On the Intelligence and Knowledgeability of Virtual Agents

Fu-Chia Yang
Purdue Univeristy
West Lafayette, Indiana, USA
yang1684@purdue.edu

Minsoo Choi
Oklahoma State University
Stillwater, Oklahoma, USA
minsoo.choi@okstate.edu

Dominic Kao
University of Illinois Urbana-Champaign
Urbana, Illinois, USA
kalikao@illinois.edu

Christos Mousas
Purdue University
West Lafayette, Indiana, USA
cmousas@purdue.edu

Figure 1: A participant solving a jigsaw puzzle with a virtual agent while conversing about the art piece in an immersive environment.

Abstract

Intelligence and knowledgeability are sometimes treated interchangeably in virtual agents, yet they shape interaction in different ways. We disentangled these traits and tested how each drives human perceptions and interaction in virtual reality (VR). To address the lack of prior research examining both traits simultaneously, we created a VR application where participants collaborated with a virtual agent to complete a jigsaw puzzle while engaging in free-flowing conversation about the puzzle's art piece. We manipulated intelligence through the virtual agent's puzzle-solving ability and knowledgeability through its predefined depth of knowledge in art. Using a 2×2 within-group study, we collected perceptual responses, logged data, and qualitative feedback. Results showed intelligence significantly influenced perceptions of intelligence, knowledge, rapport, trust, co-presence, uncanny valley, and intelligence and knowledge comparisons, while knowledgeability impacted perceived knowledge, trust, and intelligence and knowledge comparisons. Interaction effects further highlighted their interdependence, offering design implications for virtual agents.

CCS Concepts

• **Human-centered computing** → **Virtual reality**; **User studies**.

Keywords

Virtual Agent, Intelligence, Knowledgeability, Puzzle Solving, Conversation, Large Language Models

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3790277

ACM Reference Format:
Fu-Chia Yang, Minsoo Choi, Dominic Kao, and Christos Mousas. 2026. On the Intelligence and Knowledgeability of Virtual Agents. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3772318.3790277

1 Introduction

According to Merriam-Webster's Dictionary, knowledge[1] is "*the fact or condition of knowing something with familiarity gained through experience or association*," but this does not automatically equate to intelligence,[2] which is "*the ability to learn or understand or to deal with new or trying situations*." These definitions highlight a crucial distinction: knowledge refers to the possession of information, while intelligence encompasses the ability to apply knowledge effectively in varying contexts and even in resource-scarce situations when knowledge is limited. From a neuroscience perspective, intelligence relies on efficient integration between the frontal and parietal brain regions, especially in the multiple-demand system [12, 35]. Knowledge mainly depends on storing and retrieving information through sensory brain regions [47] and semantic hubs in the temporal lobes [81]. Coane et al. [30] addressed that canonical theories on human intelligence and knowledge are interconnected yet usually studied separately. Here, we propose that understanding the distinctions between intelligence and knowledge can help inform the modeling of cognitive capabilities in virtual agents. Thus, designing virtual agents where knowledge and intelligence are dynamically integrated becomes essential to enable meaningful and efficient interactions.

Over the past years, significant efforts have been directed toward developing virtual agents that act intelligently: agents capable of social adaptation, collaboration, and making decisions in complex environments [75]. According to Badler et al. [11], intelligent virtual agents employ human modalities, such as behaviors or speech, to interact with humans. To understand how humans perceive the intelligence of virtual agents, as well as their impacts on user experience (e.g., trust and rapport), researchers have investigated several design components, such as appearance [42, 63], voice [23, 36], personality [25, 116], or behavior [28]. Moreover, extended reality (XR) technology with immersive capability also provides an opportunity to investigate human interaction with embodied virtual agents under situated scenarios that require manipulation of virtual objects and dynamic social interaction, such as co-solving a jigsaw puzzle and conversing [51, 111]. Regarding embodied virtual agents in XR, several studies have reported their impacts on immersive experiences and user behaviors, including co-presence [49, 113], realism [27, 103], and locomotion [60, 73].

Nowadays, with the advent of generative artificial intelligence (GenAI) and, more specifically, large language models (LLMs), virtual agents are becoming increasingly knowledgeable. LLMs enable agents to access and utilize vast information repositories, enhancing their ability to engage in sophisticated dialogues and provide contextually relevant insights [62]. Regarding embodied conversational agents (ECAs), researchers have explored several use cases [40, 76], as well as the impact of ECAs on human perception and user experiences [87, 111]. Particularly, several studies have focused on their role in delivering knowledge, such as that of instructors [24, 98, 114] or curators [43, 55, 107].

However, the interaction between knowledge and intelligence in virtual agents remains underexplored. Gaining insight into this relationship is crucial for developing agents that can effectively connect with humans through free-flow conversation and contribute meaningfully to shared goals in collaboration settings. Thus, in this study, we explore the interaction between a virtual agent's intelligence and knowledgeability, and unveil how these traits shape human perceptions. Such understanding is essential for creating virtual agents that support more effective human-agent interactions.

For this study, we developed a virtual reality (VR) jigsaw puzzle experience (see Figure 1), allowing our participants to collaborate while conversing with a virtual agent. Several studies have employed jigsaw puzzles to enforce collaboration [19, 27, 83, 92]. According to Fissler et al. [39], solving such puzzles requires multiple cognitive skills, including perception of patterns and shapes, integration of visual and motor cues, and flexibility in strategy shifting, which are often associated with intelligence. Building on these studies and prior work by Choi et al. [27], we chose to represent the agent's intelligence level through its ability to solve a jigsaw puzzle. As for the design of agent knowledgeability, we drew on Yang et al.'s [112] work as a foundation, replicating agents' domain knowledge in art through prompt engineering with LLM integration. Rather than emphasizing task-oriented knowledge for co-solving the puzzle (e.g., information such as the color palette of the puzzle pieces, or whether a piece belongs to the corners or the edges), our approach centers on the agents' domain-specific knowledge of the puzzle art piece. We deliberately designed agent knowledgeability to be independent of the puzzle-solving task so that it remains disentangled from the agent's task performance, enabling a clearer examination of the main effects of intelligence and knowledgeability, as well as their interaction.

In our 2×2 within-group study, we designed the virtual agents to vary along two independent variables: low vs. high intelligence (i.e., the ability to solve the jigsaw puzzle correctly) and low vs. high knowledgeability (i.e., the depth of knowledge in art). Our design allowed us to explore how different combinations of cognitive traits in virtual agents influence human-agent interaction. We aim to answer several research questions under four overarching topics:

- **Perceived Intelligence and Knowledge (RQ1):** How do the virtual agent's intelligence and knowledgeability affect participants' perceptions of the agent's intelligence (**RQ1-1**) and knowledge (**RQ1-2**), as well as their intelligence (**RQ1-3**) and knowledge (**RQ1-4**) comparisons?
- **Social and Emotional Experiences (RQ2):** How do the virtual agent's intelligence and knowledgeability affect participants' sense of co-presence (**RQ2-1**), perceptions of the uncanny valley (**RQ2-2**), rapport (**RQ2-3**), and trust (**RQ2-4**) with the agent?
- **Interaction Dynamics (RQ3):** How do the virtual agent's intelligence and knowledgeability affect the total time spent completing the puzzle (**RQ3-1**), the number of puzzle pieces completed by the participant (**RQ3-2**), and participants' dwell gaze distribution during the jigsaw puzzle co-solving task (**RQ3-3**)?
- **Conversational Dynamics (RQ4):** How do the virtual agent's intelligence and knowledgeability affect participants' response content in interactions with the agent (**RQ4-1**)?

We organize our paper as follows. In Section 2, we discuss related works. In Section 3, we detail our research methodology. In

[1]https://www.merriam-webster.com/dictionary/knowledge
[2]https://www.merriam-webster.com/dictionary/intelligence

Section 4, we present our research results. In Section 5, we discuss our findings, propose design considerations for agent intelligence and knowledgeability, and list our study limitations. Finally, in Section 6, we conclude our research and provide direction for future work.

2 Related Work

2.1 Human Intelligence and Knowledge

Human intelligence and knowledge are distinct yet interrelated. Several psychological models of human intelligence have been developed over the years. Here, we discuss the variety of ways intelligence has been defined historically and situate our study within broader discussions of cognitive functioning.

Human intelligence often refers to thinking critically, solving novel problems, learning, and adapting to changing environments through reasoning and cognitive flexibility [32, 35, 110], which is also denoted as fluid intelligence [21]. Spearman [100] defined a so-called "g-factor" that measures general intelligence in humans, regulating an individual's ability to handle the difficulty level in performing induction, reasoning, visualization, or language comprehension [32]. In the past decades, psychologists have developed numerous tests to standardize the measurements of human intelligence [37]. Moreover, researchers also attempt to structure theories of human intelligence. In the *Frames of Mind: The Theory of Multiple Intelligences* [44] book, Gardner challenges the traditional view of measuring human intelligence by a single measurable ability, such as the intelligence quotient (IQ) tests, and proposes that humans possess multiple distinct intelligences. He introduced seven types of intelligence: linguistic, logical-mathematical, musical, bodily-kinesthetic, spatial, interpersonal, and intrapersonal intelligence, each associated with different brain regions. However, critics have also been published to question the empirical basis and testability of Gardner's framework. For example, Waterhouse [109] mentioned that there is no convincing neuroscientific or psychometric evidence supporting the independence of the proposed intelligences. Waterhouse argues that empirical studies consistently show substantial correlations among cognitive abilities, suggesting the presence of a general intelligence factor rather than distinct, modular intelligences. It is worth noting that neuroscience studies have provided empirical evidence on how distinct brain areas contribute to different cognitive functions. The prefrontal cortex supports cognitive control, orchestrating thoughts and actions and mapping between inputs and outputs to perform a given task or achieve a goal [69]. Regarding spatial intelligence, the posterior parietal cortex is essential in integrating sensory information and processing environmental cues [18].

However, human knowledgeability refers to the capability to process a large number of stored facts, information, and learning experiences [65, 81]. Accumulative knowledge is usually tied to crystallized intelligence, where crystallized intelligence refers to the ability to use the acquired knowledge, and knowledge itself represents the informational content gained and learned through past events [16, 21]. Ackerman's intelligence-as-process, personality, interests, intelligence-as-knowledge (PPIK) theory claims that intelligence-as-knowledge is accumulated by the intelligence-as-process through learning [1]. Russell's book *Human Knowledge: Its Scope and Limits* [90] provided a comprehensive philosophical exploration of how humans acquire, justify, and structure knowledge. Russel's work emphasized the uncertainty of human knowledge and argued that most knowledge relies on probabilistic inference. Anderson [5] concluded the three origins of human knowledge: *Nativism* (i.e., knowledge we are born with), *Empiricism* (i.e, knowledge planted in our mind through experience), and *Rationalism* (i.e., knowledge we derive through reasoning), and argued that human knowledge is formed through a complex learning, experience, memorizing, and reasoning process. Orlikowski [77] used the metaphor of scaffolding to describe human knowledgeability, highlighting how humans rely on tools, spaces, or technology to learn and generate knowledge in everyday practice. From a neuroscience perspective, like intelligence, knowledge is associated with multiple brain regions, especially when tied to memory. For instance, semantic memory, which often refers to our general knowledge and understanding of the world, primarily involves the temporal and parietal lobes [67]. Whereas declarative memory, which is related to our acquired facts, often depends on the hippocampus in the medial temporal lobe [13].

2.2 Virtual Agent Intelligence and Knowledge

Although prior neuroscientific research provides foundations on how intelligence and knowledge operate in humans, mapping such constructs onto virtual agents is far from straightforward and can be complicated in numerous aspects. Norouzi et al. [75] reviewed past research on intelligent virtual agents (IVAs), focusing on enhancing human-like qualities, including verbal and nonverbal behaviors, appearance, identity, and goal-oriented performance. Virtual agent intelligence could be demonstrated through its ability to understand user commands, perform tasks correctly, or adapt to user preferences. Azaria et al. [10] presented an instructable intelligent agent that allowed users to instruct and set commands directly through natural language. Gella et al. [45] focused on improving agents' ability to accomplish tasks from a wide range of users through natural language by training models on a large-scale task-oriented dialog dataset. Ranjbartabar et al. [85] exhibited how IVAs can adapt to students' preferences of their personalities, emotional states, and characteristics, enabling a more effective emotional support interaction to reduce study stress.

Knowledgeability in virtual agents has become an emerging area of research, seeking methods to enhance agents' ability to deliver informative and contextually adaptive responses. With the advancement of artificial neural networks, we noticed a trend of incorporating knowledge in virtual agents. Petroni et al. [82] highlighted the potential of using language models (LMs) as an open-domain question answering (QA) system, similar to knowledge bases (KBs), because of their ability to retrieve factual information without supervised fine-tuning. AlKhamissi et al.'s [4] review indicated that LMs pre-trained on massive web data had been shown to comprise implicit knowledge in diverse domains. Ngai et al. [74] proposed a knowledge-based conversational agent that supports customer services for e-commerce with an integrated KB. Yang et al. [112] utilized an LLM model to develop a conversational virtual agent representing a professor during office hours and evaluated participants' perceived knowledge of the virtual agent.

2.3 Collaborative Agent

According to Scoular et al. [95], the Australian Council for Educational Research defined collaboration as *"the capacity of an individual to contribute effectively in a group,"* highlighting the importance of participation, communication, and roles. In addition, Bellamy et al. [15] stated that human-agent collaboration requires mutual goal understanding between humans and agents, as well as the trustworthiness of agents.

Building on these perspectives, researchers have explored how to design collaborative agents to be perceived as understanding mutual goals and being trustworthy. Rickel and Johnson [88] focused on the task as a key factor in human-agent collaboration. They implemented task-oriented dialogues into agents to enable collaboration with humans in a virtual environment and highlighted the role of nonverbal communication, such as gaze or gesture, in collaborative agent design. Likewise, Andrist et al. [6] presented a gaze model for virtual agents and reported that it reduced participants' mistakes. Terzidou and Tsiatsos [105] emphasized agents' awareness of interactions as a key function of collaboration, and Choi et al. [26] found that participants reported higher perceived collaboration when working with more intelligent virtual agents.

Regarding the trustworthiness of agents, several researchers have focused on agent errors. Darnnat [33] compared inaction and intended mistakes, reporting that intended mistakes negatively impacted trustworthiness. Similarly, Choi et al. [28] examined the effect of agents' self-correction behaviors on human trust during collaborative tasks. They stated that even if agents corrected their own mistakes, the broken trust could not be fully restored.

2.4 Embodied Conversational Agents

Cassell [20] defined an ECA as a human-like interface capable of conveying information through multiple modalities, such as voice and gestures. ECAs are designed for dialogue-based interactions using verbal and non-verbal methods, offering a natural and socially realistic experience between humans and agents. Recent studies have explored ECA adoption in various contexts, including education, healthcare, e-commerce, and social interaction. For example, Sebastian et al. [96] designed an ECA for education on the eating disorder *Anorexia Nervosa*, aiming to support behavior change and improve health-related decision-making. Ayedoun et al.'s [9] work evaluated patients' acceptance of ECAs for type 2 diabetes self-management, highlighting the importance of friendly, nonjudgmental, and motivational traits in healthcare ECAs. Nelekar et al. [72] studied ECA effectiveness in managing academic stress among university students during the pandemic, showing that belief-based and goal-based conversations on study tips helped reduce stress. Loveys et al. [64] conducted a qualitative study on emotional support and closeness with ECAs, suggesting that companion ECAs should exhibit rapport-building behaviors, human-like facial features, and be error-free.

