysis. The paper focused on middle school students as compared to higher education due to the gap in the literature and participants were sampled from the central districts of Afyonkarahisar and Izmir. This article showcases students' baseline perception of AI and suggests that this perception can be changed depending on educational methods. The main limitation, though, is that the participants are from a small region and the article only investigates middle school students. Overall, the study concluded that students display both positive and negative views (but mostly positive) of AI which can be categorised to the following topics: smart, brain, nature, security, humanistic, the dilemma of good and evil, service, object, technology, life, and time.

In Mayer et al.'s article (2023), the authors consider the positive and negative effect immersive virtual reality (IVR) has on one's learning. Mayer et al. use research gained through their previous experiments to collate and summary data surrounding the topic and determine whether and IVR can be an effective tool to enhance learning. Their research focuses on immersive virtual reality compared to traditional mediums of learning. While this study developed a virtual environment for the experiment, the main focus is fallible intelligent agents. The article brings into question what could change about IVR to make learning more effective. They conclude that getting students to summarise, answer and enact are effect and can potentially be enhanced with uncertain agents that need help from students. The main limitation of the paper is that it is not a formal meta-analysis and as such the authors only cite their own work. The authors concluded that IVR can be more motivating and engaging, but it is easy for it to become distracting and in fact have a negative effect on learning. This article provides interesting points on how IVR can be improved with uncertain intelligent agents and less distractions.

In Prada & Rato's chapter (2022), they review and discuss roles and dimensionality of socially interactive agents (SIA) in games. The authors use research and data collected from various research papers within the last 15 years to try to identify the roles of SIA in games, the development of SIA throughout history, the current dilemmas surrounding SIAs and gaps in the literature. The paper concludes that SIA help create and sustain social dimensions of the game and improve social immersion for players.

In Schneider et al.'s article (2022), they review the paradigms and theories surrounding people's learning through digital materials and social interactions. The authors through the analysis of literature reviews, meta-analyses and social theories and paradigms aimed to provide evidence that social processes affect individual learning when using digital materials. The article considers a multitude of definitions of social cues and theories but worked primarily under the assumptions of the Computers-As-Social-Actors (CASA) Paradigm and CASTLE. Schneider et al. suggest that the closer learning with digital material is to a social interaction and the more anthropomorphic the digital media is, the more effective a person's learning is. The main limitation of the article is that it had no experimental or observation section to support its research finding and was not being a formal meta-

analysis. The authors summarise already established theories and comment on theories that require further investigation.

Methods

Participants

The study consisted of ten participants. There were 6 male, 3 female and 1 non-binary participants. The majority of the participants were between the ages of 25 and 34 (5); three were 18 to 24 years old; one was 45 to 54 and one was 55 to 64. All participants' vision was normal except for one where their vision impairment would not affect their experience during testing. All participants' first language was English. There was a wide range in video game experience; however, only one participant has significant experience with virtual reality games.

Self-Assessment Manikin Questionnaire

Prior to the virtual reality (VR) experience, participants completed a survey that collected demographics and their current emotional state through the self-assessment manikin (SAM) questionnaire. The SAM test consists of three questions where the participant selects an image that best represents their current emotional state for valence, arousal and dominance.

