Toward Understanding the Effects of Intelligence of a Virtual Character during an Immersive Jigsaw Puzzle Co-Solving Task

MINSOO CHOI, Purdue University, USA
DIXUAN CUI, Sam Houston State University, USA
MATIAS VOLONTE, Clemson University, USA
ALEXANDROS KOILIAS, University of the Aegean and University of the Peloponnese, Greece
DOMINIC KAO, Purdue University, USA
CHRISTOS MOUSAS, Purdue University, USA

In virtual reality, creating intelligent virtual characters has been a long-lasting endeavor. However, while researchers have investigated several aspects of a virtual character's intelligence, little attention has been paid to the impact of the implemented intelligence levels assigned to a virtual character during human-virtual character collaboration. Thus, we conducted a within-group study ($N = 24$) to explore how three different intelligence levels (low vs. medium vs. high) assigned to a virtual character can impact how study participants perceive that virtual character and interact with the task they are instructed to complete. Specifically, for our study, we developed a jigsaw puzzle game and instructed our participants to solve it with the help of a virtual character. During the jigsaw puzzle solving process, we collected application logs related to how the participants executed the task and observed the virtual environment. Moreover, after each condition, we asked the participants to respond using a questionnaire that examined their social presence, how they perceived the character's intelligence and compared it with their own, and how they rated the virtual character's realism. Our results indicated that the different intelligence levels assigned to the virtual characters impacted participants' responses on several variables, including co-presence, perceived intelligence, intelligence comparison, and character interaction and behavior realism. Moreover, based on the collected logged data, we found that the intelligence levels assigned to our virtual character significantly impacted the performance of our participants. Our results could be valuable to the research community for creating more engaging experiences with intelligent virtual characters for collaborative tasks in immersive environments.

CCS Concepts: • **Human-centered computing** → **Virtual reality**; **User studies**; • **Computing methodologies** → *Intelligent agents*; *Perception*.

Additional Key Words and Phrases: virtual character, virtual reality, intelligence, jigsaw puzzle, puzzle co-solving, collaboration

1 INTRODUCTION

Several researchers have suggested that when we immerse humans in a virtual environment, virtual reality blurs the line between real and virtual experiences to the point that the observing the distinction becomes challenging [10, 14]. Similarly, with the recent developments and the transition to an age of artificial intelligence,

Authors' addresses: Minsoo Choi, choi714@purdue.edu, Purdue University, 401 N. Grant St., West Lafayette, Indiana, USA, 47907; Dixuan Cui, dxc164@shsu.edu, Sam Houston State University, 1905 University Ave, Huntsville, Texas, USA, 77341; Matias Volonte, mvolont@clemson.edu, Clemson University, 105 Sikes Hall, Clemson, South Carolina, USA, 29634; Alexandros Koilias, alex.kilias@go.uop.gr, University of the Aegean and University of the Peloponnese, Lofos Xenia, Mytilene, Lesbos, Greece, 81100; Dominic Kao, kaod@purdue.edu, Purdue University, 401 N. Grant St., West Lafayette, Indiana, USA, 47907; Christos Mousas, cmousas@purdue.edu, Purdue University, 401 N. Grant St., West Lafayette, Indiana, USA, 47907.

ACM 1544-3965/2024/11-ART
https://doi.org/10.1145/3700822

advancements in artificial intelligence could also blur the line between real and fake [41, 114]. Considering the merging of such fields, in this paper, we examine the extent to which study participants' perceptions and interaction change when we expose them to a virtual character whose intelligence varies.

Specifically, in virtual reality, a virtual character is considered to be the representation of a human or humanoid that interacts with a particular entity (e.g., a human or other virtual characters) [16, 64], interacts and executes tasks [92], responds to events [28, 59], expresses emotions [84, 107] and intelligent behavior [76, 111], and conveys interpersonal attitudes [19, 25]—all in a human-like way. In general, researchers have proved that when virtual characters communicate information to humans, they are perceived as social agents and can influence human functioning at several cognitive-analytical processing levels [5, 35, 81, 109]. Moreover, previous studies have reported that humans can understand social interactions with virtual characters, and as a result, such social interaction can evoke a sense of social presence [11, 35]. This is especially true when virtual characters are represented anthropomorphically [35, 77, 109].

Fig. 1. First-person view of the main scene of our application. Our system placed the puzzle pieces randomly on the table. The jigsaw puzzle box is on the table's left corner. We used the semitransparent puzzle board to inform our study participants where to place the puzzle pieces. We also see the virtual character that helped study participants to co-solve the jigsaw puzzle.

In the context of this study, the variable under investigation was the intelligence level assigned to a virtual character. We defined it as a virtual character's ability to solve a cognitively demanding task: solving a jigsaw puzzle [34, 47] with the help of a human. In our study, we aimed to explore the relationship between the intelligence levels of a virtual character and participants' perceptions of this character, as well as their interactions with a task in a virtual environment. Specifically, we focused on how participants engaged with the virtual character in co-solving a jigsaw puzzle (refer to Fig. 1). To achieve this, we conducted a within-group study with

24 participants, assessing three distinct intelligence levels assigned to the virtual character: low, medium, and high.

In this project, we programmed a virtual character to assist in solving a jigsaw puzzle, with its actions determined by a user-defined probability of accurately placing pieces. While this programmed behavior may not fully embody traditional intelligence, we contend that it can still be classified as "intelligent." Drawing on Fissler et al. [34], we note that solving jigsaw puzzles requires a suite of cognitive skills, such as visual perception for identifying shapes and patterns, constructional praxis for integrating visual and motor information, mental rotation for piece alignment, and cognitive flexibility for strategy shifting. These tasks also demand cognitive speed, perceptual reasoning for strategy formulation, and both working and episodic memory to maintain associations between puzzle pieces. Given these complexities, we argue that the virtual character's behavior, designed to navigate these cognitive challenges, justifies its consideration as intelligent.

In our study, we used a female virtual character and assigned her different intelligence levels. At the low intelligence level, the virtual character cannot solve the puzzle as the probability of successfully placing a puzzle piece in the right spot is 0% (we consider this to be an unintelligent virtual character). At the medium intelligence level, the virtual character partially solves the puzzle, as the probability of successfully placing a puzzle piece in the right spot is 50%. Lastly, at the high intelligence level, the virtual character solves the puzzle successfully since she always places the puzzle piece in the right spot, as in this condition, the probability of successfully placing a puzzle piece in the right spot is 100%.

Based on the abovementioned conditions, we aimed to understand how different levels of intelligent behavior assigned to a virtual character could impact the study participants' perceptions toward that virtual character. Thus, understanding humans' ability to connect with a virtual character closely makes the evaluation of possible variations of intelligence levels of a virtual character as a function of distinct features of the virtual reality stimuli necessary.

We organized the paper into the following sections. In Section 2, we discuss related work to our project. In Section 3, we present the methodology of our study. In Section 4, we report the results of our study. In Section 5, we discuss our findings, and in Section 6, we discuss our project's limitations. Lastly, in Section 7, we provide our conclusions and potential future directions.

2 RELATED WORK

In the following subsections, we discuss related work to our project.

2.1 Interaction with Virtual Characters

Researchers have examined interaction with virtual characters in immersive environments from various perspectives ranging from simple observations to more advanced social interactions. So far, several researchers have highlighted the potential of virtual characters for delivering engaging and motivating experiences to humans and have also reported attributes of virtual characters that enhance human-virtual character interaction. Among them, we find non-verbal communication and facial expressions [80]; multimodal strategies and appearance [17]; proximity and attentional focus [104]; facial similarity between a user and a virtual character [110]; gaze, gesture, and body orientation [85]; body movements [102]; humor [75]; and smile [40].

A prior study showed that humans are more comfortable interacting with virtual characters that move like humans [101]. Moreover, researchers have indicated that humans would treat virtual characters more favorably if they believed another human controlled them. This extends to noticing positive behavior (e.g., sacrifice, protection) more often [67, 68]. On the contrary, if a human believes a virtual character to be a bot (i.e., a computer-controlled virtual character), they are more likely to blame it [69]. However, an alternate theory is that humans would respond equally to human and computer-controlled entities that exhibit similar social behavior [77], the *computers are social actors* (CASA) theory coined by Nass et al. [73].