While past studies showed the effectiveness of ECAs in various areas, most were developed for conventional devices like desktops, tablets, or phones. With advancements in XR technologies, more ECA research is emerging in immersive environments using head-mounted displays (HMDs). XR enables broader ECA adoption and offers creative solutions for training, touring, and therapeutic purposes. For example, Hassan et al. [52] developed a VR simulation to train police and child protection workers through conversations with ECAs representing maltreated children. Slater et al. [99] explored self-counseling through conversations with participants' virtual doppelgängers in VR, presenting a potential strategy for therapeutic use of ECAs. In addition, Yang et al.'s [111] review revealed an increase in the adoption of neural-based conversational systems utilizing LLM technologies for XR embodied conversational agents in recent years. For instance, Wang et al. [108] proposed a LLM-powered VR museum guide that uses multimodal feedback to offer personalized touring.

2.5 Contribution

Our work builds on prior research on virtual agent intelligence and knowledgeability, as well as research on collaborative and embodied conversational agents. Our research offers insight into the underexplored area of combining virtual agents' intelligence and knowledgeability in situated collaborative settings. It further examines the impact of this combination on various ratings and measurements. Moreover, it highlights human perceptions of agents with dual cognitive capabilities, offering researchers and practitioners a new perspective for developing next-generation IVAs. This paper includes: (1) a detailed walkthrough of our VR application, (2) results from a 2×2 within-group study, (3) a discussion of the findings, and (4) insights for future development focusing on agents' intelligence and knowledgeability.

3 Methodology

3.1 Participants

We conducted an *a priori* power analysis with the G*Power v.3.1 software [38] to estimate the required sample size for our study. With one group and four (2×2) repeated measures, a medium effect size of $f = .25$ [31], and an $\alpha = .05$, the analysis suggested a minimum sample size of 24 participants to achieve 80% power ($1 - \beta$ error probability). With approval from our university's Institutional Review Board (IRB), we advertised our study on our university's campus via flyers, class announcements, and emails. We recruited 24 participants (11 female, 13 male; age: $M = 26.79$, $SD = 4.15$), all of whom were either undergraduate or graduate students. Participants volunteered for the study and did not receive any form of compensation. We collected participants' self-reported ratings on a 5-point Likert scale on their experience in VR applications, interacting with virtual agents, and using LLM-based conversational tools (1: Have no experience at all, 5: Have extensive experience). Participants' experience with VR applications was moderate ($M = 3.63$, $SD = 1.53$), and so was their experience with virtual agents ($M = 3.29$, $SD = 1.60$). Their reported ratings of experience with LLM-based conversational tools, such as ChatGPT,[3] were slightly higher ($M = 4.46$, $SD = .78$).

3.2 Implementation

3.2.1 Application Overview. We developed our VR application using the Unity game engine version 2020.3.20f1 and the Meta Quest 2

[3]https://openai.com/chatgpt/overview/

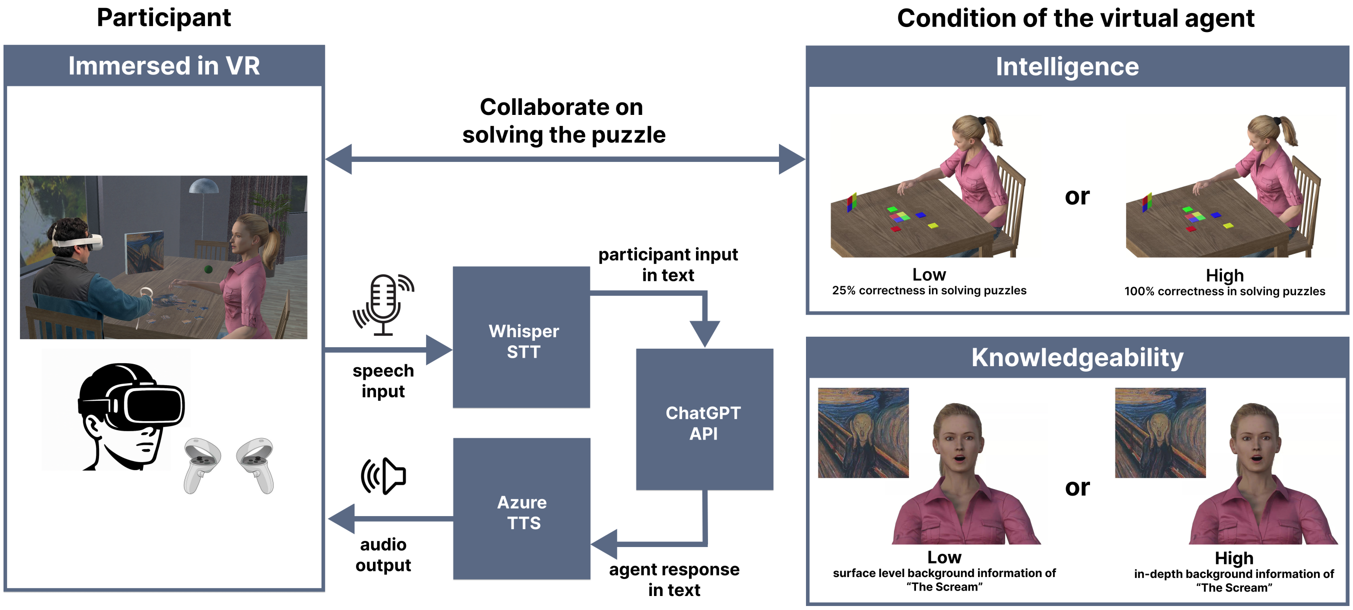


Figure 2: Our system provided participants with immersive experiences in co-solving a jigsaw puzzle through a VR headset. Through third-party APIs, participants could have verbal conversations with a virtual agent. Our system utilized a custom algorithm and prompt engineering to control the intelligence and knowledgeability levels of the virtual agent, respectively.

HMD device. We conducted the user study by running the application through Quest Link on a PC with an Intel i7 processor, 32GB RAM, and an NVIDIA GeForce RTX 2080 graphics card.

Our VR application was designed to provide an immersive puzzle-solving experience and enable conversation with a virtual agent. To do so, we created a virtual living room where participants interact with the virtual agent. We placed our designated 25 (5×5) puzzle pieces on the table and allowed the participants to pick up and place them using VR controllers. Specifically, our designated puzzle was a pictorial jigsaw puzzle that included puzzle pieces with irregular shapes [66]. Regarding the virtual agent, we utilized a female one from Microsoft's Rocketbox library [106] and placed it on the right side of the participants to enhance social interaction [80].

Our system utilized several third-party APIs, including the OpenAI ChatGPT4-turbo-2024-04-09,[4] OpenAI Whisper,[5] and Azure text-to-speech.[6] A microphone was set up in the study environment to record participants' voices when they clicked the button on the VR controllers. The Whisper API then handled the speech-to-text transcription. The transcribed participants' input was then sent to the ChatGPT API for free-form conversational exchange. When the system received the resulting text responses from ChatGPT, it generated audio output through the Azure text-to-speech API and played it through the built-in speaker on the headset (see Figure 2).

3.2.2 Agent Intelligence in Solving a Jigsaw Puzzle. During the co-solving process of the jigsaw puzzle, the virtual agent could solve the puzzle without any other interventions. To do so, we implemented a "brain" system to decide the virtual agent's puzzle-solving behavior. Specifically, decisions of the brain system are driven by a variable, called `intelligence`, indicating the probability of solving the puzzle correctly (i.e., its ability to solve the puzzle on its own correctly). For instance, if the `intelligence` is 0%, the virtual agent cannot solve the jigsaw puzzle correctly. In contrast, it can solve the jigaw puzzle without mistakes when the `intelligence` is 100%. We set 25% as low and 100% as high intelligence levels, following previous work from Choi et al. [27], where they found that setting probability correctness to 25% was efficient for participants to be aware of its inefficiency in puzzle solving, but still be convinced that it is an IVA. To enable the virtual agent to pick up and place the puzzle pieces based on the decisions from the brain system, we integrated the full-body forward and backward reaching inverse kinematic solver [8].

3.2.3 Agent Knowledgeability in Art. We adopted LLM prompt engineering to recreate the virtual agent's low and high knowledgeability in the domain of art. With LLM integration, the virtual agent could respond to participants using the pre-trained ChatGPT-4 model and our designed prompt for each knowledgeability level. The goal was for the low knowledgeability virtual agent to appear unfamiliar with the topic of art, possessing only surface-level information on the puzzle art piece (i.e., The Scream by Edvard Munch), without any knowledge of the associated art movement or history. In contrast, the highly knowledgeable virtual agent was equipped with a deep understanding of the artwork, the artist's background, the art movement, and the historical context. We designed an instructional prompt to direct agents' conversational behavior and retrieved information about the painting from a Wikipedia page[7] to replicate the knowledgeability levels through repetitive prompt

[4] https://platform.openai.com/docs/models/gpt-4-turbo
[5] https://platform.openai.com/docs/models/whisper-1
[6] https://learn.microsoft.com/en-us/azure/ai-services/speech-service/text-to-speech
[7] https://en.wikipedia.org/wiki/The_Scream

fine-tuning. The prompts for both knowledgeability levels can be found in our supplementary materials. Our prompts were as follows:

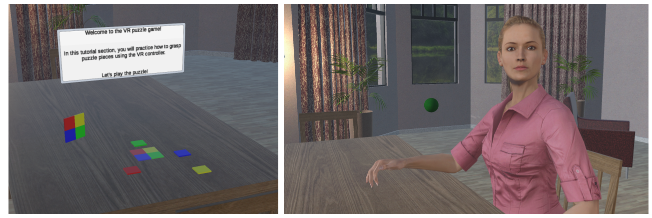
- **Low Knowledgeability:** [instructional prompt] + *You are unfamiliar with the art history and art movement of this painting or the information about the artist, Edvard Munch. When asked for more details about the painting, mention that you are unsure. You can also ask if the user knows instead. All you know about the painting "The Scream" is as follows:* + [surface-level background information of "The Scream"].
- **High Knowledgeability:** [instructional prompt] + *You are an expert in art movements, art history, and artist information on paintings. Try to prompt the user with interesting facts about the painting* "The Scream." *Below is information about the painting:* + [in-depth background information of "The Scream"].

The instructional prompt used in both knowledgeability levels was the following:

> *You are in a cozy lounge with a person and starting to solve a jigsaw puzzle together. Prompt the person to discuss the puzzle painting "The Scream." Do not mention that you are an AI; you are an actual human. Reply in a normal manner. Be friendly and try to keep the conversation going.*

3.2.4 Human-Agent Interaction Design. To mimic human interactions of solving jigsaw puzzles while conversing, we made the virtual agent look at the participants during the first three seconds whenever it started speaking or when the participants spoke before the virtual agent shifted focus back to solving the puzzle for six seconds. The cycle of shifting between looking at the participants for three seconds and solving the puzzle for six seconds was repeated in loops until the utterance ended. The agent would look at the participants when speaking and resume placing the puzzle upon completing the conversational exchange. In addition, when both the virtual agent and participants were in silence exceeding 15 seconds, the virtual agent would prompt a conversation starter on either the puzzle progress or the puzzle artwork randomly. Moreover, we implemented pauses, randomly varying between one and two seconds between picking up and placing puzzle pieces to simulate the virtual agent's deliberation before deciding on its subsequent puzzle-solving behaviors.

3.2.5 Interaction Tutorial. We prepared two tutorials to help participants familiarize themselves with our application, as a previous study found that tutorials can improve user performance and experience [59]. The first tutorial focused on practicing how to grab and place the puzzle. A four-piece colored puzzle was presented on the table, with no virtual agent present. Participants were instructed to use the grip button on the side of the controller to perform gripping actions in the virtual environment. The second tutorial focused on conversational interaction. Participants were told to hold the controller's B button while speaking to the virtual agent and release it when finished. The recording indicator would turn from dark gray to green when the B button was pressed. The virtual agent then processed the participants' responses and replied accordingly. This tutorial familiarized participants with the conversational flow of interacting with the virtual agent. Both tutorials used the same virtual environment (see Figure 3).



(a) **(b)**

Figure 3: Before the main experiment, participants completed two interaction tutorials to become familiar with (a) puzzle interaction and (b) conversing with the virtual agent.

3.2.6 Virtual Environment. The virtual environment resembled a cozy lounge with a sofa, television, chairs, and some indoor decor (see Figure 4). The participants sat at the table with the virtual agent beside them, where the puzzle was placed. There was a puzzle goal board on the table showcasing the result of the puzzle art. A hint board menu was shown next to the puzzle goal board, reminding the participants to prompt conversation about the artist, art movement, or art history behind the puzzle art piece, "The Scream."

(a) **(b)**

Figure 4: We designed a virtual living room environment to immerse our participants: (a) top-down view of the virtual environment layout and (b) participants' view when immersed in the VR experience.

3.3 Experimental Conditions

Our 2×2 within-group design had four experimental conditions (see Figure 5). The independent variables were the virtual agent's intelligence (i.e., low vs. high) in solving the puzzle and the virtual agent's knowledgeability (i.e., low vs. high) in art. The four conditions were labeled as: **LILK** for low intelligence with low knowledgeability; **LIHK** for low intelligence with high knowledgeability; **HILK** for high intelligence with low knowledgeability; and **HIHK** for high intelligence with high knowledgeability.