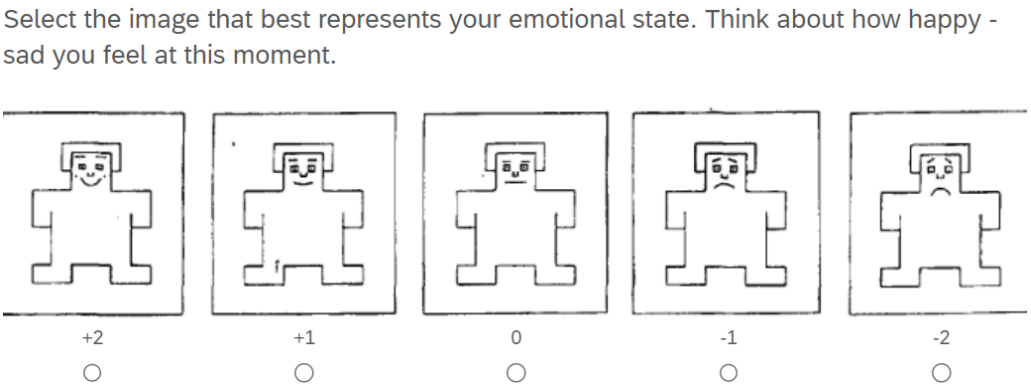

Figure 1: Valence question from SAM test

Participants completed the SAM test again after the VR experience along with other questionnaires.

Virtual Reality Experience and Embodied Agents

Participants then played a VR game, using the Meta Quest 2, for five minutes, after a user tutorial. The game was made using Unity 3D Engine. The VR experience consisted of a budgeting game and two embodied intelligent agents. These agents were called Isabelle and Freya. Isabelle expressed uncertainty about budgeting whereas Freya was certain about her budgeting knowledge and was eager to inform the player. Participants could talk to Isabelle and Freya using the Quest 2's microphone and the agents would log the conversation and respond back. After the five minutes were completed, the tester asked the participants if they wanted to play for another two minutes and recorded the participants responses.

Intrinsic Motivation Inventory, Sense of Relatedness and Sense of Presence

After the VR experience, participants completed another survey which measured their motivation, sense of relatedness and sense of presence. Motivation was measured using the Intrinsic Motivation Inventory which is a questionnaire containing statements referring to motivation to do the activity (VR game). Participants would indicate how true each statement was on a scale from 1 (not true at all) to 7 (Very true). Sense of relatedness measured how close participants felt to the agents and sense of presence measured how real the VR experience and the agents themselves felt to the participants.

Post-test Interview

The final part of the study consisted of a semi-structured interview. The tester asked the following questions:

1. Tell me about your experience with the VR app today.
    a. How did it go? What worked? What did you like? What did not work for you?
2. Why or why not did you want to continue playing the game?
3. What differences did you notice between Isabelle and Freya?
4. Which agent did you prefer?
5. Why did you prefer that agent?
6. In what ways were the agents helpful?
7. In what ways were the agents unhelpful?

Due to the small sample size, the interviews were used to collect qualitative data surrounding each participants experience to corroborate with the survey data and supply possible explanations for the results.

Findings

Survey Results

The data from the surveys was analysed using Jamovi. Paired samples t-tests were performed on the SAM test variables, sense of relatedness between Isabelle and Freya, and sense of presence between Isabelle and Freya. As seen in figure 2, valence increased after the VR game (p=0.434) and arousal increased afterwards as well (p=0.104). However, neither changed significantly. So, in general, participants were slightly happier and more excited after the experience.

Paired Samples T-Test

			statistic	df	p		Effect Size
Valence_Pre	Valence_Post	Student's t	–0.818	9.00	0.434	Cohen's d	–0.259
Arousal_Pre	Arousal_Post	Student's t	–1.809	9.00	0.104	Cohen's d	–0.572
Dominance_Pre	Dominance_Post	Student's t	0.000	9.00	1.000	Cohen's d	0.000

Note. H_a $\mu_{\text{Measure 1 - Measure 2}} \neq 0$

Figure 2: Paired samples t-test of participant SAM results before and after VR experience

From figure 3, participants sense of relatedness to Freya was higher than Isabelles (p=0.205) but not significantly. So, participants were more connected to Freya and felt they had a relationship with her. However, from the descriptives below the mean sense of relatedness for both Freya (3.73) and Isabelle (3.23) are below average (4).