2.2 Collaboration in Virtual Environments

In this paper, we deal with a human-intelligent virtual character collaboration scenario, an immersive collaborative jigsaw puzzle solving between study participants and a virtual character. Merriam-Webster's dictionary defines collaboration[1] as *"to work jointly with others or together especially in an intellectual endeavor."* So far, collaboration in virtual environments has been examined from several perspectives, ranging from human-human remote collaboration [23, 57, 82, 83] to collaboration with virtual robots [63]. We also find methods and technologies that allow users to interact and collaborate in shared virtual environments [20] and techniques that enable collaboration and communication in distributed virtual environments [54]. Moreover, researchers have also conducted several studies to understand and improve the sense of presence, trust, embodiment, or behavior during collaboration, especially when users share a common experience [53, 62, 78, 97, 98]. Fribourg et al. [36] conducted a study where their study participants were immersed in a virtual environment and played the well-known whac-a-mole game. They found that their study participants were significantly more "efficient," and accordingly more engaged, in performing the task when sharing the virtual environment, particularly for the more competitive tasks. They also reported that competition and shared experiences involving an avatar did not influence the sense of embodiment but increased user engagement. Greenwald et al. [38] conducted a study with collaboration tasks in a virtual reality telepresence system, finding that participants had better user engagement and a sense of social presence through the shared virtual environment. Other methods also introduced novel ways through avatar-mediated virtual reality systems, such as a socially immersive platform with a life-size projection [87] or a multiuser collaboration supporting system [86, 112].

2.3 Collaboration with Intelligent Virtual Characters

Regarding collaboration through artificial intelligence, Johnson et al. [49] define it as a team consisting of both humans and an artificial intelligence system, which portrays a robotic or virtual agent with a certain degree of autonomy for the computer. Moreover, during the process of the tasks, the intelligent system should be aware of the status of the tasks, predict others' actions, and direct their own behaviors and others. Regarding virtual characters, to effectively work alongside humans and other virtual teammates, they must possess the ability to comprehend the influence of previous occurrences, current situations, and future prospects on team tasks and objectives. Thus, virtual characters must possess adequate task representation and reasoning capabilities [100].

Researchers have examined different aspects of collaboration during human-virtual character interaction. Liu et al. [60] explored human-virtual character collaboration on manual transportation tasks in virtual reality. In this study, users transported a box with the help of a virtual character collaborator. The study's findings showed that participants exhibited lower cognitive load when virtual characters led the tasks. Rickel and Johnson [85] developed Steve, a tutoring system where the user is immersed in a virtual environment and interacts with a virtual character. Steve supports domain-independent capabilities and task-oriented dialogs in the virtual environment. For example, Steve can train people on operating various equipment on a virtual ship and guide them through the boat to show them where the equipment is.

Another application, called Rea, was developed by Cassell et al. [21]. Rea is a synthetic real estate agent in a virtual environment where people can enquire about buying property. The system communicates through speech, intonation, gaze direction, gesture, and facial expression. Lastly, Liu et al. [58] synthesized evacuation training drills in which humans in remote locations collaborate to help intelligent virtual characters escape the

[1]https://www.merriam-webster.com/dictionary/collaboration

building using voice commands and other predefined interaction mechanisms (e.g., opening a path in the virtual environment).

2.4 Interaction and Collaboration with Robots

We also found a few projects where humans deal with robots during puzzle solving. Burghart and Gaertner [18] analyzed the cooperative solving of a jigsaw puzzle between a robot and a human tutor. In this project, the human supports the robot when the tutor evaluates the last action of the robot as unfavorable. Giuliani and Knoll [37] observed how participants interacted in a cooperative construction task with an instructive robot, compared to the interaction with a supportive robot. While the instructive robot first instructs the user how to proceed with the construction and then supports the user by handing over building pieces, the supportive robot adopts a more passive role. It only intervenes when the user is about to make a mistake. The results of their study indicated that humans did not prefer one of the different roles. They instead adapted to the situation by taking the counterpart to the robot's role. Salem et al. [88] considered multiple modalities in their study and installed a robot in a household scenario, assisting a human participant by providing information. The participants had to place some kitchen items in a cupboard while they had to pay attention to the robot's instructions. Their finding suggested that the robot is evaluated more positively when non-verbal behaviors such as hand and arm gestures are displayed along with speech.

Häring et al. [45] explored the use of gaze and pointing gestures in scenarios where study participants had to follow the instructions of a humanoid robot and solve an abstract jigsaw puzzle with the help of the robot instructor in different grounding scenarios with varying difficulty. Their results indicated that adding gaze to the interaction improved the interaction. Ullman et al. [105] explored to what extent the type of agent (human or robot) and the type of behavior (honest or dishonest) affected perceived features of agency and trustworthiness in the context of a competitive game. They predicted that the human and robot in the dishonest manipulation would receive lower attributions of trustworthiness than the human and robot in the honest manipulation. They also perceived the robot as less intelligent and intentional than them. The human and robot in the dishonest manipulation received lower attributions of trustworthiness, and participants perceived it to be smarter than them.

2.5 Perceived Intelligence

Many studies have explored the effects of different design elements on perceived intelligence. However, how we perceive intelligent virtual characters is a pertinent topic in both computer games and virtual reality research [90]. According to Thill et al. [103], perceived intelligence is a human's perception of the agent's intelligence, irrespective of the agent's abilities or the complexity of algorithms. Contrarily, according to other researchers, while perceived intelligence is related to an agent's ability to complete tasks, it is also related to a range of human-agent interaction design elements, such as appearance [24, 55] and interaction modality [30].

Evidence suggests that the perception of the intelligence of virtual characters impacts different dimensions in users. Researchers have shown that the accuracy and errors of their responses affect the perception of virtual humans' intelligence. In the field of medicine, Bickmore et al. [9] conducted research primarily focused on conversational interactions. They discovered that participants viewed a virtual character as less intelligent when it made errors in responses, a perception that persisted even when the errors were attributed to technical difficulties, rather than the virtual character's inherent abilities. Nowak and Biocca [77] stated that when virtual characters exhibit a high degree of cognitive processing, humans may experience a higher degree of social presence or the sense of being in the presence of another person. Additionally, they showed that the perceived intelligence of virtual characters could also impact their level of trustworthiness, as humans perceive more intelligent virtual characters as more trustworthy. Furthermore, Bickmore and Cassell [8] found that virtual characters that exhibit high levels of intelligence and social skills may be better able to establish rapport and build trust with users.

However, how intelligent virtual characters are perceived could impact believability, immersion, and the usefulness of simulators, training tools, and telepresence applications. Evidence suggests that believing a virtual character is human could factor into the enjoyment of the experience, as Merritt et al. [66] indicated that humans prefer teammates they believe to be human-controlled, even if they are actually intelligent virtual characters.

2.6 Contribution

As the technologies that provide life-like characteristics to virtual characters become more accessible and powerful, more advanced interaction scenarios can be developed and studied. For this project, we considered previously conducted research on interaction with intelligent virtual characters and developed a method that controls the intelligence levels of a virtual character when scripted to solve a jigsaw puzzle with study participants collaboratively. We did so because we wanted to study how intelligence levels assigned to that virtual character could impact the perceptions of participants toward that virtual character as well as how different intelligence levels could impact the task execution process. To our knowledge, this is the first study examining how intelligence levels could impact participants' perceptions of a virtual character while collaboratively solving a jigsaw puzzle. Although collaborative jigsaw puzzle solving could not be considered the most demanding task, we think this work could serve as the basis for further research (e.g., collaboratively solving math problems with an intelligent virtual character).