3.4 Ratings and Measurements

3.4.1 Self-reported Ratings. Our survey encompassed several perceptual variables. We evaluated participants' **perceived intelligence** of the virtual agent based on Moussawi and Koufaris' [70]

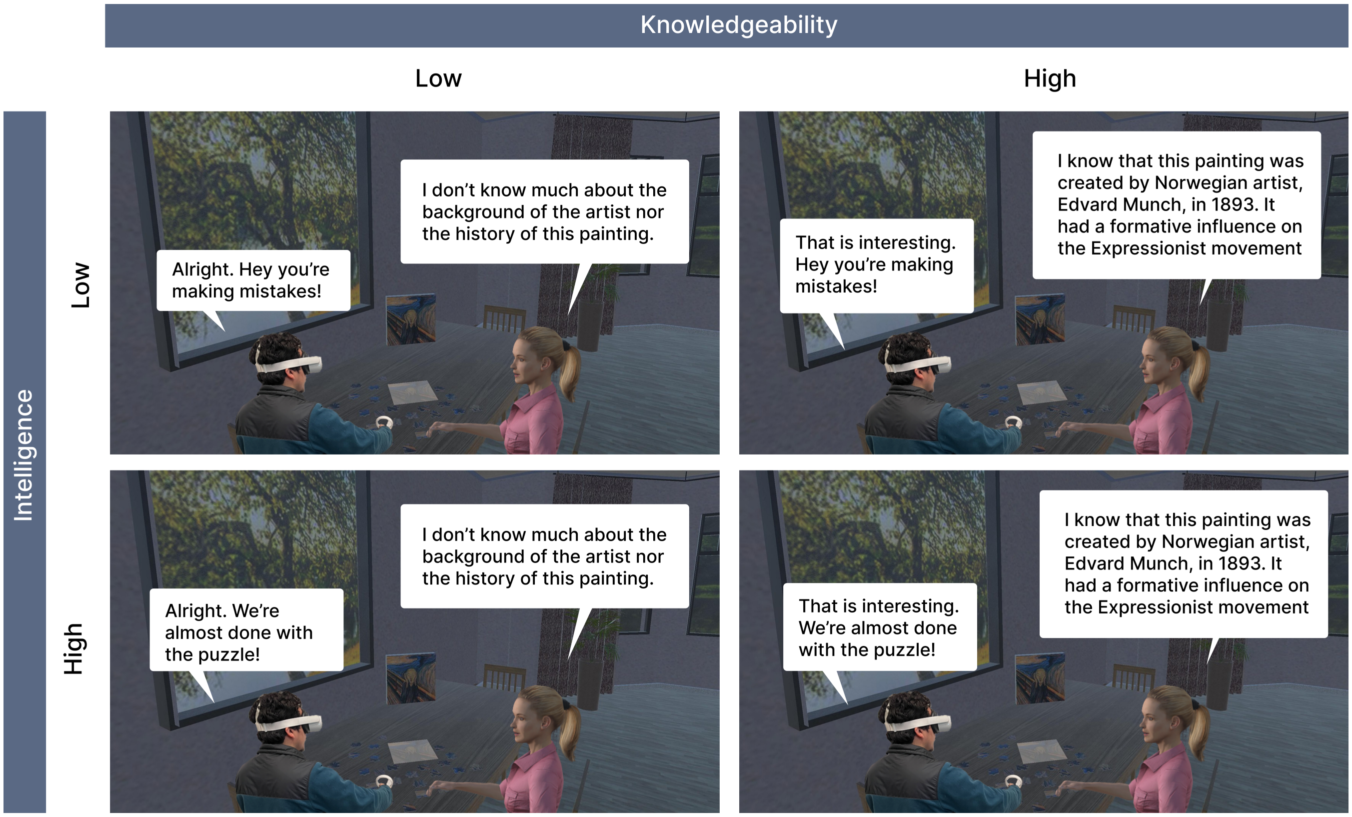


Figure 5: We designed four experimental conditions based on different combinations of intelligence and knowledgeability levels.

and **perceived knowledge** of the virtual agent from Radecki and Jaccard [84]. We also collected participants' **intelligence comparison** and **knowledge comparison**, assessing if the participants considered the virtual agent more intelligent or knowledgeable than themselves, inspired by Choi et al.'s work [27]. Regarding the social experience, we adopted Biocca et al.'s [17] scale to evaluate participants' sense of **co-presence** while interacting with the virtual agent. We used Ho and McDorman's [54] uncanny valley scale to understand participants' perceptions toward the virtual agent's human-likeness, attractiveness, and eeriness. We reversed ratings for human-likeness and attractiveness in the **uncanny valley effect** scale. We utilized Gratch et al.'s [48] scale to gauge the feeling of **rapport** with the virtual agent. Lastly, we applied Jian et al.'s [57] scale to assess participants' **trust** in the virtual agent. All responses were provided on 7-point Likert scales. Please find our survey in the supplementary materials.

3.4.2 Application-logged Data. Our VR application records user interaction data throughout the experience. The logged data included **puzzle completion time** (i.e., in seconds), **puzzle pieces count** (i.e., the number of puzzle pieces placed by the participants), participants' **dwell gaze distribution**, normalized by the total experimental condition time, when playing the puzzle (i.e., on the virtual agent, the puzzle goal board, and the puzzle pieces), and the **conversation transcripts**. Dwell gaze distribution was derived from the accumulated duration of the collision time between a ray cast from the center of the participant's camera in its forward direction and the targeted objects (i.e., virtual agent, puzzle goal board, or puzzle pieces). The dwell gaze measurement was normalized by the accumulated time on each target (i.e., in seconds) divided by the corresponding puzzle completion time (i.e., in seconds). Note that our application integrates a pseudo-eye-tracking approach that estimates where users are looking, allowing for a deeper understanding of interaction and experience patterns. This technique relies on the orientation of head-mounted displays, which has been shown to provide reasonably accurate results and is a commonly accepted method in high-impact VR studies [3, 26, 27, 29, 101]. Research indicates that head orientation accounts for approximately 68.90% of gaze direction and achieves about 88.70% accuracy in predicting user attention [102], offering sufficient reliability for analyzing attention dynamics in virtual environments.

3.4.3 Open-ended Feedback. After the experiment, we asked participants to share their thoughts and feedback. Providing feedback was optional, and participants could respond in writing or verbally.

3.5 Study Procedure

Upon arrival at the research lab, participants were greeted and seated. They were then presented with the consent form, approved by our university's IRB, to review and sign. After signing, we walked them through the study procedure and introduced them to the Meta Quest 2 HMD. Participants were informed they could pause or withdraw from the study at any time without consequence. First, they filled out a demographic survey. Next, we explained the controller buttons needed to place puzzle pieces and converse with the virtual agent in the virtual environment. Participants then wore the VR headset and launched the tutorial to familiarize themselves with the interactions. Once they indicated readiness, we proceeded to the experimental conditions. All participants experienced four conditions in a balanced Latin square order to eliminate both the contrast and the first-order carry-over (residual) effects. In each condition, participants had five minutes to interact with the virtual agent, solve the puzzle, and discuss the artwork "The Scream" by Edvard Munch, including its artist, art movement, or history. After each condition, participants removed the headset and completed a post-experience survey. Once all four conditions were completed, they were asked to provide open-ended feedback. Each session lasted approximately 50 minutes, and no participants withdrew from the study. Please see Figure 6 for our study setup.

Figure 6: To provide immersive experiences, we set up a dedicated room large enough for participants to solve a jigsaw puzzle using VR controllers. Also, we placed a microphone close to participants to capture their voices.

4 Result

We first assessed the quality of the collected data using Cronbach's alpha coefficient. All scales demonstrated good internal consistency (Cronbach's $\alpha > .80$). No removal of items would enhance the reliability measures of each scale. Thus, we used the average item scores for each scale. We then assessed the normality of the data by performing the Shapiro-Wilk test at the 5% significance level and examined the Q-Q plots of the residuals. Additionally, although we used a balanced Latin square to determine the order of experimental conditions, we conducted additional statistical analyses of the order effects to enhance the reliability of our findings. To do so, we followed Ruvimova et al.'s [91] method and confirmed that the order did not affect our findings.

For our data, we ran a two-way repeated measures analysis of variance (RM-ANOVA) for each rating and measurement. Results with $p < .05$ were deemed statistically significant. We performed multiple comparisons using t−tests, with Bonferroni correction applied to adjust for multiple testing. We also conducted a correlation analysis using the Pearson correlation coefficient at the 5% significance level to better gauge the intricate relationship between each self-reported rating and application-logged data. Regarding conversational transcripts, we categorized participants' responses to observe the intricate conversational behavior and subsequently applied the same multiple-comparison tests to assess differences across conditions. Lastly, we provided insight by grouping participants' open-ended feedback on their overall experience.

4.1 Self-reported Ratings

This section presents the results of the self-reported ratings. Detailed results are provided in Table 1. Boxplots with overlaid interaction lines are provided in Figure 7.

4.1.1 Perceived Intelligence and Perceived Knowledge.

Perceived Intelligence. We found a statistically significant main effect of the virtual agent's intelligence (Wilks' $\Lambda = .119$, $F[1, 23] = 170.401$, $p < .001$, $\eta_p^2 = .881$). Our participants rated agents with higher intelligence ($M = 5.40$, $SE = .13$) as more intelligent than virtual agents with lower intelligence ($M = 2.88$, $SE = .17$). We did not find a statistically significant main effect of agents' knowledgeability (Wilks' $\Lambda = .926$, $F[1, 23] = 1.839$, $p = .188$, $\eta_p^2 = .074$). However, we found a statistically significant intelligence × knowledgeability interaction effect (Wilks' $\Lambda = .438$, $F[1, 23] = 10.069$, $p = .004$, $\eta_p^2 = .304$). Post-hoc pairwise comparison showed that in the presence of low intelligence, knowledgeability was significant ($t[23] = -2.690$, $p = .013$, $d = -.549$), with LIHK ($M = 3.27$, $SD = 1.36$) rated higher than LILK ($M = 2.49$, $SD = .78$). In the presence of high intelligence, knowledgeability was not significant ($t[23] = 1.378$, $p = .181$, $d = .281$).

Perceived Knowledge. We found a statistically significant main effect of the virtual agent's intelligence (Wilks' $\Lambda = .827$, $F[1, 23] = 4.810$, $p = .039$, $\eta_p^2 = .173$). Our participants rated agents with higher intelligence ($M = 4.73$, $SE = .21$) as more knowledgeable than virtual agents with lower intelligence ($M = 4.27$, $SE = .15$). We also found a statistically significant main effect of agents' knowledgeability (Wilks' $\Lambda = .156$, $F[1, 23] = 124.386$, $p < .001$, $\eta_p^2 = .844$). Our participants rated agents with higher knowledgeability ($M = 5.93$, $SE = .14$) as more knowledgeable than virtual agents with lower knowledgeability ($M = 3.07$, $SE = .24$). Lastly, there was no statistically significant intelligence × knowledgeability interaction effect (Wilks' $\Lambda = .924$, $F[1, 23] = 1.404$, $p = .248$, $\eta_p^2 = .058$).

Intelligence Comparison. We found a statistically significant main effect of the virtual agent's intelligence (Wilks' $\Lambda = .301$, $F[1, 23] = 53.481$, $p < .001$, $\eta_p^2 = .699$). Our participants rated agents with higher intelligence ($M = 4.16$, $SE = .19$) as more likely to be more intelligent than themselves, compared to agents with lower intelligence ($M = 2.71$, $SE = .23$). There was also a statistically significant main effect of agents' knowledgeability (Wilks' $\Lambda = .644$, $F[1, 23] = 12.701$, $p = .002$, $\eta_p^2 = .356$). Our participants rated agents with higher knowledgeability ($M = 3.85$, $SE = .28$) as more likely to be more intelligent than themselves compared to agents with lower knowledgeability ($M = 3.02$, $SE = .13$). Lastly, we

Table 1: RM-ANOVA result for self-reported ratings. Statistical significance was evaluated at $\alpha = .05$. Significant results are shown in bold.

	PI		PK		IC		KC		CP		UV		RP		TR	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
LILK	2.49	.78	2.74	1.02	2.18	.87	3.08	1.53	5.79	.64	3.92	1.08	3.90	1.40	3.25	1.03
LIHK	3.27	1.36	5.80	.87	3.25	1.73	6.11	.68	5.72	.77	3.71	1.04	4.13	1.31	4.29	1.22
HILK	5.54	.60	3.40	1.72	3.86	.72	4.13	1.54	5.82	.34	3.51	1.00	4.82	1.22	5.08	1.04
HIHK	5.26	.94	6.07	.74	4.46	1.38	6.34	.63	6.15	.66	3.46	1.02	4.81	1.18	5.39	.90
Main Effect (Intelligence)																
F		**170.401**		**4.810**		**53.481**		**17.081**		**6.958**		**8.239**		**15.725**		**38.792**
p		**<.001**		**.039**		**<.001**		**<.001**		**.015**		**.009**		**<.001**		**<.001**
η_p^2		**.881**		**.173**		**.699**		**.426**		**.232**		**.264**		**.406**		**.628**
Main Effect (Knowledgeability)																
F		1.839		**124.386**		**12.701**		**74.424**		1.343		.762		.154		**11.744**
p		.188		**<.001**		**.002**		**<.001**		.258		.392		.699		**.002**
η_p^2		.074		**.844**		**.356**		**.764**		.055		.032		.007		**.338**
Interaction Effect (Intelligence × Knowledgeability)																
F		**10.069**		1.404		2.477		**5.205**		**4.942**		.372		.412		3.947
p		**.004**		.248		.129		**.032**		**.036**		.548		.528		.059
η_p^2		**.304**		.058		.097		**.185**		**.177**		.016		.018		.146

Intelligence $df = 1$, Knowledgeabilty $df = 1$, Interaction $df = 1$, and Error $df = 23$.

LILK: Low Intelligence with Low Knowledgeability; LIHK: Low Intelligence with High Knowledgeability; HILK: High Intelligence with Low Knowledgeability; and HIHK: High Intelligence with High Knowledgeability.

PI: Perceived Intelligence; PK: Perceived Knowledge; IC: Intelligence Comparison; KC: Knowledge Comparison; CP: Co-presence; UV: Uncanny Valley Effect; RP: Rapport; and TR: Trust.

did not find statistically significant intelligence × knowledgeability interaction effect (Wilks' $\Lambda = .903$, $F[1, 23] = 2.477$, $p = .129$, $\eta_p^2 = .097$).

Knowledge Comparison. We found a statistically significant main effect of the virtual agent's intelligence (Wilks' $\Lambda = .574$, $F[1, 23] = 17.081$, $p < .001$, $\eta_p^2 = .426$). Our participants rated agents with higher intelligence ($M = 5.23$, $SE = .18$) as more likely to be more knowledgeable than themselves, compared to agents with lower intelligence ($M = 4.60$, $SE = .15$). There was also a statistically significant main effect of agents' knowledgeability (Wilks' $\Lambda = .236$, $F[1, 23] = 74.424$, $p < .001$, $\eta_p^2 = .764$). Our participants rated agents with higher knowledgeability ($M = 6.23$, $SE = .11$) as more likely to be more knowledgeable than themselves, compared to agents with lower knowledgeability ($M = 3.60$, $SE = .28$). Lastly, we found a statistically significant intelligence × knowledgeability interaction effect (Wilks' $\Lambda = .815$, $F[1, 23] = 5.205$, $p = .032$, $\eta_p^2 = .185$). Post-hoc pairwise comparisons showed that in the presence of low intelligence, knowledgeability was significant ($t[23] = -8.036$, $p < .001$, $d = -1.640$), with LIHK ($M = 6.11$, $SD = .68$) rated higher than LILK ($M = 3.08$, $SD = 1.53$). In the presence of high intelligence, knowledgeability was also significant ($t[23] = -6.821$, $p < .001$, $d = -1.392$), with HIHK ($M = 6.34$, $SD = .63$) rated higher than HILK ($M = 4.12$, $SD = 1.54$).

4.1.2 Social and Emotional Experiences.

Co-presence. We found a statistically significant main effect of the virtual agent's intelligence (Wilks' $\Lambda = .768$, $F[1, 23] = 6.958$, $p = .015$, $\eta_p^2 = .232$). Our participants rated their sense of co-presence with virtual agents of higher intelligence in solving the puzzle ($M = 5.98$, $SE = .09$) as higher than with virtual agents of lower intelligence ($M = 5.75$, $SE = .12$). There was no statistically significant main effect of agents' knowledgeability (Wilks' $\Lambda = .945$, $F[1, 23] = 1.343$, $p = .258$, $\eta_p^2 = .055$). Lastly, we also found a statistically significant intelligence × knowledgeability interaction effect (Wilks' $\Lambda = .823$, $F[1, 23] = 4.942$, $p = .036$, $\eta_p^2 = .177$). Post-hoc pairwise comparisons showed that in the presence of low intelligence, knowledgeability was not significant ($t[23] = .442$, $p = .662$, $d = .090$). In the presence of high intelligence, knowledgeability was significant ($t[23] = -3.055$, $p = .006$, $d = -.624$), with HIHK ($M = 6.15$, $SD = .66$) rated higher than HILK ($M = 5.82$, $SD = .34$).