Paired Samples T-Test

			statistic	df	p		Effect Size
SOR_Freya	SOR_Isabelle	Student's t	1.37	9.00	0.205	Cohen's d	0.432

Note. H_a $\mu_{Measure\ 1 - Measure\ 2} \neq 0$

Descriptives

	N	Mean	Median	SD	SE
SOR_Freya	10	3.73	3.67	1.22	0.384
SOR_Isabelle	10	3.23	3.17	1.24	0.391

Figure 3: Summary of participants' sense of relatedness to Isabelle compared to Freya

From the Intrinsic Motivation Inventory (figure 4), the mean motivation to partake in the activity – the VR experience – was 4.91 which was above average (4) and the median was 5.43. For participants' motivation to do the budgeting task, the mean was 3.97 and the median was 4.50. From figure 5, participants' motivation to engage in the VR experience was significantly higher than average (p=0.021) and their motivation to engage with the budgeting task was not significantly below average.

Descriptives

	N	Mean	Median	SD
IMI_Average	10	4.91	5.43	1.22
IMI_B_Average	10	3.97	4.50	1.95

Figure 4: Intrinsic Motivation Inventory questionnaire summary

One Sample T-Test

		Statistic	df	p		Effect Size
IMI_Average	Student's t	2.3771	9.00	0.021	Cohen's d	0.7517
IMI_B_Average	Student's t	-0.0540	9.00	0.521	Cohen's d	-0.0171

Note. H_a $\mu > 4$

One Sample T-Test

		Statistic	df	p		Effect Size
IMI_B_Average	Student's t	-0.0540	9.00	0.479	Cohen's d	-0.0171

Note. H_a $\mu < 4$

Figure 5: One Sample T-Test to determine whether participants' motivation differed from the average (4)

As for sense of presence (figure 6), there was a very slight preference for Isabelle (Agent 1) over Freya (p=0.693). This meant that more participants thought that Isabelle seemed more real than Freya, but it was not significant.

Paired Samples T-Test

			statistic	df	p		Effect Size
iGroup_Agents1	iGroup_Agents2	Student's t	0.408	9.00	0.693	Cohen's d	0.129

Note. H_a $\mu_{\text{Measure 1 - Measure 2}} \neq 0$

Descriptives

	N	Mean	Median	SD
iGroup_Agents1	10	4.30	4.50	0.823
iGroup_Agents2	10	4.10	4.50	1.287
iGroup_Average	10	4.39	4.42	0.713

Figure 6: Summary of sense of presence towards the VR experience and the agents

VR Experience and Interviews

From the interviews, many participants expressed that they were intrigued by the novelty of the experience. Yet, there were some issues with the controls that confused them. Specifically, there were issues with the microphone picking up their voices and displaying the microphone user interface. This made some participants talk to an agent more than the other because the microphone seemed to work better with a specific agent.

Five participants expressed that they preferred Isabelle and the other five preferred Freya. Common descriptors of Isabelle when she was preferred included "friendlier", "personable", "more human-like", "curious" and "helpful" whereas when she was not preferred common descriptors included "dismissive", "knew nothing" and "robotic." When Freya was preferred descriptors included "friendlier", "engaging", "helpful", "back and forth dialogue", "spoke clearly." However, when not preferred she was described as "more formal", "condescending", "more robotic", "unreal" and "about the numbers." Finally, seven out of the ten participants recognised that Isabelle was uncertain about budgeting while Freya was certain.

Discussion

Discussion of Results

The most significant outcome measure was the general intrinsic motivation inventory ($p=0.021$). This result suggests that participants were highly motivated to engage in the activity (the VR experience with the intelligent agents). However, the questionnaire does not indicate what the reason is for this motivation. From the interviews, participants suggested that they were motivated by the novelty of the VR experience and the intelligent agents – finding out what the agents knew and the limits of the interactions with them. Conversely, motivation to partake in the budgeting task was low. The interviews suggested the participants general lack of interest in the subject and the agents did not increase their desire to do the budgeting task.

From the findings, the agent that expressed certainty (Freya) was preferred. However, this difference was not statistically significant. An agent was preferred if they appeared friendlier, more approachable, or more human to a person. What a participant considered to be friendly or approachable depended on person to person. Therefore, agents expressing uncertainty do not necessarily evoke a positive emotional response. People's perception of artificial intelligence may depend more on personality or subjective factors.