2.7 Research Hypotheses

In this study, we formulated six overarching hypotheses to explore the influence of intelligence levels on various aspects of participants' interactions with a virtual character. Our hypotheses were derived from existing literature. Considering that prior research has shown that a virtual character's intelligence can significantly impact users' social presence [77], we hypothesized that higher intelligence levels would increase co-presence, attentional allocation, and behavioral independence ratings (**H1.1**, **H1.2**, and **H1.3**). Additionally, considering that researchers indicated that the ability of a virtual character to complete tasks efficiently enhances perceived intelligence and comparison ratings [60, 103], we hypothesized that intelligence levels would positively influence participants' perceived intelligence and intelligence comparison ratings (**H2.1** and **H2.2**). Furthermore, the literature suggests that interaction and behavior realism are closely tied to a character's intelligence [21, 58]. Thus, we hypothesized that higher intelligence levels would enhance interaction, appearance, behavior, and movement realism ratings (**H3.1**, **H3.2**, **H3.3**, **H3.4**). Moreover, considering the impact of intelligence on attentional focus [85] and how participants allocate their gaze [77], we hypothesized that intelligence levels would influence participants' dwell gaze at the virtual character's body and puzzle-related objects (**H4.1**, **H4.2**, and **H4.3**). Finally, considering that intelligent systems improve task performance and user engagement [60, 66], we hypothesized that intelligence levels would influence the number of corrections made, the number of puzzle piece pickups, the completion time of the jigsaw puzzle, and the perceived difficulty of manipulating the puzzle pieces (**H5.1**, **H5.2**, **H5.3**, and **H6**). Our hypotheses were the following:

- **H1.1:** Intelligence levels will influence study participants' co-presence ratings.
- **H1.2:** Intelligence levels will influence study participants' attentional allocation ratings.
- **H1.3:** Intelligence levels will influence study participants' perceived behavioral independence ratings.
- **H2.1:** Intelligence levels will influence study participants' perceived intelligence ratings.
- **H2.2:** Intelligence levels will influence study participants' intelligence comparison ratings.
- **H3.1:** Intelligence levels will influence study participants' interaction realism ratings toward the virtual character.

- **H3.2:** Intelligence levels will not influence study participants' appearance realism ratings.
- **H3.3:** Intelligence levels will influence study participants' behavior realism ratings.
- **H3.4:** Intelligence levels will not influence study participants' movement realism ratings.
- **H4.1:** Intelligence levels will influence study participants' dwell gaze at the virtual character's body.
- **H4.2:** Intelligence levels will not influence study participants' dwell gaze at the jigsaw puzzle pieces.
- **H4.3:** Intelligence levels will not influence study participants' dwell gaze at the jigsaw puzzle box.
- **H5.1:** Intelligence levels will influence the number of corrections made by the study participants.
- **H5.2:** Intelligence levels will influence the number of puzzle piece pickups made by the study participants.
- **H5.3:** Intelligence levels will influence the completion time of the jigsaw puzzle.
- **H6:** Intelligence levels will not influence study participants' ratings on the difficulty of manipulating the puzzle pieces.

3 MATERIALS AND METHODS

This section presents the methodology and details of this study.

3.1 Participants

We conducted an *a priori* power analysis to determine the sample size for our study using the G*Power v. 3.1 software [27]. Based on one group with three repeated measures, a medium effect size $f = .30$ [32], a non-sphericity correction $\epsilon = .80$, and an $\alpha = .05$, to achieve an 80% power ($1 - \beta$ error probability), the calculation recommended 23 participants. We recruited participants through class announcements, posters placed in different departments across our campus, and emails sent to our university's students' listservs. In this study, the participant group comprised 24 students (age: $M = 21.79$, $SD = 2.76$), all undergraduate and graduate students at a Midwest U.S. university. Of the sample, six were female (age: $M = 21.66$, $SD = 2.58$), and 18 were male (age: $M = 21.83$, $SD = 2.89$). Moreover, all of our participants had experienced virtual reality at least once before participating in our study. All participants were volunteers and there was no type of compensation involved. Lastly, note that all participants completed the study.

3.2 Virtual Realty Application

We designed a semi-realistic living room where we placed the participants and all interactions occurred. The living room was lit with sunlight coming from the windows and light sources in the living room (see Fig. 2). Both the participant and the virtual character were seated at a table. Following an L-shaped formation based on the F-formations social interaction model [79], we placed the virtual character on a chair on the participant's right side. We downloaded the virtual character from Microsoft's Rocketbox library [108]. On the table, we placed all puzzle pieces in random positions (the initial position of all puzzle pieces was the same for each condition), a semitransparent board of the jigsaw puzzle that the participants and the virtual character should solve, and the box of the puzzle showing the completed image. We used the semitransparent board to indicate where exactly our participants should place each puzzle piece and the puzzle box to provide our participants with a clearer image of the puzzle.

We designed a puzzle with 25 pieces—the size of each puzzle piece was 4×4 cm. During our application's development and testing process, which our lab's members conducted to explore our application's flows, bugs, and other essential issues, we realized that fewer pieces would make the puzzle easier, and more pieces would frustrate participants. Therefore, we considered 25 pieces to be the optimal number for our study. During the application's runtime, a participant can collaborate to solve the jigsaw puzzle with a virtual character whose intelligence level varies. The participant can pick up and place the puzzle pieces using the virtual reality controllers. Note that our application detects whether the participant puts the puzzle pieces in the appropriate position through collision

Fig. 2. We designed and used a living room virtual environment in our study: (left) top view, (middle) front view, and (right) back view.

events between the puzzle pieces and targets on the puzzle board. We decided not to represent our participants with a self-avatar, since we wanted them not to be distracted by a virtual body they might perceive as not their own. We think that this decision helped them focus on the task.

We integrated a short talk at the beginning, middle, and toward the end of the jigsaw puzzle co-solving process between the participant and the virtual character to create a realistic and more engaging experience [33] as well as to help build trust and rapport [7, 48]. For the small talk, we assigned speech and lip sync to the virtual character and allowed the participant to choose an answer from a pop-up graphical user interface using the virtual reality controller (see Table 1 for the small talk we created and Fig. 3 for an in-app example). We used the Salsa LipSync Suite[2] from Unity Asset Store for the lip sync. In addition, we generated the speech assigned to the virtual character using Microsoft's Azure text-to-speech[3] using the Jenny actor (female adult from the United States voice model) with an exciting speaking style. Lastly, during the puzzle co-solving process, we assigned humming to the virtual character in random time steps between picking up and placing puzzle pieces to indicate that the character was "thinking" about her decision on where to place the puzzle piece.

Table 1. The talk we implemented in our application.

#	Time Step	Character Dialog	User Options	Result
T1	Once application started	Hi, my name is Jenny. I am happy to meet you. How was your day?	1. It was good! 2. A little bit bored.	Continue Continue
T2	After T1	Okay, let's play this puzzle together!	1. Good, let's do it! 2. No, I don't want to play.	Continue Application stops
T3	After solving 50% of the puzzle	We are halfway done! Let's cheer up!	1. Nice! Go for it! 2. I don't want to play anymore.	Continue Application stops
T4	After solving 100% of the puzzle	Good job! We did it!	1. Good job! 2. Bye.	Application stops Application stops

We developed a virtual reality application in the Unity games engine (version 2020.3.20) using the Oculus Integration Toolkit. Both for our implementation and study, we used a Dell Alienware Aurora R7 desktop computer (Intel Core i7, NVIDIA GeForce RTX 2080, 32GB RAM). Also, we used Meta's Quest 1 as our virtual reality head-mounted display.

3.2.1 Character Intelligence . Our virtual character can collaboratively solve the jigsaw puzzle with the participant based on the different intelligence levels assigned to her. Specifically, our algorithm drives the virtual character to make the right or wrong decisions on where to place the puzzle pieces. The virtual character in our application solves the jigsaw puzzles based on a loop-based function called CharacterAI (see Algorithm 1 in Appendix A).

[2]https://assetstore.unity.com/packages/tools/animation/salsa-lipsync-suite-148442
[3]https://azure.microsoft.com/en-us/products/cognitive-services/text-to-speech

Fig. 3. An example graphical user interface of the small talk, showing the choices for an answer.

This function has six inputs: R, A, P, I, V_c, and S_C. R is a list including puzzle targets not yet solved, A is a list of puzzle pieces that the virtual character can pick up, P is a list including pairs of puzzle pieces and their answers, I is a variable showing the intelligence of the virtual character with a range between 0% (not intelligent at all; always places the picked puzzle piece in a wrong target spot) and 100% (highly intelligent; always places the puzzle pieces in the right target spot), V_c is a variable showing the current puzzle piece interacted with by the virtual character, and S_C is a variable indicating the current state of the virtual character, which can be either $PickUp$, $Place$, or $Wait$.

During the application's runtime, the actions of the participant or the virtual character update R and A. Also, the outputs of our algorithm, V_N and S_N, update S_C and V_c; however, P and I remain constant. Our algorithm also has two local variables: T and D. T is a variable showing the targeted puzzle piece of the current state, and D is a variable to determine whether the virtual character should place the puzzle piece in the right or wrong target spot.