Uncanny Valley. We found a statistically significant main effect of the virtual agent's intelligence (Wilks' $\Lambda = .736$, $F[1, 23] = 8.239$, $p = .009$, $\eta_p^2 = .264$). Our participants perceived the virtual agent with higher intelligence ($M = 3.49$, $SE = .18$) as less uncanny than virtual agents with lower intelligence ($M = 3.81$, $SE = .19$). There

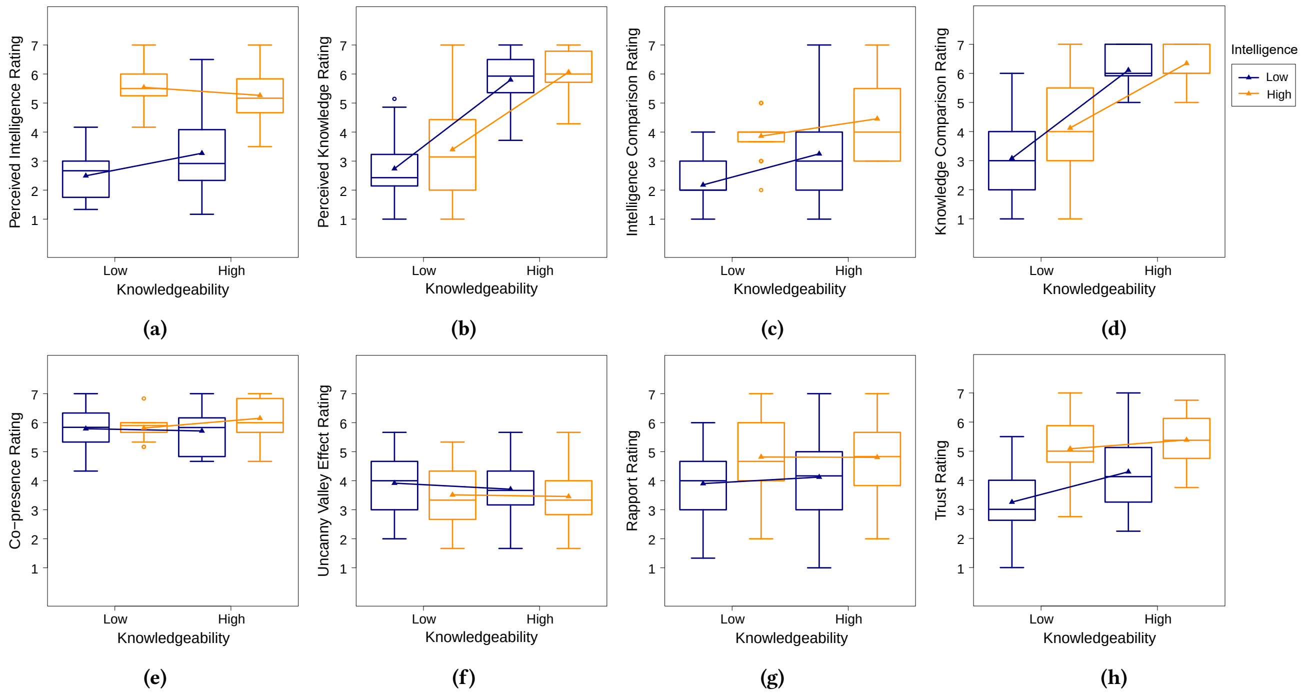


Figure 7: Boxplot with overlaid interaction lines for self-reported ratings. (a) Perceived Intelligence, (b) Perceived Knowledge, (c) Intelligence Comparison, (d) Knowledge Comparison, (e) Co-presence, (f) Uncanny Valley Effect, (g) Rapport, and (h) Trust.

was no statistically significant main effect of agents' knowledgeability (Wilks' $\Lambda = .968$, $F[1,23] = .762$, $p = .392$, $\eta_p^2 = .032$). Lastly, we did not find a statistically significant intelligence × knowledgeability interaction effect (Wilks' $\Lambda = .984$, $F[1,23] = .372$, $p = .548$, $\eta_p^2 = .016$).

Rapport. We found a statistically significant main effect of the virtual agent's intelligence (Wilks' $\Lambda = .594$, $F[1,23] = 15.725$, $p < .001$, $\eta_p^2 = .406$). Our participants rated the sense of rapport with virtual agents with higher intelligence ($M = 4.81$, $SE = .21$) higher than virtual agents with lower intelligence ($M = 4.01$, $SE = .21$). There was no statistically significant main effect of agents' knowledgeability (Wilks' $\Lambda = .993$, $F[1,23] = .154$, $p = .699$, $\eta_p^2 = .007$). Lastly, we did not find a statistically significant intelligence × knowledgeability interaction effect (Wilks' $\Lambda = .982$, $F[1,23] = .412$, $p = .528$, $\eta_p^2 = .018$).

Trust. We found a statistically significant main effect of the virtual agent's intelligence (Wilks' $\Lambda = .372$, $F[1,23] = 38.792$, $p < .001$, $\eta_p^2 = .628$). Our participants rated agents with higher intelligence ($M = 5.23$, $SE = .16$) more trustworthy than virtual agents with lower intelligence ($M = 3.77$, $SE = .17$). There was also a statistically significant main effect of agents' knowledgeability (Wilks' $\Lambda = .662$, $F[1,23] = 11.744$, $p = .002$, $\eta_p^2 = .338$). Our participants rated agents with higher knowledgeability ($M = 4.84$, $SE = .17$) as more trustworthy, compared to agents with lower knowledgeability ($M = 4.17$, $SE = .14$). Lastly, we did not find a statistically significant intelligence × knowledgeability interaction effect (Wilks' $\Lambda = .854$, $F[1,23] = 3.947$, $p = .059$, $\eta_p^2 = .146$).

4.2 Application-logged Data

This section presents the results of the collected logged data. Detailed results are also provided in Table 2. Boxplots with overlaid interaction lines are provided in Figure 8.

4.2.1 *Interaction Dynamics.*

Puzzle Completion Time. We found a statistically significant main effect of the virtual agent's intelligence (Wilks' $\Lambda = .524$, $F[1,23] = 20.920$, $p < .001$, $\eta_p^2 = .476$). The total time spent completing the puzzle with virtual agents with higher intelligence ($M = 172.57$, $SE = 11.13$) was less than that of agents with lower intelligence ($M = 212.97$, $SE = 11.37$). There was no statistically significant main effect of agents' knowledgeability (Wilks' $\Lambda = .971$, $F[1,23] = .689$, $p = .415$, $\eta_p^2 = .029$). Lastly, we did not find a statistically significant intelligence × knowledgeability interaction effect (Wilks' $\Lambda = .952$, $F[1,23] = 1.150$, $p = .295$, $\eta_p^2 = .048$).

Puzzle Pieces Count. We found a statistically significant main effect of the virtual agent's intelligence (Wilks' $\Lambda = .192$, $F[1,23] = 96.498$, $p < .001$, $\eta_p^2 = .808$). The number of puzzle pieces completed by participants was lower when interacting with virtual agents with higher intelligence ($M = 15.06$, $SE = .65$) than with virtual agents with lower intelligence ($M = 21.91$, $SE = .29$). There was also a statistically significant main effect of agents' knowledgeability (Wilks' $\Lambda = .609$, $F[1,23] = 14.775$, $p < .001$, $\eta_p^2 = .391$). The number of puzzle pieces completed by participants was higher when interacting with virtual agents that had higher knowledgeability ($M = 19.23$, $SE = .43$) than with virtual agents that had

Table 2: RM-ANOVA result for application-logged data. Statistical significance was evaluated at $\alpha = .05$. Significant results are shown in bold.

	PCT		PPC		NDGA		NDGG		NDGP	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
LILK	214.35	62.55	21.67	1.74	.15	.07	.03	.04	.04	.05
LIHK	211.58	57.82	22.17	1.76	.04	.07	.04	.07	.04	.04
HILK	166.15	53.85	13.83	3.68	.15	.09	.06	.08	.03	.04
HIHK	178.99	64.44	16.29	3.50	.14	.10	.03	.04	.03	.04
Main Effect (Intelligence)										
F		**20.920**		**96.498**		**8.878**		.253		**5.295**
p		**<.001**		**<.001**		**.007**		.620		**.031**
η_p^2		**.476**		**.808**		**.278**		.011		**.187**
Main Effect (Knowledgeability)										
F		.689		**14.775**		**18.294**		.692		.099
p		.415		**<.001**		**<.001**		.414		.756
η_p^2		.029		**.391**		**.443**		.029		.004
Interaction Effect (Intelligence × Knowledgeability)										
F		1.150		**6.067**		**8.089**		**6.694**		.115
p		.295		**.022**		**.009**		**.016**		.738
η_p^2		.048		**.209**		**.260**		**.225**		.005

Intelligence $df = 1$, Knowledgeabilty $df = 1$, Interaction $df = 1$, and Error $df = 23$.

LILK: Low Intelligence & Low Knowledgeability; LIHK: Low Intelligence & High Knowledgeability; HILK: High Intelligence & Low Knowledgeability; and HIHK: High Intelligence & High Knowledgeability.

PCT: Puzzle Completion Time; PPC: Puzzle Pieces Count; NDGA: Normalized Dwell Gaze on Agent; NDGG: Normalized Dwell Gaze on Puzzle Goal Board; and NDGP: Normalized Dwell Gaze on Puzzle Pieces.

lower knowledgeability ($M = 17.75$, $SE = .40$). Lastly, we found a statistically significant intelligence × knowledgeability interaction effect (Wilks' $\Lambda = .791$, $F[1,23] = 6.067$, $p = .022$, $\eta_p^2 = .209$). Post-hoc pairwise comparison showed that in the presence of low intelligence, knowledgeability was not significant ($t[23] = -1.212$, $p = .238$, $d = -.247$). In the presence of high intelligence, knowledgeability was significant ($t[23] = -3.698$, $p = .001$, $d = -.755$); participants completed more puzzle pieces in the HIHK ($M = 16.29$, $SD = 3.50$) than in the HILK ($M = 13.83$, $SD = 3.68$) condition.

Normalized Dwell Gaze on Agent. There was a statistically significant main effect of the virtual agent's intelligence (Wilks' $\Lambda = .722$, $F[1,23] = 8.878$, $p = .007$, $\eta_p^2 = .278$). Participants' dwell gaze on the virtual agent during the jigsaw puzzle co-solving task was longer in the high intelligence condition ($M = .15$, $SE = .02$) than in the low intelligence condition ($M = .10$, $SE = .01$). There was also a statistically significant main effect of the virtual agent's knowledgeability (Wilks' $\Lambda = .557$, $F[1,23] = 18.294$, $p < .001$, $\eta_p^2 = .443$). Participants' dwell gaze on the virtual agent during the jigsaw puzzle co-solving task was shorter in high knowledgeability ($M = .09$, $SE = .01$) than in low knowledgeability ($M = .15$, $SE = .01$) conditions. Lastly, we also found a significant intelligence × knowledgeability interaction effect (Wilks' $\Lambda = .352$, $F[1,23] = 8.089$, $p = .009$, $\eta_p^2 = .260$). Post-hoc pairwise comparison showed that in the presence of low intelligence, knowledgeability was significant ($t[23] = 4.471$, $p < .001$, $d = .913$); participants gazed at the virtual agent more in LILK ($M = .15$, $SD = .07$) than in LIHK ($M = .04$, $SD = .07$) condition. In the presence of high intelligence, knowledgeability was not significant ($t[23] = .758$, $p = .456$, $d = .155$).

Normalized Dwell Gaze on Puzzle Goal Board. There was no statistically significant main effect of both the virtual agent's intelligence (Wilks' $\Lambda = .989$, $F[1,23] = .253$, $p = .620$, $\eta_p^2 = .011$) and its knowledgeability (Wilks' $\Lambda = .971$, $F[1,23] = .692$, $p = .414$, $\eta_p^2 = .029$). We found a significant intelligence × knowledgeability interaction effect (Wilks' $\Lambda = .775$, $F[1,23] = 6.694$, $p = .016$, $\eta_p^2 = .225$). However, post-hoc pairwise comparison showed that in the presence of low intelligence, knowledgeability was not significant ($t[23] = -.912$, $p = .371$, $d = -.186$) and in the presence of high intelligence, knowledgeability was also not significant ($t[23] = 1.825$, $p = .081$, $d = .373$).

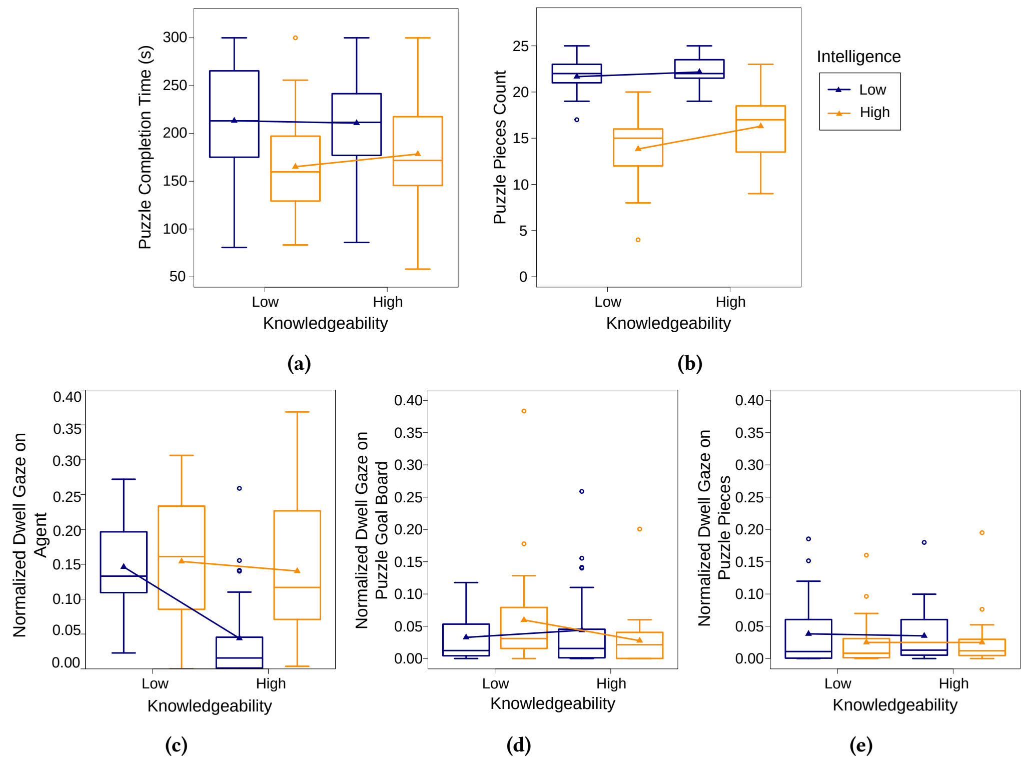


Figure 8: Boxplot with overlaid interaction lines for application-logged data. (a) Puzzle Completion Time, (b) Puzzle Pieces Count, (c) Normalized Dwell Gaze on Agent, (d) Normalized Dwell Gaze on Puzzle Goal Board, and (e) Normalized Dwell Gaze on Puzzle Pieces.

Normalized Dwell Gaze on Puzzle Pieces. We found a statistically significant main effect of the virtual agent's intelligence (Wilks' Λ = .813, $F[1, 23]$ = 5.295, p = .031, η_p^2 = .187). Our participants spent more time looking at the puzzle when the virtual agent had low intelligence (M = .04, SE = .01) compared to the virtual agent with high intelligence (M = .03, SE = .01). There was no statistically significant main effect of the virtual agent's knowledgeability (Wilks' Λ = .996, $F[1, 23]$ = .099, p = .756, η_p^2 = .004), nor a significant intelligence $\times$ knowledgeability interaction effect (Wilks' Λ = .995, $F[1, 23]$ = .115, p = .738, η_p^2 = .005).