Strengths and Limitations

This study provides a basis for further investigation on how embodied agents can evoke positive responses from users. The results suggest that uncertainty may be one factor as some participants appreciated the interpreted curiosity and humility of the uncertain agent. However, others found it dismissive and lacking self-confidence. Future research should investigate combined factors (such as personality type with fallible agents) that evoke positive emotional responses to intelligent agents.

Some of the results were skewed by the novelty of the VR experience for many participants. The findings from the SAM test suggests that participants' arousal increased somewhat. However, from the responses in the interviews this is due to the novelty of the VR experience. Furthermore, although 9 out of 10 participants wanted to play in the virtual world for 2 extra minutes, the interviews revealed that this was because participants wanted to get a better understanding of the game controls and virtual world rather than being motivated to engage in the budgeting activity or interact with the intelligent agents.

If this study was to be done again, it would be conducted as a within subjects study to separate interactions with Freya from Isabelle since many participants interpreted Isabelle telling them to talk to Freya about budgeting as telling them to go away. Notably, the system evoked some confusion for participants due to the lack of experience with VR and lack of instruction prior to testing.

Conclusion

The preliminary findings from this study suggest that embodied intelligent agents do impact users' emotional responses in VR. However, the type of response varies from person to

person. Reiterating this study with more structured VR experience and considering participants' personality type may further the literature on intelligent agents' interactions with users.

References

Cassel, J. (2022). Socially Interactive Agents as Peers. *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics, 2*, 331-365. doi:https://doi.org/10.1145/3563659.3563670

Demir, K., & Guraksin, G. E. (2022). Determining middle school students' perceptions of the concept of artificial intelligence: A metaphor analysis. *Participatory Educational Research, 9*(2), 297-312. doi:http://dx.doi.org/10.17275/per.22.41.9.2

Goode, L. (2019). *Get Ready to Hear a Lot More about XR.* Wired. https://www.wired.com/story/what-is-xr/

The jamovi project (2022). jamovi. (Version 2.3) [Computer Software]. https://www.jamovi.org.

Jiafei, D., Samson, Y., Hui Li, T., Hongyuan, Z., Cheston, T. (2022). A survey of embodied ai: From simulators to research tasks. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *6*(2), 230–244

Mayer, R. E., Makransky, G., & Parong, J. (2023). The Promise and Pitfalls of Learning in Immersive Virtual Reality. *International Journal of Human-Computer Interaction, 39*(11), 2229-2238. doi:https://doi.org/10.1080/10447318.2022.2108563

OpenAI. (2022). *ChatGPT.* https://chat.openai.com.

Prada, R., & Rato, D. (2022). Socially Interactive Agents in Games. *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics, 2*, 493–526. doi:https://doi.org/10.1145/3563659.3563675

R Core Team (2021). R: A Language and environment for statistical computing. (Version 4.1) [Computer software]. https://cran.r-project.org.

TED-Ed. (2021, March 12). *How does artificial intelligence learn? - Briana Brownell* [Video]. YouTube. https://www.youtube.com/watch?v=0yCJMt9Mx9c

TED-Ed. (2022, December). *How will AI change the world?* [Video]. TED. https://www.ted.com/talks/ted_ed_how_will_ai_change_the_world?language=en

Schneider, S., Beege, M., Nebel, S., Schnaubert, L., & Rey, G. D. (2022). The Cognitive-Affective-Social Theory of Learning in digital Environments (CASTLE). *Educational Psychology Review, 34*, 1-38. doi:https://doi.org/10.1007/s10648-021-09626-5

This Week in Enterprise Tech. (2023, June 3). *The ChatGPT Balancing Act - AI Challenges, human fallibility, maintaining data agility with Talend* [Video]. YouTube. https://www.youtube.com/watch?v=K3C_vi_Gl_Y.