After the participant answers the first two questions in the small talk (see T1 and T2 in Table 1), our virtual reality application sets S_C to $PickUp$. Then, the system calls the CharacterAI function to let the virtual character solve the jigsaw puzzle. When our system sets S_C to $PickUp$, there are two cases. If there is at least one available puzzle piece, the virtual character finds a puzzle piece T from A and picks it up. Then, V_N is set to T (the picked puzzle piece) and S_N is set to $Place$. However, if no puzzle pieces are available, S_N is set to $Wait$, and the virtual character waits until more puzzle pieces are available to pick up.

When S_N is set to $Place$, our algorithm chooses a random value D (from 0 to 100). If the D is less than I, our algorithm drives the virtual character to place V_c in the right spot; else, the virtual character places V_c in the wrong spot. Regarding the right decision, our algorithm finds the answer T from P by using V_c. However, when the virtual character has to make a wrong decision, our algorithm finds the wrong answer T from R. After the virtual character places V_c in any spot, V_N is set to $NULL$ and S_N is set to $PickUp$. This function iterates until the participant or the virtual character places all the puzzle pieces in the right target spot.

3.2.2 Animation System. The algorithm mentioned above decides how the virtual character should interact with the puzzle pieces. Based on this decision, our animation system animates the virtual character using motion data and inverse kinematics. We assigned a sitting idle motion to our virtual character, which we downloaded from Mixamo. We also implemented a full-body forward and backward inverse kinematics solver [3] to allow the character to perform a reaching task to grab a puzzle piece from the table and leave it on a target spot on the puzzle board that we placed on the table. We assign targets for the right hand and gaze of the virtual character so the gazing of the virtual character would be associated with the reaching task. The targets include puzzle pieces and target spots on the puzzle board on the table so the virtual character can move the right hand to puzzle pieces or target spots while gazing at the chosen targets (see Fig. 4). Also, we assigned eye blinking and head idle motion to make our character look as realistic as possible. Lastly, we developed a predefined delay time between the picking and placing activity to make our virtual character look like she is thinking. In our case, our virtual character waits for two seconds before picking up a puzzle piece and three seconds before placing the puzzle piece.

Fig. 4. The picking and placing activity of a puzzle piece was performed by our virtual character and controlled by our algorithm.

3.3 Experimental Conditions

We developed three experimental conditions to explore how different intelligence levels assigned to our virtual character could impact how study participants perceive and interact with that character, as well as how they interact and accomplish the task we asked them to complete. Specifically, we developed and examined three experimental conditions:

- **Low Intelligence (LI):** This condition represents an unintelligent virtual character. Specifically, in this experimental condition, we assigned the virtual character a 0% probability of correctly placing any puzzle piece on the puzzle board. This means the character will always pick and place a puzzle piece in the wrong spot. We assume this virtual character to be a non-intelligent one.
- **Medium Intelligence (MI):** In this condition, we assigned a 50% probability to our character to correctly place a puzzle piece on the puzzle board. Our algorithm tracks the ratio of wrong puzzle pieces placed on the board so far and decides whether the virtual character should place a new puzzle piece in the right or wrong spot. We assume this virtual character to be moderately intelligent.
- **High Intelligence (HI):** In this condition, we assigned the virtual character a 100% probability of placing a puzzle piece in the right spot on the puzzle board. This means that the virtual character never makes mistakes. Thus, we assume this virtual character to be highly intelligent.

We ensured the balance for first-order carry-over effects across the experimental conditions using Latin squares [113]. We used the same puzzle across all three conditions to standardize our experiment. Note that although there are multiple intervals between 0% and 100% intelligence that we could have explored, we limited our study to two extremes (i.e., 0% and 100% intelligence) and a moderate value (i.e., 50% intelligence) since more conditions could have made our participants lose interest and get bored.

3.4 Measurements

To understand how the intelligence levels assigned to the virtual character impacted our participants across the three experimental conditions, we collected both subjective (i.e., questionnaire responses) and objective (i.e., application logs) data. We describe the collected data in the following subsections.

3.4.1 Questionnaires. We developed a questionnaire to collect responses from participants to investigate how they interacted with the three intelligence levels we assigned to the virtual character. The developed questionnaire (see Appendix B) included 30 items. Specifically, the questionnaire included six items to measure co-presence, six for attentional allocation, six for perceived behavioral independence—all 18 items from Biocca et al. [12]—and six for perceived intelligence from Moussawi and Koufaris [72]. We asked the participants to indicate if the virtual character was more intelligent than them and to specify the virtual character's interaction, appearance, behavior, and movement realism. We also asked them to rate the difficulty of manipulating the puzzle pieces. The authors of this paper developed the questions, using a seven-point Likert scale for the questionnaire responses. Lastly, we asked the participants an open-ended question to provide comments about their experience and the application in a designated space. We distributed the questionnaire to the participants in a computer-based survey using the Qualtrics online survey tool.

3.4.2 Application Logs. We collected data from the virtual reality application to further understand how our participants interacted with the virtual character and the tasks we instructed them to accomplish. Specifically, we collected the following data:

- **Virtual Character Dwell Gazing:** We collected how much dwell time a participant spent gazing at the virtual character's body (including the upper body trunk, face, and both arms) over the total time in the virtual environment.
- **Puzzle Box Dwell Gazing:** We collected how much dwell time a participant spent gazing at the puzzle box over the total time in the virtual environment.
- **Puzzle Pieces Dwell Gazing:** We collected how much dwell time a participant spent gazing at the puzzle pieces over the total time in the virtual environment.
- **Corrections:** We collected how many times participants corrected puzzle pieces that the virtual character placed in the wrong spots on the puzzle board.
- **Puzzle Pieces Pickups:** We counted how many times participants picked up puzzle pieces.
- **Completion Time:** We collected how much time participants spent finishing the puzzle, measured in seconds.

For the gazing data, we did not use an eye-tracking solution; instead, we implemented a pseudogazing method that returns the time that the forward vector of the virtual reality headset collides with either the virtual character (we placed colliders on the upper body trunk, face, and both arms) or the puzzle's box and puzzle pieces through raycasting. Raycasting[4] detects and returns if a collision occurs with the colliders assigned to all objects of interest or the virtual character. If there was a collision, we set a counter to collect the duration of this collision in milliseconds. All gazing measurements were calculated as percentage time spent (i.e., total time collected gazing at an item over total time in the virtual environment).

[4]https://docs.unity3d.com/ScriptReference/Physics.Raycast.html

3.5 Procedure

The participants scheduled a day and time with the research team in advance. Once they arrived at the lab space where the study was taking place, the researchers provided information about the experimental procedure. The researchers provided a consent form for the participants to read and sign upon agreement. The Institutional Review Board of our university approved the consent form. The researchers then asked the participants to provide demographic data in the Qualtrics survey tool. Then, the researchers helped the participants put on the head-mounted display.

Before beginning the experiment, we immersed our participants in the virtual environment used for the main study, but no virtual character was present at this point. We placed a simple jigsaw puzzle of four pieces on the table and asked the participants to solve it. We did so because we wanted to introduce the controllers to the participants and ensure they knew how the jigsaw puzzle interaction worked. A previous study showed that such tutorials about virtual reality controllers could improve participants' performance and user experience [50]. Once our participants indicated that they were ready for the study, the research team told them that in the new scene they would sit in front of a table in a virtual living room where a virtual character would be sitting next to them. The research team also informed the participants that the task they had to perform was to solve the jigsaw puzzle on the table. We did not provide further information to our participants about the environment, the virtual character and its intelligence, or the task. We show our experimental setup in our lab in Fig. 5.

Fig. 5. A participant in our lab during the study.

The research team also informed the participants that once they finished the puzzle, an on-screen instruction would tell them to remove the headset and continue with the next part of the study, which was to fill out the provided questionnaire. After completing the questionnaire, the research team told the participants they would proceed with the subsequent condition of the experiment. We repeated this process for each of the three conditions of our experiment.

The research team informed the participants that they could have additional breaks between each condition and had full permission to terminate the study at any time. After the end of the study, the research team was willing to answer questions about the experiment. At that time, the research team also asked the participants to express their thoughts about the experience and provide additional feedback regarding their interaction with the virtual character. None of the participants quit the study, and none spent more than 60 minutes in the lab to complete the study.