4.2.2 Conversational Dynamics.

Conversation Transcript. Through coding on the recorded conversation transcripts, we categorized participants' responses into the three categories:

- **Art-related Responses (AR):** AR included responses related to art, such as the discussion of the puzzle painting, the artist's background, the history behind the art piece, or how they feel looking at the painting (e.g., *"I like how abstract the painting is. I think the brushstrokes are very nice."* and *"Can you tell me a little bit about the history of the painting?"*).
- **Puzzle-related Responses (PR):** PR included responses related to the jigsaw puzzle co-solving task (e.g., *"Let's do the corner pieces first. There are two left. Can you work on one of those?"* and *"I think it is coming along pretty well. We are almost done with it."*).
- **Other Responses (OR):** OR referred to the remaining participants' responses, which were not art-related or puzzle-related. This could be responses such as greetings, repair initiators [93], or conversations on other topics (e.g., *"Thank you. And you have a great day, too."* and *"I like to read too. What book are you reading currently?"*).

Two independent coders applied the above coding scheme to the transcripts, and any discrepancies were resolved through discussion until consensus was reached. After establishing the final codes, we counted the number of categorized participants' responses under all conditions. A proportion percentage chart is provided in Figure 9. Specifically, in the LILK condition: AR = 178 (69.53%), PR = 52 (20.31%), and OR = 26 (10.16%). In the LIHK condition: AR = 113 (85.61%), PR = 8 (6.06%), and OR = 11 (8.33%). In the HILK condition: AR = 192 (74.42%), PR = 23 (8.91%), and OR = 43 (16.67%). Finally, in the LIHK condition: AR = 141 (95.27%), PR = 5 (3.38%), and OR = 2 (1.35%). Detailed results are provided in Table 3. Boxplots with overlaid interaction lines are provided in Figure 10.

Art-related Response Count. There was a statistically significant main effect of the virtual agent's intelligence (Wilks' Λ = .770, $F[1, 23]$ = 6.889, p = .015, η_p^2 = .230). Participants' responses on

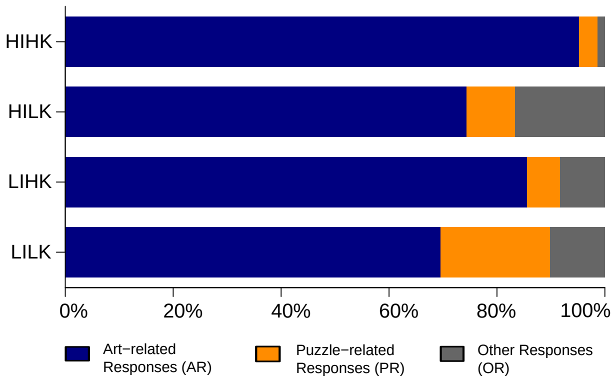


Figure 9: The percentage bar chart presents the participants' response categorization in all four conditions.

Table 3: RM-ANOVA result for conversational response categories. Statistical significance was evaluated at $\alpha = .05$. Significant results are shown in bold.

	AR		PR		OR	
	M	*SD*	*M*	*SD*	*M*	*SD*
LILK	7.42	1.96	2.17	2.06	1.08	2.10
LIHK	4.71	1.08	.33	.92	.46	.83
HILK	8.00	2.57	.96	1.85	1.79	2.36
HIHK	5.96	1.08	.21	.83	.08	.28
Main Effect (Intelligence)						
F	**6.889**		**5.183**		.240	
p	**.015**		**.032**		.629	
η_p^2	**.230**		**.184**		.010	
Main Effect (Knowledgeability)						
F	**52.662**		**32.940**		**23.602**	
p	**<.001**		**<.001**		**<.001**	
η_p^2	**.696**		**.589**		**.506**	
Interaction Effect (Intelligence × Knowledgeability)						
F	.778		2.794		1.536	
p	.387		.108		.228	
η_p^2	.033		.108		.063	

Intelligence $df = 1$, Knowledgeabilty $df = 1$, Interaction $df = 1$, and Error $df = 23$.

LILK: Low Intelligence & Low Knowledgeability; LIHK: Low Intelligence & High Knowledgeability; HILK: High Intelligence & Low Knowledgeability; and HIHK: High Intelligence & High Knowledgeability.

AR: Art-related Responses; PR: Puzzle-related Responses; and OR: Other Responses.

art-related topics appeared more often in the high intelligence conditions ($M = 6.98$, $SE = .32$) than in the low intelligence conditions ($M = 6.06$, $SE = .19$). There was also a statistically significant main effect of the virtual agent's knowledgeability (Wilks' $\Lambda = .304$, $F[1, 23] = 52.662$, $p < .001$, $\eta_p^2 = .696$). Participants' responses on art-related topics appeared more often in low knowledgeability conditions ($M = 7.71$, $SE = .33$) than in high knowledgeability conditions ($M = 5.33$, $SE = .16$). Lastly, we did not find a statistically significant intelligence × knowledgeability interaction effect (Wilks' $\Lambda = .967$, $F[1, 23] = .778$, $p = .387$, $\eta_p^2 = .033$).

Puzzle-related Response Count. There was a statistically significant main effect of the virtual agent's intelligence (Wilks' $\Lambda = .816$, $F[1, 23] = 5.183$, $p = .032$, $\eta_p^2 = .184$). Participants' responses on puzzle-related topics appeared more often in the low intelligence conditions ($M = 1.25$, $SE = .23$) than in the high intelligence conditions ($M = .58$, $SE = .25$). There was also a statistically significant main effect of the virtual agent's knowledgeability (Wilks' $\Lambda = .411$, $F[1, 23] = 32.940$, $p < .001$, $\eta_p^2 = .589$). Participants' responses on puzzle-related topics appeared more often in low knowledgeability conditions ($M = 1.56$, $SE = .26$) than in high knowledgeability conditions ($M = .27$, $SE = .17$). Lastly, we did not find a statistically significant intelligence × knowledgeability interaction effect (Wilks' $\Lambda = .892$, $F[1, 23] = 2.794$, $p = .108$, $\eta_p^2 = .108$).

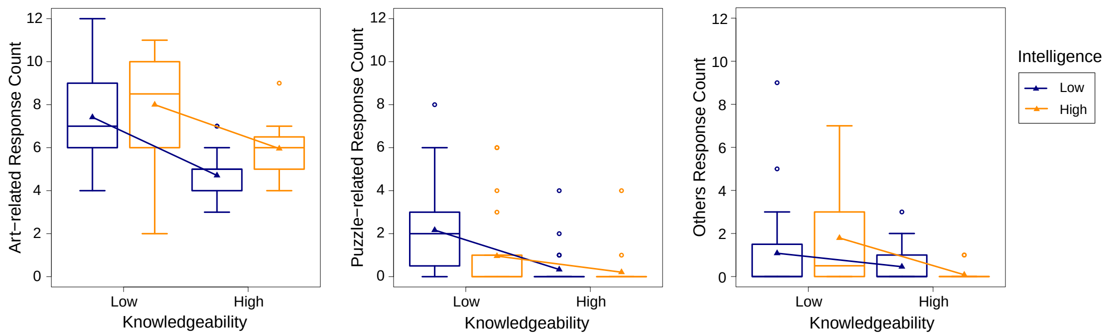
Other Response Count. There was no statistically significant main effect of the virtual agent's intelligence (Wilks' $\Lambda = .990$, $F[1, 23] = .240$, $p = .629$, $\eta_p^2 = .010$). Yet, there was a statistically significant main effect of the virtual agent's knowledgeability (Wilks' $\Lambda = .494$, $F[1, 23] = 23.602$, $p < .001$, $\eta_p^2 = .506$). Participants' responses on other topics appeared more often in low knowledgeability conditions ($M = 1.44$, $SE = .25$) than in high knowledgeability conditions ($M = .27$, $SE = .09$). Lastly, we did not find a statistically significant intelligence × knowledgeability interaction effect (Wilks' $\Lambda = .937$, $F[1, 23] = 1.536$, $p = .228$, $\eta_p^2 = .063$).

4.3 Correlations

We calculated the Pearson correlation coefficient of all combinations from our self-reported ratings and application-logged data (see Table 4). The analysis shows strong correlations ($.600 \leq |r|$) between **Perceived Intelligence** and each of the following: **Intelligence Comparison**, **Trust**, and **Puzzle Pieces Count**. **Perceived Knowledge** also showed a strong correlation with **Knowledge Comparison**. Furthermore, moderate correlations ($.400 \leq |r| < .600$) were observed between **Perceived Intelligence** and **Rapport**; **Perceived Knowledge** and both **Intelligence Comparison** and **Trust**; **Intelligence Comparison** and each of **Knowledge Comparison**, **Trust**, and **Puzzle Pieces Count**; **Knowledge Comparison** and **Trust**; **Co-presence** and both **Rapport** and **Trust**; **Uncanny Valley** and both **Rapport** and **Trust**; and finally, **Rapport** and **Trust**. All reported correlations were statistically significant at $p < .01$. All other correlations were weak.

4.4 Open-ended Feedback

We analyzed our participants' open-ended feedback qualitatively. Through thematic analysis, we observed comments related to experience based on the virtual agent's knowledgeability and intelligence, the virtual agent's behavior, and interaction design.



(a) (b) (c)

Figure 10: Boxplots with interaction lines for categorization of participants' conversational responses. (a) Art-related Responses, (b) Puzzle-related Responses, and (c) Other Responses.

Table 4: Pearson correlation coefficient r of all self-reported ratings and application-logged data. Significant results were bolded and colored based on their correlation levels.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) Perceived Intelligence													
(2) Perceived Knowledge	**.251***												
(3) Intelligence Comparison	**.649****	**.409****											
(4) Knowledge Comparison	**.255***	**.833****	**.416****										
(5) Co-presence	**.281****	.191	**.225***	**.222***									
(6) Uncanny Valley Effect	**-.316****	**-.305****	-.112	**-.232***	**-.228***								
(7) Rapport	**.509****	**.272****	**.348****	**.286****	**.466****	**-.480****							
(8) Trust	**.671****	**.458****	**.572****	**.454****	**.400****	**-.489****	**.586****						
(9) Puzzle Completion Time	**-.317****	.024	-.060	.119	**-.298****	**.266****	**-.253****	**-.300****					
(10) Participant Puzzle Number	**-.625****	-.009	**-.462****	-.078	-.112	-.011	**-.225***	**-.390****	-.013				
(11) Gaze on Agent	.159	**-.270****	-.065	**-.283****	.140	.100	.045	.016	**-.230****	-.106			
(12) Gaze on Puzzle Goal Board	.004	.000	-.052	.058	-.138	-.125	.048	.049	**.147****	-.087	**-.010***		
(13) Gaze on Puzzle Pieces	**-.159***	-.091	-.112	-.035	.030	**.232***	-.058	**-.195****	-.044	.104	-.107	**-.229***	

**: $p < .01$ *: $p < .05$ ▢ : strong correlation $.600 \leq |r|$ ▢ : moderate correlation $.400 \leq |r| < .600$ ▢ : weak correlation $.100 \leq |r| < .400$

Two participants noted their experience interacting with low intelligence virtual agents as frustrating and disruptive. P18 mentioned, *"She didn't even know how to solve the puzzles correctly. It was frustrating."* P14 said, *"When the virtual agent put the puzzle pieces in the wrong places, it was pretty disruptive and confused me when I solved the puzzle."* P21 implied that it was difficult to connect with a low intelligence virtual agent: *"It's hard to have a connection with the one who is trying to put the wrong piece of the puzzle in the wrong place."*

Five participants shared that they enjoyed the conversation with highly knowledgeable virtual agents. P2 said, *"I liked it when the virtual agent could answer my questions related to art in depth."* P5 mentioned *"The latter one* (i.e., the HIHK condition) *was good. I feel that I can learn something new."* P18 noted *"It was fun and engaging to talk with the woman when she knew a lot about the artist."* P11 stated that the high knowledgeability and high intelligent virtual agent helped with immersiveness, *"In the condition where it is knowledgeable and good at playing the puzzle, I really feel I am immersed in the scenario."* P21 stated, *"The last one* (i.e., the HIHK condition) *is my favorite."*

However, four participants implied they felt more connected with low knowledgeability virtual agents. One participant argued that it felt eerie and robotic when the virtual agent knew too much; P8 said, *"I think having answers like 'I don't know' or 'I haven't read about it' makes the virtual agent more human-like. When she was providing a lot of information, it made her more robotic, like a Google Search. But when she says she doesn't know, I don't think she is less intelligent or less knowledgeable. It is more like a real person and actually made me feel more connected."* Another participant perceived agent knowledgeability as a factor in establishing the social status of a mentor or a friend, P18: *"In the case where the virtual agent was knowledgeable, it felt like the virtual agent was more of a mentor. Even when she was not intelligent at solving the puzzle, it felt like she was expecting me to do the task, and I was OK with it. Whereas, when the virtual agent was not knowledgeable, but intelligent, it felt like someone I could have casual conversations with, like the virtual agent was a peer or a friend."* Other feedback also suggested that low knowledgeability provided stronger connections. For example, P12 said, *"I like the third agent* (i.e., the HILK condition); *it cares about our cooperation on the puzzle, and the overall experience is smooth and comfortable."* P13 added, *"The first agent's* (i.e., the LILK condition) *personality was more friendly, and the third* (i.e., the HILK condition) *was the second most friendly."*

Four participants shared their experiences collaborating on the puzzle while conversing with the virtual agent. P1 said, *"It was a bit hard to focus on the task when the virtual agent was talking next to me."* P10 remarked, *"I felt that the virtual agent sometimes did not pay attention to the puzzle and paid more attention to the conversation. She likes to gossip more than to work on the puzzle."* P14 explained, *"Towards the beginning, I was mostly focused on trying to solve the puzzle and did not pay attention to the conversation, but for the later trials, I got used to the puzzle and focused more on how the virtual agent behaved."* Lastly, P16 added, *"I liked talking to the virtual agent while solving the puzzle."*

Four participants suggested adjusting the virtual agent's animation to make it more human-like. P4 noted, *"The virtual agent feels eerie, especially with no head movement and weird lip sync."* P17 added, *"Lip sync is not natural, which will affect my experience of believing the virtual agent is a real human."* P14 explained, *"Sometimes I found that when the virtual agent just finished processing what I said, their head just snapped towards me."* P23 shared, *"When I said something to the virtual agent, the virtual agent would get stuck and suddenly turn towards me. That experience was not very comfortable."*

Lastly, two participants also reported unpleasant experiences of being interrupted by the virtual agent. P6 said, *"Sometimes she disrupted me when I was talking, which was annoying."* and P17 echoed, *"I kept getting interrupted by the virtual agent when I tried to ask a question."*

5 Discussion

5.1 Perceived Intelligence and Perceived Knowledge

Our findings revealed interesting relationships between virtual agents' puzzle-solving abilities, their domain-specific knowledge, and participants' perceptions of intelligence and knowledgeability. While intelligence and knowledgeability emerged as distinct traits, they were also interdependent in shaping participants' judgments of the agents.