4 RESULTS

For our statistical analyses, we used the three experimental conditions as independent variables and the self-reported ratings and logged data as dependent variables. We examined the normality of our data graphically using Q-Q plots of the residuals and with the Shapiro-Wilk test at the 5% level. We used one-way repeated measures analysis of variance (RM-ANOVA) with post hoc Bonferroni-corrected estimates for the normally distributed data (most of our dependent variables). We used Friedman's test for data that violated the normality criteria (**intelligence comparison**, **puzzle piece manipulation**, and **puzzle box dwell gazing** measurements) and the Wilcoxon signed-rank tests with Bonferroni corrections for the post hoc analysis. Lastly, for the **corrections** measurement, since the participants did not correct the virtual character at all during the high intelligence condition (the character performed the task with 100% correct choices), we compared the data between the low and medium conditions using a paired sample t-test. We provide boxplots for the self-reported ratings in Fig. 6 and the application's logged data in Fig. 7.

4.1 Self-Reported Data

Co-Presence: We found a statistically significant result (Wilks' $\Lambda = .682$, $F[2,22] = 5.136$, $p = .015$, $\eta_p^2 = .318$) across the three conditions (see Fig. 6a). The post hoc pairwise comparison indicated that participants in the HI ($M = 5.80$, $SD = 1.07$) condition rated their co-presence higher than in the LI ($M = 5.29$, $SD = 1.07$) condition at $p = .018$.

Attentional Allocation: We did not find a statistically significant result (Wilks' $\Lambda = .896$, $F[2,22] = 1.281$, $p = .298$, $\eta_p^2 = .104$) on participants' attentional allocation ratings across the three conditions (see Fig. 6b).

Behavioral Independence: We did not find a statistically significant result (Wilks' $\Lambda = .835$, $F[2,22] = 2.174$, $p = .138$, $\eta_p^2 = .165$) on participants' behavioral independence ratings across the three conditions (see Fig. 6c).

Perceived Intelligence: We found a statistically significant result (Wilks' $\Lambda = .255$, $F[2,22] = 32.176$, $p < .001$, $\eta_p^2 = .745$) across the three conditions (see Fig. 6d). The post hoc pairwise comparison indicated that participants in the HI ($M = 5.81$, $SD = 1.00$) condition rated the perceived intelligence higher than in the MI ($M = 4.07$, $SD = 1.35$) condition at $p < .001$ and in the LI ($M = 3.06$, $SD = 1.36$) condition at $p < .001$. Moreover, participants in the MI condition rated the perceived intelligence higher than in the LI condition at $p = .01$.

Intelligence Comparison: We found a statistically significant result ($\chi^2[2] = 31.460$, $p < .001$) across the three examined conditions (see Fig. 6e). The post hoc analysis indicated that participants rated the intelligence of the virtual character higher in the HI ($M = 3.41$, $SD = 1.90$) condition than in the MI ($M = 2.08$, $SD = 1.41$) condition ($Z = -3.673$, $p < .001$) and in the LI ($M = 1.29$, $SD = .75$) condition ($Z = -3.749$, $p < .001$). Moreover,

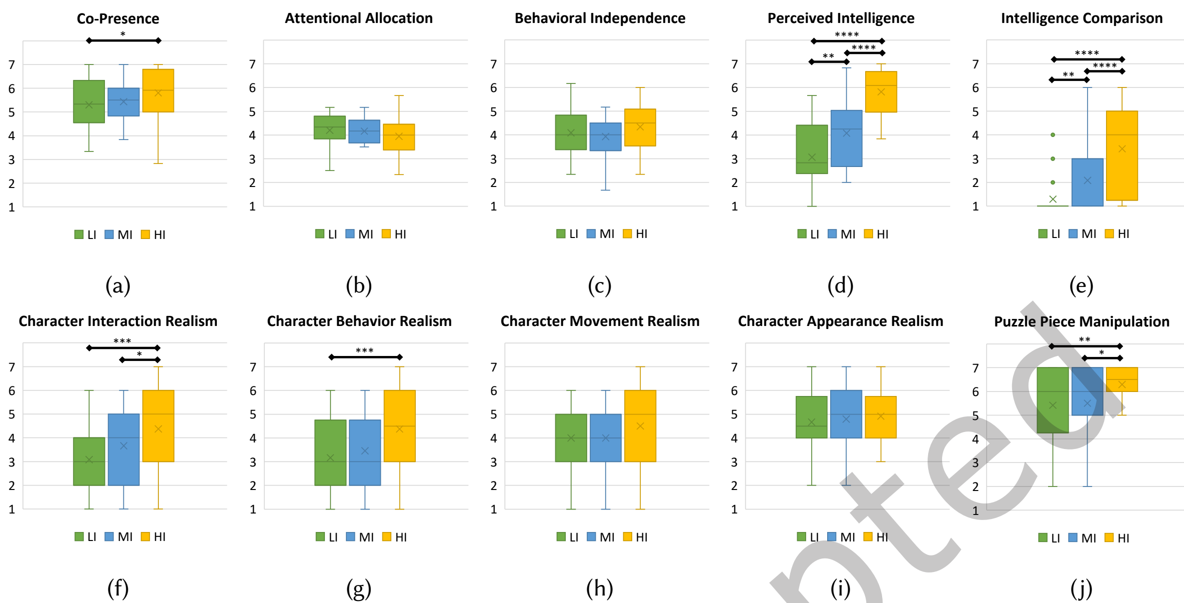

Fig. 6. Boxplots of self-reported data. Boxes enclose the middle 50% of the data. The median is denoted by a thick horizontal line. For statistically significant results: * $p < .05$, ** $p < .01$, *** $p < .005$, and **** $p < .001$.

participants rated the intelligence of the virtual character higher in the MI condition than in the LI condition ($Z = -2.694$, $p = .007$).

Character Interaction Realism: We found a statistically significant result (Wilks' $\Lambda = .606$, $F[2, 22] = 7.166$, $p = .007$, $\eta_p^2 = .367$) across the three examined conditions (see Fig. 6f). The post hoc pairwise comparison indicated that participants in the HI ($M = 4.37$, $SD = 1.68$) condition rated the interaction realism with the virtual character higher than in the MI ($M = 3.66$, $SD = 1.49$) condition at $p = .02$ and in the LI ($M = 3.08$, $SD = 1.66$) condition at $p = .004$.

Character Behavior Realism: We found a statistically significant result (Wilks' $\Lambda = .633$, $F[2, 22] = 6.382$, $p = .004$, $\eta_p^2 = .394$) across the three examined conditions (see Fig. 6g). The post hoc pairwise comparison indicated that participants in the HI ($M = 4.37$, $SD = 1.66$) condition rated the realism of the virtual character's behavior higher than in the LI ($M = 3.16$, $SD = 1.63$) condition at $p = .004$.

Character Movement Realism: We did not find a statistically significant result on how participants rated the realism of the virtual character's movement (Wilks' $\Lambda = .839$, $F[2, 22] = 2.106$, $p = .146$, $\eta_p^2 = .161$) across the three examined conditions (see Fig. 6h).

Character Appearance Realism: We did not find a statistically significant result on how participants rated the realism of the virtual character's appearance (Wilks' $\Lambda = .875$, $F[2, 22] = 1.571$, $p = .230$, $\eta_p^2 = .125$) across the three examined conditions (see Fig. 6i).

Puzzle Piece Manipulation: We found a statistically significant result ($\chi^2[2] = 12.864$, $p = .002$) on participants' puzzle piece manipulation ratings across the three examined conditions (see Fig. 6j). The post hoc analysis indicated that participants rated it easier to manipulate a puzzle piece in the HI ($M = 6.29$, $SD = .80$) condition than in the MI ($M = 5.50$, $SD = 1.47$) condition ($Z = -2.451$, $p = .014$) and the LI ($M = 5.41$, $SD = 1.58$) condition ($Z = -2.692$, $p = .007$).

4.2 Logged Data

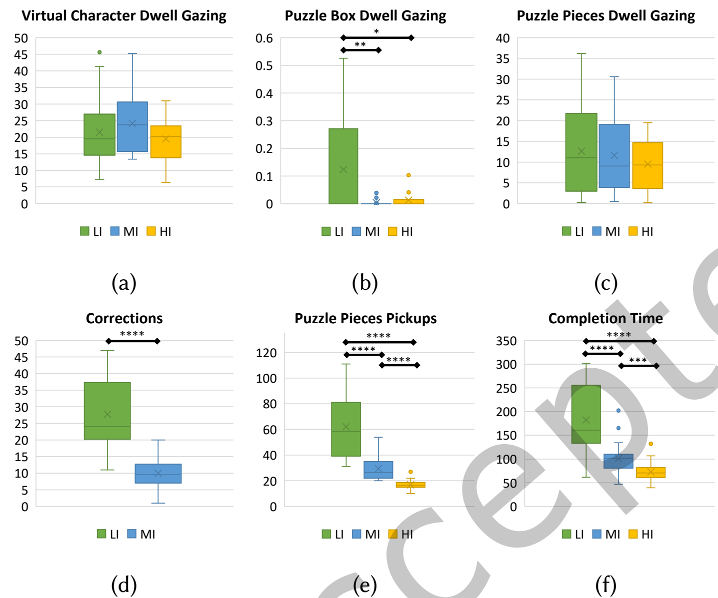

Fig. 7. Boxplots of the application's logged data. Boxes enclose the middle 50% of the data. The median is denoted by a thick horizontal line. For statistically significant results: $^{*}\ p < .05$, $^{**}\ p < .01$, $^{***}\ p < .005$, and $^{****}\ p < .001$.

Virtual Character Dwell Gazing: We did not find a statistically significant result on how participants gazed at the virtual character (Wilks' $\Lambda = .784$, $F[2, 22] = 3.036$, $p = .069$, $\eta_p^2 = .216$) across the three examined conditions (see Fig. 7a).

Puzzle Box Dwell Gazing: We found a statistically significant result ($\chi^2[2] = 9.640$, $p = .008$) across the three examined conditions (see Fig. 7b). The post hoc analysis indicated that participants gazed for more time at the puzzle box in the LI ($M = .12$, $SD = .18$) condition than in the MI ($M = .003$, $SD = .009$) condition ($Z = -2.756$, $p = .006$) and the HI ($M = .01$, $SD = .02$) condition ($Z = -2.552$, $p = .011$).

Puzzle Pieces Dwell Gazing: We did not find a statistically significant result (Wilks' $\Lambda = .870$, $F[2, 22] = 1.636$, $p = .217$, $\eta_p^2 = .130$) on how participants gazed at the puzzle pieces across the three examined conditions (see Fig. 7c).

Corrections: We found a statistically significant result ($t[23] = 9.183$, $p < .001$) in the corrections made by our participants between the LI ($M = 27.70$, $SD = 9.83$) condition and the MI ($M = 9.95$, $SD = 4.73$) condition (see Fig. 7d).

Puzzle Pieces Pickups: We found a statistically significant result (Wilks' $\Lambda = .199$, $F[2, 22] = 33.314$, $p < .001$, $\eta_p^2 = .801$) across the three examined conditions (see Fig. 7e). The post hoc pairwise comparison indicated that participants in the LI (M = 62.08, SD = 24.05) condition picked more puzzle pieces than in the MI (M = 29.45, SD = 8.82) condition at $p < .001$ and in the HI (M = 16.87, SD = 4.48) condition at $p < .001$. Moreover, participants in the MI condition picked more puzzle pieces than in the LI condition at $p < .001$.

Completion Time: We found a statistically significant result (Wilks' $\Lambda = .215$, $F[2, 22] = 40.065$, $p < .001$, $\eta_p^2 = .785$) across the three examined conditions (see Fig. 7f). The post hoc pairwise comparison indicated that participants in the LI (M = 182.25, SD = 69.33) condition spent more time completing the puzzle than in the MI (M = 101.19, SD = 32.20) condition at $p < .001$ and in the HI (M = 72.57, SD = 19.93) condition at $p < .001$. Moreover, participants in the MI condition spent more time completing the puzzle than in the LI condition at $p = .001$.

4.3 Qualitative Data

In addition to the quantitative data, we gathered comments from our participants about anything they liked or disliked and suggestions about the intelligence levels of the virtual character. Nine participants commented about the different intelligence levels assigned to our virtual character. Moreover, nine participants stated that it was a bit challenging to interact with the virtual character during the low intelligence condition. Among them, five participants stated that, in most cases, they enjoyed the high intelligence condition more.

> P7 said, *"The second condition [low intelligence condition] made me feel confused about solving the puzzle because I had to monitor the performance of the virtual character as she kept on putting the puzzle pieces in the wrong places, which confused me about the whole puzzle itself. The first [high intelligence] and last one [medium intelligence] were good in terms of collaboration, and I felt like I was working with another person in the environment."*
>
> P8 reported, *"During the first test [low intelligence condition], there was definitely a disconnect between me and her because we seemed to be trying to accomplish completely different goals. It seemed to me that the more intelligently the other person solving the puzzle acted, the more realistic she seemed."*
>
> P15 said, *"On tests one [medium intelligence condition] and three [low intelligence condition], the individual continued to block my view of pieces and would place them in the wrong locations, making the puzzle more difficult to complete. In the second test [high intelligence condition], the individual and I worked together, where I did not have to continually assist the other individual in all of their actions, and it was more natural as a result."*
>
> P16 reported, *"Throughout the trials, I noticed that I was getting more frustrated with the other person doing the puzzle with me because I had to constantly keep an eye out for where she would place the pieces. At first, I thought she was doing well because she was putting pieces down faster, but it took me some time to realize she was just putting pieces down wherever, and I had to be more mindful of what she was doing."*
>
> P20 mentioned, *"The first condition [low intelligence condition], it was clear to me that the NPC [i.e., non-player character] was doing things without "thinking" it through, as if it was trying to intentionally sabotage our best time to solve the puzzle."*
>
> P22 said, *"She also really frustrated me in the last round [low intelligence condition], where she kept picking up the pieces and putting them in the wrong location."*
>
> Lastly, several of our participants mentioned they enjoyed co-solving the jigsaw puzzle with the highly intelligent virtual character. Specifically, P1 said, *"I enjoyed the last one [high intelligence condition] the most because the virtual character actually responded to my actions."* P8 reported, *"I enjoyed this more capable virtual character [high intelligence condition]."* Finally, P15 wrote, *"I*

would like to try again but only with the most intelligent character. This condition [high intelligence condition] made me enjoy the gameplay more."

5 DISCUSSION

For this project, we collected both questionnaire responses and application logs to understand how our study participants perceived the different levels of intelligence assigned to the virtual character and how the three intelligence levels impacted how they executed the jigsaw puzzle solving. The statistical analyses revealed several interesting results, which we discuss in the following subsections.

5.1 Social Presence

The most unexpected result was the one found in the **co-presence** variable. Our participants provided significantly higher ratings when we exposed them to the high intelligence condition than the low intelligence one. This finding indicates that interacting with a highly intelligent and capable virtual character to solve a given task enhanced the participants' perception of being with another individual in that virtual environment. We know that co-presence depends on having another individual to connect with through a communication medium [77] and different behaviors that show cognitive capabilities [31]. For example, establishing and maintaining eye contact between people involved in interactions is a crucial factor causing "social engagement" [1, 15].

Our results prove that intelligence assigned to a virtual character could impact the sense of co-presence (**H1.1** is supported). However, based on our findings, we see that moderate intelligence was not enough to provide significant results against the low intelligence level, meaning that altering the sense of co-presence through an intelligent collaborator is not because of "some intelligence" assigned to that virtual individual. Thus, we argue that only a high level of intelligence and, therefore, cognitive capabilities could increase the sense of co-presence in virtual environments.

The other two dimensions of social presence we examined did not reveal significant results across the three conditions (**H1.2** and **H1.3** are not supported). On the one hand, although a prior study reported differences in **attentional allocation** during interaction with an intelligent virtual agent [29], to our knowledge, attentional allocation differences were more common when study participants observed only the face of virtual characters [89]. Thus, on the basis of non-significant results, we argue that the intelligence levels assigned to the virtual character were not enough to make our participants rate their attentional allocation differently. On the other hand, we say that we were partially anticipating the **behavioral independence** results obtained from our study.

In all three examined conditions, we scripted our virtual character to perform a task independently without considering what the study participants do. Thus, although the study participants' behavior was in direct response to the virtual character's behavior (especially in the low and medium intelligence conditions, in which the participants were trying to correct the mistakes of the virtual character), our participants noticed that the character's behavior was not in direct response to them. A few of our participants also mentioned this in the comments they provided. This limited bidirectional dependence across all three conditions made our participants not provide significantly different ratings across the three intelligence levels.