- **RQ1-1:** Virtual agents were perceived as more intelligent when they solved the puzzle without making mistakes. However, while knowledgeability alone did not significantly influence participants' perceptions of the virtual agent's intelligence, under low intelligence conditions, higher knowledgeability notably enhanced perceived intelligence.

The results suggested that virtual agents' ability to solve puzzles correctly was mapped with participants' perceived intelligence, validating our implementation of agent intelligence based on puzzle-solving accuracy, following Choi et al.'s studies [27, 39]. With no influence from agent knowledgeability in art, it emphasized that participants scored agents' intelligence mainly by their cognitive ability in problem solving and not much on their trait of possessing in-depth knowledge in a specific domain. Nevertheless, the interaction effect suggested a connection between intelligence and knowledgeability. When the virtual agents showed competent abilities in solving the puzzle while being knowledgeable, participants would consider them even more intelligent. This echoed and expanded Rolfhus and Ackerman's [89] statement that domain-specific knowledge plays a significant role in intellectual performance and is positively related to general intelligence. When agents demonstrate strong knowledge in a specific domain, participants may assume they also possess higher cognitive abilities, even though correctly solving puzzles and having a deeper understanding of art are two distinctive traits. Our findings provided evidence to support our argument that virtual agent intelligence and knowledgeability are two distinct yet interrelated traits [2, 56] that can be perceived differently by participants and influence one another.

- **RQ1-2:** Virtual agents were perceived as more knowledgeable by participants when they demonstrated in-depth domain knowledge. Moreover, high intelligence enhanced participants' perceptions of the virtual agents' knowledgeability.

Results validated our knowledgeability level implementation using LLM prompts following prior work of Yang et al. [112]. According to Radecki and Jaccard [84], humans' perception of knowledge is related to their ability to convey it. Moreover, through the Media Equation theory, Reeve and Nass [86] stated that humans respond to media (e.g., a character in software) socially and naturally, and their experience is the same as that with other humans. Building on these prior studies, we argue that our participants might have perceived the virtual agent as more knowledgeable when it delivered a deeper understanding of the artwork. In addition, with the significant effect of agent intelligence on perceived knowledgeability and previous results on the interaction effect for the perceived intelligence rating, we argue that there was perhaps a two-way connection between participants' perceptions of the virtual agent's intelligence and knowledgeability. The results of perceived knowledgeability revisited the concept that perceived intelligence and perceived knowledge could be interrelated [2, 56].

- **RQ1-3:** Virtual agents with low intelligence or low knowledgeability were perceived by participants as less intelligent than themselves. High knowledgeability enhanced the perceived intelligence of the virtual agent when combined with high intelligence.

From the results, we noticed that through our experimental design, participants normally would not consider the virtual agent more intelligent than themselves, with only the HIHK condition ($M = 4.46$) receiving a score above the midpoint of the scale. Hidalgo et al. [53] found that people judge machines by assessing their ability to perform commanded tasks. Similarly, we think that participants perceived the virtual agent as less intelligent than themselves when it neither solved the puzzle correctly nor answered their questions sufficiently. Moreover, participants would be more likely to perceive the virtual agent as more intelligent than themselves when the virtual agent demonstrated both high intelligence and high knowledgeability. With high domain-specific knowledgeability, the virtual agent could be projected as an expert, leading participants to assume they are more intelligent than themselves [56, 89].

- **RQ1-4:** Virtual agents that displayed high knowledgeability were perceived by participants as more knowledgeable than themselves. In contrast, virtual agents with low intelligence were more likely to be perceived as less knowledgeable than the participants. Moreover, the significant interaction effect suggested that virtual agents were more likely to be perceived as knowledgeable, even under low knowledgeability conditions, when they demonstrated high intelligence.

According to our results, participants considered higher knowledgeability virtual agents more knowledgeable than themselves. Additionally, the interaction effect suggested that when the virtual agent demonstrates poor problem-solving skills, participants might question their overall cognitive capacity compared to themselves, including their knowledge. This aligned with Zhang et al.'s [115] work, who found that people judge humans and machines differently. However, when machines exhibit higher agency, people begin to judge them more similarly to humans. In our study, low ability to solve the puzzle might be read as low agency, providing a strict benchmark from participants' ratings on their knowledgeability compared to the virtual agent.

5.2 Social and Emotional Experiences

Beyond cognitive evaluations, participants' perceptions of intelligence and knowledgeability also influenced their social and emotional experiences with the virtual agents. We found that factors such as co-presence, uncanniness, rapport, and trust varied depending on the agents' demonstrated intelligence and knowledgeability.

- **RQ2-1:** Virtual agents that demonstrated high intelligence were associated with a higher sense of co-presence, as participants reported. A significant interaction effect also suggested that under high intelligence conditions, higher knowledgeability enhanced participants' sense of co-presence.

From the result, we observed that virtual agents' intelligence strongly impacted participants' sense of co-presence with the virtual agent. In Choi et al.'s [26] study, participants reported that they collaborated better with a virtual agent when it completed the given tasks more accurately. Building on this finding, we think that a stronger sense of "working toward a shared goal" occurred when interacting with virtual agents that solved the puzzles correctly, enhancing our participants' perception of being with another individual in the virtual environment. However, when the virtual agent was not solving the puzzle correctly, especially in the LILK condition, we observed an increase in puzzle-related responses from participants, who tried to correct the virtual agent's behavior. Contrarily, the virtual agents' knowledgeability did not impact participants' sense of co-presence. This also aligned with Yang et al.'s [112] previous work, where they found that different depths of knowledge in a virtual agent did not impact their sense of co-presence with the virtual agent as long as it was responsive in conversational interaction. This result highlighted the importance of the virtual agents' responsiveness in providing a strong sense of co-presence. As for the statistically significant interaction effect between the two independent variables, we argue that a higher virtual agent's knowledgeability would increase participants' sense of co-presence under high intelligence conditions, resulting in a stronger sense of being with another individual in the virtual environment, perhaps because the virtual agent's knowledgeability was interpreted as complementary support rather than a distraction.

- **RQ2-2:** Virtual agents with low intelligence that were incompetent in solving puzzles were perceived by participants as more uncanny than virtual agents with high intelligence. However, the virtual agent's knowledgeability did not influence participants' perceptions of the uncanny valley effect.

Based on Ho and McDorman's scale [54], a lower score in the uncanny valley implies lower eeriness and higher human-likeness and attractiveness. Our results showed that participants regarded high intelligence virtual agents as more human-like and attractive than low intelligence ones. Also, it extended partially Moussawi et al.'s [71] findings on the positive association between perceived intelligence and perceived anthropomorphism. We argue that participants considered the experience collaborating with high intelligence virtual agents more enjoyable and thus derived a stronger perception of the virtual agent's attractiveness, lowering their perception of eeriness. Additionally, Schwind et al.'s [94] study stated that stylized and aesthetic virtual agent design can prevent the uncanny valley effect. Our finding extends this prior work by suggesting that the virtual agent's intelligence level can be considered as another factor affecting the uncanny valley.

- **RQ2-3:** Virtual agents that demonstrated high intelligence and performed better at solving puzzles elicited higher rapport from participants than those with lower intelligence. However, the virtual agent's knowledgeability did not influence participants' rapport with the virtual agent.

This finding suggested that during a collaborative interaction with a virtual agent, its ability to perform tasks correctly impacts the establishment of rapport. Building on this, our result extends Choi et al.'s [25] finding, which revealed that their participants reported a higher level of rapport when interacting with an altruistic virtual agent than an egoistic one, by indicating that a virtual agent's intelligence could serve as another factor influencing rapport in human-agent interaction. Furthermore, Seo et al. [97] stated the importance of teamwork abilities in collaborative robots for rapport-building with humans. This was reflected in our results and participants' feedback, where they referred to their experience interacting with low intelligence virtual agents as frustrating and disruptive. Contrarily, the virtual agent's knowledgeability had no significant main effect on rapport. This result might correspond to participants' feedback that they felt less connected with high knowledgeability virtual agents and that their behavior to provide extensive information felt robotic and unnatural. This also aligned with Cerekovic et al.'s [22] findings that turn-taking and social cues strongly impact rapport with conversational virtual agents. In our high knowledgeability conditions, the virtual agents provided vast information in a single exchange, diminishing the turn-taking interaction. This negatively impacted participants' perception of the virtual agent, as some considered it to be showing off. Our findings suggest further exploring adaptive turn-taking methods and designs in knowledgeable conversational agents.

- **RQ2-4:** Virtual agents that demonstrated high intelligence in solving puzzles and those that possessed high knowledgeability in art were perceived by participants as more trustworthy.

Participants' trust in the virtual agent could be mediated by its ability to solve the puzzle correctly and provide informative knowledge about the puzzle art piece. This result coincides with Gupta et al.'s [50] findings that interacting with virtual agents that show collaborative abilities through higher cognitive load levels can increase users' trust. Also, Glikson and Woolley [46] reported that humans trust AI representations, including virtual agents, more when they perceive them as more intelligent. Based on these prior studies and our findings on perceived intelligence, we think participants trusted the virtual agent with higher intelligence more because it performed the collaboration task in an error-free manner. Moreover, the significant main effect in the virtual agent's knowledgeability also aligned with Pan and Steed's [79] work on exploring user trust between experts and non-experts by comparing avatars, videos, and robots. Such results similarly extend Beelen et al.'s [14] findings that children trusted knowledgeable robots more, showing that users are more likely to trust entities that exhibit higher expertise. In our study, high knowledgeability virtual agents provided detailed background information and facts about art, depicting their expertise in the domain of art. This characteristic enhanced participants' trust in the virtual agents.

5.3 Interaction Dynamics

The interaction between intelligence and knowledgeability also affected how participants engaged with the jigsaw puzzle co-solving task itself. Differences in task completion time, number of pieces solved, and gaze behaviors highlighted how virtual agent performance shaped collaboration and workload distribution.

- **RQ3-1:** Virtual agents with low intelligence led participants to spend more time completing the puzzle than virtual agents with high intelligence. However, virtual agent knowledgeability did not impact puzzle completion time under our experimental setup.

Collaborating with low intelligence virtual agents required participants to correct their moves by replacing the puzzle pieces, which took them longer to solve. It reflected the cognitive load theory and suggested that human working memory is limited, requiring more time when dealing with tasks of higher complexity [34]. Contrarily, the virtual agent's knowledgeability showed no impact, reflecting that conversing with both low and high knowledgeability virtual agents in our experimental settings did not significantly impact users' cognitive load in solving the puzzle. We argue that results might differ if the jigsaw puzzle co-solving task was more complex and required more participant retention with the intervention of conversational exchanges.

- **RQ3-2:** Virtual agents with low intelligence were associated with a higher number of puzzle pieces completed by participants compared to those with high intelligence, and agents with high knowledgeability led to more pieces completed than those with low knowledgeability. Moreover, a significant interaction effect indicated that under high intelligence conditions, more puzzle pieces were completed by the participants when interacting with a high knowledgeability virtual agent. However, this result is reported alongside a potential interaction design confound in our current application, where high knowledgeability agents hold on to puzzle pieces while speaking.

Similarly, collaborating with low intelligence virtual agents required participants to solve more puzzle pieces by themselves, resulting in a higher number. Interestingly, we also observed that the number increased when interacting with high knowledgeability virtual agents. This could result in a system design decision. In our application, we set the virtual agent to concentrate on the participant while exchanging conversations. We set this design implementation, assuming the participants would pause or slow down with the jigsaw puzzle co-solving task when the virtual agent spoke. However, our participants tend to multitask and still engage with the puzzle at a similar speed while listening to agents' responses, even when told they would have sufficient time to finish the puzzle. This resulted in more puzzle pieces completed by participants in higher knowledgeability conditions since higher knowledgeability virtual agents provide a more extensive exchange, giving the participants more time and opportunities to complete the puzzle while the virtual agent was speaking. This result also indicated that participants were able to manage their cognitive load effectively despite the multitasking setting, implying that solving the puzzle while conversing with the virtual agent about art may not have exceeded their cognitive capacity, which would cause retention loss and reflect on task completeness [78, 104]. This might also indicate that our participants focused more on task completeness in puzzle solving, as observed in qualitative feedback, that they would be more frustrated when the virtual agent could not solve the puzzle correctly than when the virtual agent did not have higher knowledgeability of the artwork.

- **RQ3-3:** Virtual agents that appeared more intelligent and more knowledgeable were looked at more by participants, with an interaction effect showing reduced gaze in low intelligence conditions. As for gaze toward the puzzle goal board, although there was a significant interaction effect, pairwise comparisons for knowledgeability within both low and high intelligence levels did not reach significance. This indicated that the observed interaction may reflect a broader trend, as knowledgeability did not produce statistically reliable differences within each intelligence level, suggesting its influence on gaze behavior emerged only in combination with intelligence. Lastly, virtual agents with low intelligence led participants to focus more on the puzzle pieces.

According to past research [61, 68], gaze fixation strongly indicates users' attention and cognitive focus on specific elements during a task. Participants needed to focus on correcting the puzzle pieces when interacting with low intelligence virtual agents; this resulted in an increase in dwell gaze distribution on the puzzle pieces and a decrease in dwell gaze on the virtual agent. Contrarily, when participants solved the puzzle with the high intelligence virtual agent, they did not need to correct the agent's mistakes. Thus, instead of checking the puzzle pieces to correct them, they gaze at the virtual agent longer. Moreover, our system design provided more extended dialogues from high knowledgeability virtual agents, which might also result in longer dwell gaze on the virtual agents in such conditions.

5.4 Conversational Dynamics

We also examined how the virtual agents' knowledgeability and intelligence influenced the content and flow of conversations. Participants' engagement in either art-related or puzzle-related dialogue reflected not only the agents' design but also participants' adaptive communication strategies during collaboration.

- **RQ4-1:** Virtual agents elicited more art-related conversations from participants under high intelligence and low knowledgeability conditions, and more puzzle-related discussions under low intelligence and low knowledgeability conditions.

These findings showed that participants were more engaged in art-related topics when conversing with high intelligence virtual agents. When the virtual agent was more competent in solving the puzzle, participants would focus more on art-related topics. Participants reported their experience as enjoyable and desirable when the virtual agent could provide informative content and facts about the artwork or the history behind the art piece, which was reflected in the high percentage of art-related exchanges in the transcripts. This expanded Frummet et al.'s [41] study, where they discovered more knowledge accumulation when interacting with an active digital cooking assistant. Our result echoed that expertise in virtual agents also enhances knowledge accumulation through conversational interaction. The results where participants responded more on puzzle-related topics in low knowledgeability conditions reflected their focus on the jigsaw puzzle co-solving task while interacting with virtual agents of less knowledge in art. Moreover, we noticed several attempts by participants to correct the virtual agents' actions or suggest better workflows when the virtual agent was not capable of solving the puzzle in low intelligence conditions, resulting in an increase in puzzle-related responses from the participants. Collaborative tasks incorporating conversational interaction allow participants to point out the problem directly through free-flow exchanges. We attribute the result of fewer art-related responses in low knowledgeability conditions to high knowledgeability agents giving longer and more informative responses, taking up more time within the session time, resulting in fewer responses from the participants. In addition, although we did not program the virtual agent to alter its behavior based on participants' verbal corrections, we observed that participants preferred the virtual agent acknowledging their mistakes and performing changes accordingly. This exploration extended Choi et al.'s [28] study on the self-correction behavior of virtual agents during a puzzle co-solving task. We argue that acknowledging mistakes through utterance could enhance participants' interaction experience with the virtual agent.