We argue that the intelligence assigned to a virtual character is not enough in itself to make participants think they work together toward a common task. There is a need for that virtual character to directly respond to the participant's actions or perform a behavior closer to the one that the participants' performed (e.g., in low and medium intelligence conditions, the virtual character could also correct the mistakes) to be able to enhance such perception. Our finding agrees with Johnson et al.'s [49] description of intelligent systems (i.e., intelligent systems should be aware of the status of the tasks, predict others' actions, and direct their own behaviors) and extends it to intelligent virtual characters.

5.2 Perceived Intelligence and Intelligence Comparison

Regarding **perceived intelligence**, we know that researchers have explored the effects of different design elements assigned to virtual agents to make them "intelligent" and how such elements affect perceived intelligence [115]. For example, previously conducted studies concluded that both anthropomorphic appearance [46] and interaction modality [56] could impact study participants' perceptions of intelligence. Our study extends such findings (**H2.1** is supported) by reporting that perceived intelligence is related to a character's ability to complete a given task efficiently (i.e., the more efficiently a virtual character completes a task, the more intelligent the virtual character is perceived). We think such a finding could be of great importance in the human-intelligent virtual character interaction community in which characters work toward co-solving problems with humans.

We also asked the participants to compare their intelligence against the virtual character's intelligence (**intelligence comparison**). When observing the results, we see that the participants' responses differed significantly: as the intelligence of the virtual character increased, the rating that the virtual character was more intelligent than the participants also increased (**H2.2** is supported). However, although our participants increased their ratings from the low to medium intelligence conditions, and from the medium to high intelligence conditions, they also clearly stated that the virtual character was still less intelligent than them (the average scores in the three conditions are below 3.5 on the seven-point Likert scale we used). Our finding agrees with previous studies that explored how humans compared themselves with robots [6, 91, 105] and expands it by showing that our participants reported that they are smarter than a virtual character to which we assigned high intelligence.

For us, this finding indicates that a virtual character's efficiency in solving a problem might not be enough to make study participants indicate that such a virtual character could be more intelligent than them. Perhaps additional elements should be integrated into a virtual character to make that virtual character look and behave in a more human-like way. Another interpretation could be the easiness of the tasks that the study participants and our virtual character interacted with. Perhaps a more difficult task, such as solving a complex math problem or assembling an abstract 3D structure, could make study participants compare their intelligence differently against an intelligent virtual character. However, we should always keep in mind that personality styles (i.e., arrogance, conceitedness, egotism, and narcissism) could always be present when we ask participants how they perceive themselves compared to others [2, 43, 93] and, therefore, might impact how they rate such virtual entities.

5.3 Virtual Character's Realism

The **character interaction realism** revealed statistically significant results (**H3.1** is supported). Specifically, our participants indicated that when they interacted with a virtual character with high intelligence assigned to her, the interaction realism increased compared to the conditions in which we assigned low or medium intelligence levels. Although prior research indicated that the interaction realism between agents (e.g., in multi-agent simulations) increases when they are assigned an intelligent behavior [96], our study verifies and expands such findings by indicating that intelligence assigned to a virtual character also increases the realism of the interaction between a human and that virtual character.

In our study, the participants distinguished the similarities and dissimilarities of the different variables of character realism we examined. They provided similar ratings for the **appearance** and **movement realism** of the virtual character in all three conditions (**H3.2** and **H3.4** are supported). We know from previous research that when we assign virtual characters different appearances [71, 74, 106, 116] and motions [70, 117], this could impact how study participants perceive and interact with them. However, in our study, we used the same virtual character across all three conditions. Moreover, we used an animation controller and inverse kinematics implementation to animate the virtual character, which was the same across all conditions. Thus, we consider this to be an expected outcome.

Moreover, according to our results, when we assigned a high intelligence level to our virtual character, our participants reported higher **character behavior realism**, which we think is another interesting result (**H3.3** is supported). Although several studies discuss different factors (including intelligence) that researchers should consider if they want to improve the behavior realism of a virtual character [13, 39, 61, 94], our study proves that different intelligence levels can also impact how study participants rate the behavior realism of that virtual character, indicating that a more realistic behavior is the one that is expressed through a more intelligent virtual character, agreeing with the norm of the uncanny valley of artificial intelligence [44, 95, 99].

5.4 Gazing

Although we know that the gazing data we collected is not accurate, we think our data could still reveal interesting insights into how our participants interacted with the virtual character and the task we assigned them to accomplish. In alignment with the **attentional allocation** result, we found the **virtual character dwell gazing** to be similar across the three examined conditions (**H4.1** is not supported). Our participants in all three conditions spend only a very small amount of time gazing at the body of the virtual character. We consider this a partially unexpected result mainly because of the high co-presence ratings. Interestingly, the average gaze scores directed at the virtual character across all conditions were higher than those directed at the puzzle. We had anticipated that our study participants would spend more time gazing at the puzzles rather than the virtual human. This unexpected and intriguing finding warrants further research. However, it is worth noting that we did not place the virtual character in a position that was in the immediate view of the participant but on the periphery of our participant's field of view (we placed the virtual character on the right side of the participant following an L-shape formation). Thus, our participants had to slightly turn their heads or shoulders to direct their gaze at the virtual character. A similar result was also found for the **puzzle pieces dwell gazing**, showing that our participants spent a similar percentage of their time gazing at the puzzle pieces in all three conditions (**H4.2** is supported).

Interestingly, we found that when we assigned a low intelligence to our virtual character, the percentage time for the **puzzle box dwell gazing** was significantly increased compared to the other two intelligence levels (**H4.3** is not supported). To our understanding, when our participants noticed that the virtual character was making many mistakes, they had to correct her. At the same time, they were trying to minimize their chances of placing puzzle pieces in the wrong target spots on the puzzle board. Thus, they strategized to gaze for more time at the puzzle box, first to evaluate whether a puzzle piece already placed on the puzzle board was in the right spot and second to ensure that when they placed a puzzle piece, it would be better to place it in the right spot.

Based on the gazing data, although we know that humans gaze more at people who are within their immediate field of view [4, 65], we can conclude that our participants focused more on the task than the virtual character, without this impacting their perception toward the virtual character. After all, the only instruction they received from us was to solve the puzzle and not to engage or socialize with the virtual character. However, we still think we should conduct further experimentation regarding where we should place the virtual character and how the character's position impacts whether and how study participants focus on that virtual character or the task.

5.5 Task Execution and Puzzle Interaction

Although the virtual character did not make mistakes in the high intelligence condition and, therefore, the participants did not correct the virtual character when she was assigned high intelligence, we see that the participants made significantly more **corrections** to the mistakes made by the virtual character during the low intelligence condition compared to the medium one (**H5.1** is supported). We have seen in previously conducted studies that intelligent assistants [42, 52] and task efficiency [22, 26] occur in faster completion time and task execution. However, our study extends such findings by proving that a virtual character's intelligence level is a factor that impacts task efficiency and optimal task accomplishment and execution—participants picked up pieces and corrected the virtual character fewer times and spent less time finishing the puzzle in the high intelligence condition.

The inability of the virtual character to efficiently co-solve the puzzle with the participants resulted in expected results that indicate that, indeed, a less intelligent virtual character can impact the task execution process. Specifically, we found a similar pattern between **puzzle piece pickups** and **completion time**, showing that a less intelligent virtual character made our participants pick up more puzzle pieces and spend more time accomplishing the task than when the intelligence of the virtual character was increased (**H5.2** and **H5.3** are supported).

We also see that participants rated the **puzzle pieces manipulation** to be significantly easier during the high intelligence condition than the other two conditions (**H6** is not supported). Our participants used the virtual reality controllers to manipulate the puzzle pieces, and in all three conditions, they used the same mechanism for grabbing and placing the puzzle pieces. Although we see that the average scores are quite high in all three conditions, we suspect that two main reasons made our participants rate the easiness of puzzle pieces manipulation the way they did. First, it is because of the repeated activity (picking up puzzle pieces from the puzzle board and placing them on the table) that our participants performed to correct the mistakes of the virtual character when we assigned medium and low intelligence levels to her. Second, the total time they spent in the virtual environment during the low and medium intelligence conditions could have affected their mental and physical stamina.