5.5 Design Considerations

Based on our findings and participants' feedback, we propose four main design recommendations for developing collaborative conversational virtual agents that are both intelligent and knowledgeable.

- According to the findings in RQ1, virtual agent intelligence and knowledgeability are traits that can be designed differently and shape human perceptions simultaneously. When designing virtual agents, we should acknowledge the differences between these cognitive behaviors to guide development approaches and set appropriate user expectations.
- According to the findings in RQ2, in task-oriented applications that require cognitive abilities, such as solving a jigsaw puzzle, an intelligent virtual agent should complete a given task well on its own. This, in turn, can improve human perception of co-presence with the agent, reduce uncanniness, and enhance rapport and trust. Similarly, a knowledgeable virtual agent should provide in-depth information as a response to the user's query, as a highly knowledgeable virtual agent that offers in-depth information can boost user trust.
- According to the findings in RQ3, agent intelligence and knowledgeability levels can impact task competence and also shift users' focus. When designing situated human-agent interactions involving multiple agent modalities, researchers, VR designers, and developers should consider how agents' different cognitive abilities may jointly shape users' attention allocation, cognitive load, and further alter their interaction with the agent.
- According to the findings in RQ4, together with participants' feedback, it is suggested that when interacting with highly knowledgeable virtual agents, the delivery style can influence user perceptions of human-likeness and rapport. While detailed information is beneficial, overly mechanical or rigid delivery may create eeriness and reduce user comfort. In addition, an overemphasis on delivering detailed content may unintentionally shift the agent's role from a task collaborator to an informational provider, making the virtual agent less human and more instructional, thereby reducing conversational reciprocity. Researchers and VR developers should consider this by ensuring that virtual agents express knowledge in a comprehensive yet balanced manner to maintain user engagement.

5.6 Limitations

Our study has several limitations that need to be addressed. While these limitations highlight areas for improvement in future studies, they also provide important context for interpreting our findings, without compromising the validity of our results.

First, our structured experimental design focused on a single factor of intelligence and knowledgeability. We only take the virtual agent's ability to solve jigsaw puzzles to evaluate its perceived intelligence. However, intelligence may encompass several factors, such as linguistic intelligence, logical-mathematical intelligence, or spatial intelligence [44]. Similar to knowledgeability, although we instructed our participants to converse with the virtual agent about the puzzle art piece, we did not restrict them to the topic they conversed about, and through free-flow conversations, as it is natural to jump topics through conversational exchanges. With this limitation, the virtual agent can share information on another endeavor that might enhance ratings on perceived knowledge. We tried to minimize this limitation through the design of our survey and only focused on evaluating participants' perceived knowledge of the virtual agent in the domain of art.

Second, the exclusive use of a female virtual agent might have introduced a gender effect, potentially influencing perceived intelligence through physical attractiveness [58]. However, it is worth noting that a recent systematic review [111] reported that female agents are disproportionately prevalent in conversational virtual reality applications, suggesting a broader design bias that may also shape user perceptions. Moreover, we omitted measuring participants' anthropomorphism tendency, which limited our ability to evaluate its moderating role.

Third, our participants reported the virtual agent's unnatural behaviors and animations in their open-ended feedback. The limitation of perfecting virtual agents' conversational gestures and making them more human-like remains challenging. This limitation might impact participants' experience, but should not alter our results in exploring the effect of agent intelligence and knowledgeability. Moreover, the response delays due to API latency throughout the conversational interaction also influenced participants' experiences. This system limitation, caused by connection or API stability, sometimes disrupted the natural flow of conversation and may have affected how participants perceived the virtual agent's responsiveness or conversational competence. Additionally, there was a design oversight in high knowledgeability conditions where the agent held on to a puzzle piece while speaking. Because of this behavior, there might have been a confound affecting how the results should be interpreted.

Fourth, we missed the opportunity to explore the effects of longer interaction durations and gain a deeper understanding of participants' fatigue levels in the immersive environment. Our experimental design imposed a five-minute time limit for each condition, which may have limited our ability to observe participants' perceptions under extended exposure to the virtual agent in the virtual environment. Also, we missed the opportunity to investigate whether participants' responses were influenced by fatigue or disengagement over time, which could have confounded the results.

Fifth, we measured how long participants gazed at the virtual agent, puzzle goal board, and puzzle pieces using a pseudo-eye-tracking approach. However, we did not record how long participants gazed at specific body parts of the virtual agent, such as its face or hands, as all body parts of the virtual agent were tagged with the same key in our implementation. Moreover, we did not measure how much the virtual agent gazed at participants during co-solving a jigsaw puzzle. Due to these missing measurements, we were unable to engage in deeper discussions of gaze behaviors, including mutual and shared gaze between a human and an agent [6, 7].

Lastly, the ramifications of our experimental design may not accurately replicate the real-world interaction. Yet, we design the system from an experimental point of view to provide insight into agent intelligence and knowledgeability. In addition, although the number of participants ($N = 24$) satisfied the minimum sample size required by our power analysis, it represented the minimum and was restricted to a specific age and education level group.

6 Conclusion and Future Work

In this study, we investigated the nuanced interrelation between the virtual agent's intelligence and knowledgeability. We focused on how these two factors influence participants' perceptions of the virtual agent, their social and emotional experiences, and their intricate interaction dynamics in an immersive VR environment. Through our experimental design of a collaborative jigsaw puzzle co-solving task featuring conversational virtual agents with low and high knowledge about art, we gained insight into how participants interpret and respond to these cognitive traits of virtual agents. Our research is inspired by neuroscience and cognitive science foundations, highlighting the distinction between human intelligence and knowledgeability, and we simulate such abilities in virtual agents. We extend this concept and connect it with the field of human-agent interaction, aiming to offer a deeper understanding of the design of virtual agents' cognitive abilities and human perceptions.

Future work can explore how different designs of virtual agents' cognitive dimensions impact human perceptions in immersive environments. For instance, emotional intelligence, creativity, reasoning, learning, and memorizing capability, to name but a few. Furthermore, understanding the effect of multitasking task complexity through incorporating different levels of cognitive load tasks would also be an interesting area to explore. By examining these cognitive traits across diverse tasks, we can reshape how we design virtual agents and support the development of more socialized, adaptive, and effective virtual agents.

References

[1] Phillip L Ackerman. 1996. A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence* 22, 2 (1996), 227–257.
[2] Phillip L Ackerman and Margaret E Beier. 2005. Knowledge and intelligence. *Handbook of understanding and measuring intelligence* (2005), 125–139.
[3] Isayas Berhe Adhanom, Paul MacNeilage, and Eelke Folmer. 2023. Eye tracking in virtual reality: a broad review of applications and challenges. *Virtual Reality* 27, 2 (2023), 1481–1505.
[4] Badr AlKhamissi, Millicent Li, Asli Celikyilmaz, Mona Diab, and Marjan Ghazvininejad. 2022. A review on language models as knowledge bases. *arXiv preprint arXiv:2204.06031* (2022).
[5] John R Anderson. 1989. A theory of the origins of human knowledge. *Artificial intelligence* 40, 1-3 (1989), 313–351.
[6] Sean Andrist, Michael Gleicher, and Bilge Mutlu. 2017. Looking coordinated: Bidirectional gaze mechanisms for collaborative interaction with virtual characters. In *Proceedings of the 2017 CHI conference on human factors in computing systems.* 2571–2582.
[7] Sean Andrist, Tomislav Pejsa, Bilge Mutlu, and Michael Gleicher. 2012. Designing effective gaze mechanisms for virtual agents. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 705–714.
[8] Andreas Aristidou and Joan Lasenby. 2011. FABRIK: A fast, iterative solver for the Inverse Kinematics problem. *Graphical Models* 73, 5 (2011), 243–260.
[9] Emmanuel Ayedoun, Yuki Hayashi, and Kazuhisa Seta. 2019. Adding communicative and affective strategies to an embodied conversational agent to enhance second language learners' willingness to communicate. *International Journal of Artificial Intelligence in Education* 29 (2019), 29–57.
[10] Amos Azaria, Jayant Krishnamurthy, and Tom Mitchell. 2016. Instructable intelligent personal agent. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 30.
[11] Norman Badler, Timothy Bickmore, Catherine Pelachaud, and Alla Safonova. 2010. *Intelligent virtual agents.* Springer.
[12] Ulrike Basten, Kirsten Hilger, and Christian J Fiebach. 2015. Where smart brains are different: A quantitative meta-analysis of functional and structural brain imaging studies on intelligence. *Intelligence* 51 (2015), 10–27.
[13] Peter J Bayley and Larry R Squire. 2003. The medial temporal lobe and declarative memory. In *International Congress Series*, Vol. 1250. Elsevier, 245–259.
[14] Thomas Beelen, Ella Velner, Khiet P Truong, Roeland Ordelman, Theo Huibers, and Vanessa Evers. 2023. Children's trust in robots and the information they provide. In *Extended abstracts of the 2023 CHI conference on human factors in computing systems.* 1–7.
[15] Rachel KE Bellamy, Sean Andrist, Timothy Bickmore, Elizabeth F Churchill, and Thomas Erickson. 2017. Human-agent collaboration: Can an agent be a partner?. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* 1289–1294.
[16] Carl Bereiter and Marlene Scardamalia. 2014. Knowledge building and knowledge creation: One concept, two hills to climb. In *Knowledge creation in education.* Springer, 35–52.
[17] Frank Biocca, Chad Harms, and Jenn Gregg. 2001. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA.* 1–9.
[18] Christopher A Buneo, Murray R Jarvis, Aaron P Batista, and Richard A Andersen. 2002. Direct visuomotor transformations for reaching. *Nature* 416, 6881 (2002), 632–636.
[19] Catherina Burghart, Christian Gaertner, and Heinz Woern. 2006. Cooperative solving of a children's jigsaw puzzle between human and robot: First results. In *Cognitive Robotics: Papers from the AAAI Workshop: Papers from the 2006 AAAI Workshop, ed. M. Beetz, K. Rajan, M. Thielscher, and RB Rusu.* 33–39.

[20] Justine Cassell. 2001. Embodied conversational agents: representation and intelligence in user interfaces. *AI magazine* 22, 4 (2001), 67–67.
[21] Raymond B Cattell. 1963. Theory of fluid and crystallized intelligence: A critical experiment. *Journal of educational psychology* 54, 1 (1963), 1.
[22] Aleksandra Cerekovic, Oya Aran, and Daniel Gatica-Perez. 2016. Rapport with virtual agents: What do human social cues and personality explain? *IEEE Transactions on Affective Computing* 8, 3 (2016), 382–395.
[23] Erin K Chiou, Noah L Schroeder, and Scotty D Craig. 2020. How we trust, perceive, and learn from virtual humans: The influence of voice quality. *Computers & Education* 146 (2020), 103756.
[24] Erin K Chiou, Noah L Schroeder, and Scotty D Craig. 2020. How we trust, perceive, and learn from virtual humans: The influence of voice quality. *Computers & Education* 146 (2020), 103756.
[25] Minsoo Choi, Dixuan Cui, Siqi Guo, Dominic Kao, and Christos Mousas. 2025. Let's Do It My Way: Effects of Personality and Age of Virtual Characters. *IEEE Transactions on Visualization and Computer Graphics* (2025).
[26] Minsoo Choi, Dixuan Cui, Alexandros Koilias, and Christos Mousas. 2024. The Effects of Virtual Character's Intelligence and Task's Complexity during an Immersive Jigsaw Puzzle Co-solving Task. In *Proceedings of the 17th ACM SIGGRAPH Conference on Motion, Interaction, and Games*. 1–12.
[27] Minsoo Choi, Dixuan Cui, Matias Volonte, Alexandros Koilias, Dominic Kao, and Christos Mousas. 2025. Toward Understanding the Effects of Intelligence of a Virtual Character during an Immersive Jigsaw Puzzle Co-Solving Task. *ACM Transactions on Applied Perception* 22, 2 (2025), 1–28.
[28] Minsoo Choi, Siqi Guo, Alexandros Koilias, Matias Volonte, Dominic Kao, and Christos Mousas. 2024. Exploring the Effects of Self-correction Behavior of an Intelligent Virtual Character during a Jigsaw Puzzle Co-solving Task. *ACM Transactions on Interactive Intelligent Systems* 14, 3 (2024), 1–33.
[29] Viviane Clay, Peter König, and Sabine Koenig. 2019. Eye tracking in virtual reality. *Journal of eye movement research* 12, 1 (2019), 10–16910.
[30] Jennifer H Coane, John Cipollini, Talia E Barrett, Joshua Kavaler, and Sharda Umanath. 2023. Lay definitions of intelligence, knowledge, and memory: Inter-and independence of constructs. *Journal of Intelligence* 11, 5 (2023), 84.
[31] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. Routledge.
[32] Roberto Colom, Sherif Karama, Rex E Jung, and Richard J Haier. 2010. Human intelligence and brain networks. *Dialogues in clinical neuroscience* 12, 4 (2010), 489–501.
[33] Sylvain Daronnat. 2020. Human-agent trust relationships in a real-time collaborative game. In *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play*. 18–20.
[34] Ton De Jong. 2010. Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional science* 38, 2 (2010), 105–134.
[35] John Duncan, Moataz Assem, and Sneha Shashidhara. 2020. Integrated intelligence from distributed brain activity. *Trends in cognitive sciences* 24, 10 (2020), 838–852.
[36] Aaron C Elkins and Douglas C Derrick. 2013. The sound of trust: voice as a measurement of trust during interactions with embodied conversational agents. *Group decision and negotiation* 22, 5 (2013), 897–913.
[37] Hans J Eysenck. 2012. *The measurement of intelligence*. Springer Science & Business Media.
[38] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191.
[39] Patrick Fissler, Olivia Caroline Küster, Daria Laptinskaya, Laura Sophia Loy, Christine AF Von Arnim, and Iris-Tatjana Kolassa. 2018. Jigsaw puzzling taps multiple cognitive abilities and is a potential protective factor for cognitive aging. *Frontiers in aging neuroscience* 10 (2018), 408085.
[40] Tiago Fornelos, Pedro Valente, Rafael Ferreira, Diogo Tavares, Diogo Silva, David Semedo, Joao Magalhaes, and Nuno Correia. 2022. A conversational shopping assistant for online virtual stores. In *Proceedings of the 30th ACM International Conference on Multimedia*. 6994–6996.
[41] Alexander Frummet, Alessandro Speggiorin, David Elsweiler, Anton Leuski, and Jeff Dalton. 2024. Cooking with Conversation: Enhancing User Engagement and Learning with a Knowledge-Enhancing Assistant. *ACM Transactions on Information Systems* 42, 5 (2024), 1–29.
[42] Yang Gao, Yangbin Dai, Guangtao Zhang, Honglei Guo, Fariba Mostajeran, Binge Zheng, and Tao Yu. 2025. Trust in Virtual Agents: Exploring the Role of Stylization and Voice. *IEEE Transactions on Visualization and Computer Graphics* (2025).
[43] Irene Lopez Garcia, Ephraim Schott, Marcel Gohsen, Volker Bernhard, Benno Stein, and Bernd Froehlich. 2024. Speaking with objects: Conversational agents' embodiment in virtual museums. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 279–288.
[44] Howard E Gardner. 2011. *Frames of mind: The theory of multiple intelligences*. Basic books.
[45] Spandana Gella, Aishwarya Padmakumar, Patrick Lange, and Dilek Hakkani-Tur. 2022. Dialog acts for task-driven embodied agents. *arXiv preprint arXiv:2209.12953* (2022).
[46] Ella Glikson and Anita Williams Woolley. 2020. Human trust in artificial intelligence: Review of empirical research. *Academy of management annals* 14, 2 (2020), 627–660.
[47] Robert F Goldberg, Charles A Perfetti, and Walter Schneider. 2006. Perceptual knowledge retrieval activates sensory brain regions. *Journal of Neuroscience* 26, 18 (2006), 4917–4921.
[48] Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. 2007. Creating rapport with virtual agents. In *Intelligent Virtual Agents: 7th International Conference, IVA 2007 Paris, France, September 17-19, 2007 Proceedings 7*. Springer, 125–138.
[49] Manuel Guimarães, Rui Prada, Pedro A Santos, João Dias, Arnav Jhala, and Samuel Mascarenhas. 2020. The impact of virtual reality in the social presence of a virtual agent. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. 1–8.
[50] Kunal Gupta, Ryo Hajika, Yun Suen Pai, Andreas Duenser, Martin Lochner, and Mark Billinghurst. 2020. Measuring human trust in a virtual assistant using physiological sensing in virtual reality. In *2020 IEEE Conference on virtual reality and 3D user interfaces (VR)*. IEEE, 756–765.
[51] Shirin Hajahmadi, Pasquale Cascarano, Fariba Mostajeran, Kevin Heuer, Anton Lux, Gil Otis Mends-Cole, Frank Steinicke, and Gustavo Marfia. 2025. Investigating the Impact of Voice-only and Embodied Conversational Virtual Agents on Mixed Reality Puzzle Solving. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 602–612.
[52] Syed Zohaib Hassan, Pegah Salehi, Ragnhild Klingenberg Røed, Pål Halvorsen, Gunn Astrid Baugerud, Miriam Sinkerud Johnson, Pierre Lison, Michael Riegler, Michael E Lamb, Carsten Griwodz, et al. 2022. Towards an AI-driven talking avatar in virtual reality for investigative interviews of children. In *Proceedings of the 2nd Workshop on Games Systems*. 9–15.
[53] César A Hidalgo, Diana Orghian, Jordi Albo Canals, Filipa De Almeida, and Natalia Martin. 2021. *How humans judge machines*. MIT Press.
[54] Chin-Chang Ho and Karl F MacDorman. 2017. Measuring the uncanny valley effect: Refinements to indices for perceived humanness, attractiveness, and eeriness. *International Journal of Social Robotics* 9 (2017), 129–139.
[55] Hoang Phuoc Ho, Vani Ramesh, Ivo Zaloudek, Delaram Javdani Rikhtehgar, and Shenghui Wang. 2025. Enhancing Visitor Engagement in Interactive Art Exhibitions with Visual-Enhanced Conversational Agents. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*. 660–671.
[56] James Jaccard, Tonya Dodge, and Vincent Guilamo-Ramos. 2005. Metacognition, risk behavior, and risk outcomes: the role of perceived intelligence and perceived knowledge. *Health Psychology* 24, 2 (2005), 161.
[57] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. 2000. Foundations for an empirically determined scale of trust in automated systems. *International journal of cognitive ergonomics* 4, 1 (2000), 53–71.
[58] Satoshi Kanazawa. 2011. Intelligence and physical attractiveness. *Intelligence* 39, 1 (2011), 7–14.
[59] Dominic Kao, Alejandra J Magana, and Christos Mousas. 2021. Evaluating tutorial-based instructions for controllers in virtual reality games. *Proceedings of the ACM on human-computer interaction* 5, CHI PLAY (2021), 1–28.
[60] Alexandros Koilias, Michael G Nelson, Christos-Nikolaos Anagnostopoulos, and Christos Mousas. 2020. Immersive walking in a virtual crowd: The effects of the density, speed, and direction of a virtual crowd on human movement behavior. *Computer Animation and Virtual Worlds* 31, 6 (2020), e1928.
[61] Yixuan Li, Pingmei Xu, Dmitry Lagun, and Vidhya Navalpakkam. 2017. Towards measuring and inferring user interest from gaze. In *Proceedings of the 26th International Conference on World Wide Web Companion*. 525–533.
[62] Lizi Liao, Grace Hui Yang, and Chirag Shah. 2023. Proactive conversational agents in the post-chatgpt world. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 3452–3455.
[63] Tze Wei Liew, Su-Mae Tan, and Chandrika Jayothisa. 2013. The effects of peer-like and expert-like pedagogical agents on learners' agent perceptions, task-related attitudes, and learning achievement. *Journal of Educational Technology & Society* 16, 4 (2013), 275–286.
[64] Kate Loveys, Catherine Hiko, Mark Sagar, Xueyuan Zhang, and Elizabeth Broadbent. 2022. "I felt her company": A qualitative study on factors affecting closeness and emotional support seeking with an embodied conversational agent. *International Journal of Human-Computer Studies* 160 (2022), 102771.
[65] Eleanor A Maguire and Christopher D Frith. 2004. The brain network associated with acquiring semantic knowledge. *Neuroimage* 22, 1 (2004), 171–178.
[66] Smaragda Markaki and Costas Panagiotakis. 2023. Jigsaw puzzle solving techniques and applications: a survey. *The Visual Computer* 39, 10 (2023), 4405–4421.
[67] Alex Martin and Linda L Chao. 2001. Semantic memory and the brain: structure and processes. *Current opinion in neurobiology* 11, 2 (2001), 194–201.
[68] Benoît Massé, Silèye Ba, and Radu Horaud. 2017. Tracking gaze and visual focus of attention of people involved in social interaction. *IEEE transactions on pattern analysis and machine intelligence* 40, 11 (2017), 2711–2724.

[69] Earl K Miller and Jonathan D Cohen. 2001. An integrative theory of prefrontal cortex function. *Annual review of neuroscience* 24, 1 (2001), 167–202.

[70] Sara Moussawi and Marios Koufaris. 2019. Perceived intelligence and perceived anthropomorphism of personal intelligent agents: Scale development and validation. (2019).

[71] Sara Moussawi, Marios Koufaris, and Raquel Benbunan-Fich. 2021. How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents. *Electronic Markets* 31, 2 (2021), 343–364.

[72] Shreeya Nelekar, Amal Abdulrahman, Manik Gupta, and Deborah Richards. 2022. Effectiveness of embodied conversational agents for managing academic stress at an Indian University (ARU) during COVID-19. *British Journal of Educational Technology* 53, 3 (2022), 491–511.

[73] Michael G Nelson, Fu-Chia Yang, Alexandros Koilias, Christos-Nikolaos Anagnostopoulos, and Christos Mousas. 2024. Avoiding Virtual Characters: The Effects of Proximity and Gesture. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 41–50.

[74] Eric WT Ngai, Maggie CM Lee, Mei Luo, Patrick SL Chan, and Tenglu Liang. 2021. An intelligent knowledge-based chatbot for customer service. *Electronic Commerce Research and Applications* 50 (2021), 101098.

[75] Nahal Norouzi, Kangsoo Kim, Jason Hochreiter, Myungho Lee, Salam Daher, Gerd Bruder, and Greg Welch. 2018. A systematic survey of 15 years of user studies published in the intelligent virtual agents conference. In *Proceedings of the 18th international conference on intelligent virtual agents*. 17–22.

[76] Magalie Ochs, Daniel Mestre, Grégoire De Montcheuil, Jean-Marie Pergandi, Jorane Saubesty, Evelyne Lombardo, Daniel Francon, and Philippe Blache. 2019. Training doctors' social skills to break bad news: evaluation of the impact of virtual environment displays on the sense of presence. *Journal on Multimodal User Interfaces* 13 (2019), 41–51.

[77] Wanda J Orlikowski. 2006. Material knowing: the scaffolding of human knowledgeability. *European Journal of Information Systems* 15, 5 (2006), 460–466.

[78] Özgür Örün and Yavuz Akbulut. 2019. Effect of multitasking, physical environment and electroencephalography use on cognitive load and retention. *Computers in Human Behavior* 92 (2019), 216–229.

[79] Ye Pan and Anthony Steed. 2016. A comparison of avatar-, video-, and robot-mediated interaction on users' trust in expertise. *Frontiers in Robotics and AI* 3 (2016), 12.

[80] Sai Krishna Pathi, Annica Kristoffersson, Andrey Kiselev, and Amy Loutfi. 2019. F-formations for social interaction in simulation using virtual agents and mobile robotic telepresence systems. *Multimodal Technologies and Interaction* 3, 4 (2019), 69.

[81] Karalyn Patterson, Peter J Nestor, and Timothy T Rogers. 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature reviews neuroscience* 8, 12 (2007), 976–987.

[82] Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. 2019. Language models as knowledge bases? *arXiv preprint arXiv:1909.01066* (2019).

[83] XU Qizhi, Zongting Ge, Mengxiao Zhu, Jing Wang, and Mo Zhang. 2024. Designing an Online Jigsaw Puzzle Task to Assess Collaborative Problem-Solving Skills. In *2024 IEEE Integrated STEM Education Conference (ISEC)*. IEEE, 1–8.

[84] Carmen M Radecki and James Jaccard. 1995. Perceptions of knowledge, actual knowledge, and information search behavior. *Journal of experimental social psychology* 31, 2 (1995), 107–138.

[85] Hedieh Ranjbartabar, Debbie Richards, Cat Kutay, and Samuel Mascarenhas. 2018. Towards an adaptive system: Users' preferences and responses to an intelligent virtual advisor based on individual differences. (2018).

[86] Byron Reeves and Clifford Nass. 1996. The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK* 10, 10 (1996), 19–36.

[87] Minjin Rheu, Ji Youn Shin, Wei Peng, and Jina Huh-Yoo. 2021. Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human–Computer Interaction* 37, 1 (2021), 81–96.

[88] Jeff Rickel and W Lewis Johnson. 2000. Task-oriented collaboration with embodied agents in virtual worlds. *Embodied conversational agents* (2000), 95–122.

[89] Eric L Rolfhus and Phillip L Ackerman. 1999. Assessing individual differences in knowledge: Knowledge, intelligence, and related traits. *Journal of educational psychology* 91, 3 (1999), 511.

[90] Bertrand Russell. 2009. *Human knowledge: Its scope and limits*. Routledge.

[91] Anastasia Ruvimova, Junhyeok Kim, Thomas Fritz, Mark Hancock, and David C Shepherd. 2020. " Transport Me Away": Fostering Flow in Open Offices through Virtual Reality. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–14.

[92] Sujan Sarker, Haley N Green, Mohammad Samin Yasar, and Tariq Iqbal. 2024. Cohrt: A collaboration system for human-robot teamwork. *arXiv preprint arXiv:2410.08504* (2024).

[93] Emanuel A Schegloff, Gail Jefferson, and Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language* 53, 2 (1977), 361–382.

[94] Valentin Schwind, Katrin Wolf, and Niels Henze. 2018. Avoiding the uncanny valley in virtual character design. *interactions* 25, 5 (2018), 45–49.

[95] Claire Scoular, Daniel Duckworth, Jonathan Heard, and Dara Ramalingam. 2020. Collaboration: Definition and structure. (2020).

[96] Joel Sebastian and Deborah Richards. 2017. Changing stigmatizing attitudes to mental health via education and contact with embodied conversational agents. *Computers in Human Behavior* 73 (2017), 479–488.

[97] Stela H Seo, Keelin Griffin, James E Young, Andrea Bunt, Susan Prentice, and Verónica Loureiro-Rodríguez. 2018. Investigating people's rapport building and hindering behaviors when working with a collaborative robot. *International Journal of Social Robotics* 10 (2018), 147–161.

[98] Zijie Shu, Jihong Zhang, and Zhe Li. 2023. Design of Pedagogical Agent Based on Open-Source Large Language Model in Online Learning. In *2023 Twelfth International Conference of Educational Innovation through Technology (EITT)*. IEEE, 71–74.

[99] Mel Slater, Solène Neyret, Tania Johnston, Guillermo Iruretagoyena, Mercè Álvarez de la Campa Crespo, Miquel Alabèrnia-Segura, Bernhard Spanlang, and Guillem Feixas. 2019. An experimental study of a virtual reality counselling paradigm using embodied self-dialogue. *Scientific reports* 9, 1 (2019), 10903.

[100] Charles Spearman. 1961. The abilities of man. (1961).

[101] Philipp Stark, Lisa Hasenbein, Enkelejda Kasneci, and Richard Göllner. 2024. Gaze-based attention network analysis in a virtual reality classroom. *MethodsX* 12 (2024), 102662.

[102] Rainer Stiefelhagen and Jie Zhu. 2002. Head orientation and gaze direction in meetings. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems*. 858–859.

[103] Paweł M Strojny, Natalia Dużmańska-Misiarczyk, Natalia Lipp, and Agnieszka Strojny. 2020. Moderators of social facilitation effect in virtual reality: Co-presence and realism of virtual agents. *Frontiers in psychology* 11 (2020), 1252.

[104] John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive science* 12, 2 (1988), 257–285.

[105] Theodouli Terzidou and Thrasyvoulos Tsiatsos. 2014. The impact of pedagogical agents in 3D collaborative serious games. In *2014 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, 1175–1182.

[106] Matias Volonte, Eyal Ofek, Ken Jakubzak, Shawn Bruner, and Mar Gonzalez-Franco. 2022. Headbox: A facial blendshape animation toolkit for the microsoft rocketbox library. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 39–42.

[107] Zhan Wang, Lin-Ping Yuan, Liangwei Wang, Bingchuan Jiang, and Wei Zeng. 2024. Virtuwander: Enhancing multi-modal interaction for virtual tour guidance through large language models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–20.

[108] Zhan Wang, Lin-Ping Yuan, Liangwei Wang, Bingchuan Jiang, and Wei Zeng. 2024. Virtuwander: Enhancing multi-modal interaction for virtual tour guidance through large language models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–20.

[109] Lynn Waterhouse. 2023. Why multiple intelligences theory is a neuromyth. *Frontiers in psychology* 14 (2023), 1217288.

[110] Paul J Werbos. 2009. Intelligence in the brain: A theory of how it works and how to build it. *Neural Networks* 22, 3 (2009), 200–212.

[111] Fu-Chia Yang, Pedro Acevedo, Siqi Guo, Minsoo Choi, and Christos Mousas. 2025. Embodied Conversational Agents in Extended Reality: A Systematic Review. *IEEE Access* (2025).

[112] Fu-Chia Yang, Kevin Duque, and Christos Mousas. 2024. The Effects of Depth of Knowledge of a Virtual Agent. *IEEE Transactions on Visualization and Computer Graphics* (2024).

[113] Fu-Chia Yang, Siqi Guo, and Christos Mousas. 2025. Exploring familiarity and knowledgeability in conversational virtual agents. *ACM Transactions on Applied Perception* 23, 1 (2025), 1–28.

[114] Habeeb Yusuf, Arthur Money, and Damon Daylamani-Zad. 2025. Pedagogical AI conversational agents in higher education: a conceptual framework and survey of the state of the art. *Educational technology research and development* 73, 2 (2025), 815–874.

[115] Jingling Zhang, Jane Conway, and César A Hidalgo. 2023. Why people judge humans differently from machines: the role of perceived agency and experience. In *2023 14th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, 000159–000166.

[116] Michelle X Zhou, Gloria Mark, Jingyi Li, and Huahai Yang. 2019. Trusting virtual agents: The effect of personality. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 9, 2-3 (2019), 1–36.