5.6 Implications and Design Considerations

Our findings revealed that intelligence levels assigned to a virtual character increased the co-presence, perceived intelligence, and interaction and behavior realism as reported by the study participants. Although this study aligns with previous studies that show that intelligence is effective at increasing behavior realism [51, 76, 96], we show that intelligence levels can also impact other aspects of perceptions regarding a virtual character. Moreover, our results suggest that even a simple intelligent system (certainly, there are more advanced artificial intelligence systems that can benefit from recent advances in deep and reinforcement learning) and a relatively simple interaction task are sufficient to increase perceived intelligence and the interaction and behavior realism toward a virtual character. However, these perceptions (co-presence, perceived intelligence, and interaction and behavior realism) are sometimes but not always consistently impacted by the intelligence level of a virtual character, so this finding is particularly notable. We argue that additional behavior and cognitive capabilities assigned to a virtual character might facilitate even higher perceived intelligence levels.

The data extracted from examining how study participants interacted with different levels of intelligence assigned to a virtual character should be documented for future consideration when developing virtual reality applications. Thus, we would like to reflect on how our conclusions can aid the design of intelligent virtual characters that behave more realistically. Since task execution efficiency was not enough to convince the study participants about the intelligence of a virtual character, researchers must consider and implement additional functionalities in a virtual character that co-solves a task with humans. Among other essential functionalities, we think that variations in behavior and motion, task awareness, social engagement, and behavior dependence (i.e., the behavior of the virtual character is in direct response to the behavior of the human) could enhance behavior realism and the perceived intelligence of a virtual character. Moreover, implementing environmental events and making the virtual character aware of them (i.e., a character might respond to a sound coming from a sound source away from the task) could enhance spatial and environmental awareness and, therefore, could enhance the realism of the virtual character and provide a more human-like reaction to that virtual character. All of the previously mentioned design decisions could help create highly behaviorally realistic intelligent virtual characters.

6 LIMITATIONS

Designing and developing intelligent virtual characters is a complex process that requires several components assigned to them to work harmoniously. Thus, although we were able to develop a method that controls the intelligence of a virtual character on how to solve a jigsaw puzzle, there are several limitations. Note that these limitations do not invalidate our approach toward developing an intelligent virtual character; instead, they can help future research toward the advancement of the design of intelligent virtual characters and the evaluation of how study participants perceive and interact with such a virtual character.

Our virtual character was able to solve the jigsaw puzzle based on our fixed loop. Specifically, the virtual character randomly picks up puzzle pieces from the table and places them on the targets on the puzzle board based on the intelligence level (correct placement probability) assigned to her. As a result, the behavior sequence of our virtual character is always the same, and therefore, it could have made our character look less realistic. Moreover, it does not make sense for the virtual character to think about where to place a puzzle piece if only one is left. Thus, we think that to increase the believability of our character's behavior, we need to study and understand how to assign "thinking" timings based on the progress of solving the jigsaw puzzle. Moreover, we think that implementing additional variations of behaviors not necessarily related to the task would make the virtual character behave more realistically.

In our application, we only implemented a relatively small degree of conversation with the virtual character used in all three conditions. We argue that researchers should consider a more active dialog to increase engagement with the virtual character. We think that implementing a chatbot in conjunction with a speech recognition method (to substitute the graphical user interface) would increase the interaction realism with the virtual character and the overall experience of the study participants.

Another limitation relates to the dialog system's inability to adapt according to the virtual character's intelligence level. A more effective approach might have been for the virtual character to express uncertainty or admit to not knowing how to solve the jigsaw puzzle, rather than providing incorrect solutions that could potentially detract from the participant's experience. This aspect of the character's interaction is a critical consideration in our findings. A nuanced and context-aware dialogue from the virtual character could have offered a more authentic experience, thereby enriching our understanding of the role played by the character's intelligence. However, it is essential to explain why we chose not to vary the dialogues based on different conditions. While customizing dialogues to each intelligence level seems beneficial, it could have introduced unwanted variability into the study. Our main objective was to isolate and assess the impact of the virtual character's intelligence on puzzle-solving. Modifying dialogues for each condition would have added complexity and risked confounding the results, making it difficult to determine if observed effects were due to intelligence alone.

Lastly, we did not assess our participants' perceived workload (e.g., using NASA's Task-Load Index scale), which we consider a significant omission. Future studies should consider including such a questionnaire to understand better how study participants cope with the task's demands and how a repeated activity (picking up and placing jigsaw puzzle pieces) could potentially impact them.

7 CONCLUSIONS AND FUTURE WORK

Intelligence assigned to a virtual character is known to positively affect that virtual character's realism. However, despite studies on intelligent virtual characters that have focused on numerous domains, it is unknown whether and how intelligence levels could impact humans when they closely collaborate toward the same task. Thus, in this paper, we explored how study participants perceived and interacted with different levels of intelligence assigned to a virtual character. We found that assigning a high intelligence level to our virtual character positively impacted the overall virtual reality experience. However, it is not enough to convince people how realistic is the whole experience since more factors should be considered for people to perceive such a virtual character as an intelligent entity.

Although this work takes an important first step in developing a baseline understanding of how intelligence levels assigned to a virtual character could impact human-intelligent virtual character interaction, we think there is a lack of a comprehensive framework or guidelines to guide the design of immersive human-intelligent virtual human interaction to optimize perceived intelligence. Thus, in our future studies, we plan to explore further additional functionalities and factors that can be assigned to virtual characters as well as interaction scenarios to understand better how to build intelligent virtual characters for collaborative problem solving with humans in immersive environments. Moreover, although none of our participants mentioned anything about the missing self-avatar, in our future studies, we would like to explore whether variations of a self-avatar (e.g., assigning the body of Einstein to our participants) could impact how participants perceived themselves and compare it against intelligent virtual characters in which their appearance also varies. Lastly, we would like to explore the reasons behind the large inter-individual variability in our results. We think this variability could stem from several factors, such as the diversity of participant demographics, behavioral differences between participants, bias errors, and lack of skill. While exploring these variations is intriguing, it fell outside the scope of our current work. Nonetheless, it would be valuable to investigate these factors in future research to gain a deeper understanding of why participants responded as they did—particularly how demographic or other factors influence the perception of a virtual character's intelligence.

REFERENCES

[1] Sames Al Moubayed, Malek Baklouti, Mohamed Chetouani, Thierry Dutoit, Ammar Mahdhaoui, J-C Martin, Stanislav Ondas, Catherine Pelachaud, Jérôme Urbain, and Mehmet Yilmaz. 2009. Generating robot/agent backchannels during a storytelling experiment. In *2009 IEEE International Conference on Robotics and Automation*. IEEE, 3749–3754.

[2] Theo Araujo, Natali Helberger, Sanne Kruikemeier, and Claes H De Vreese. 2020. In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & society* 35 (2020), 611–623.

[3] Andreas Aristidou and Joan Lasenby. 2011. FABRIK: A fast, iterative solver for the Inverse Kinematics problem. *Graphical Models* 73, 5 (2011), 243–260.

[4] Peter Auer and Elisabeth Zima. 2021. On word searches, gaze, and co-participation. *Gesprächsforschung. Online-Zeitschrift Zur Verbalen Interaktion* 22 (2021), 390–425.

[5] Jeremy N Bailenson, Jim Blascovich, Andrew C Beall, and Jack M Loomis. 2003. Interpersonal distance in immersive virtual environments. *Personality and social psychology bulletin* 29, 7 (2003), 819–833.

[6] Maxwell Bennett, Tom Williams, Daria Thames, and Matthias Scheutz. 2017. Differences in interaction patterns and perception for teleoperated and autonomous humanoid robots. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 6589–6594.

[7] Timothy Bickmore and Justine Cassell. 1999. Small talk and conversational storytelling in embodied conversational interface agents. In *AAAI fall symposium on narrative intelligence*. 87–92.

[8] Timothy Bickmore and Justine Cassell. 2005. Social dialongue with embodied conversational agents. *Advances in natural multimodal dialogue systems* (2005), 23–54.

[9] Timothy W Bickmore, Laura M Pfeifer, and Michael K Paasche-Orlow. 2007. Health document explanation by virtual agents. In *Intelligent Virtual Agents: 7th International Conference, IVA 2007 Paris, France, September 17-19, 2007 Proceedings 7*. Springer, 183–196.

[10] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2001. The magicbook-moving seamlessly between reality and virtuality. *IEEE Computer Graphics and applications* 21, 3 (2001), 6–8.

[11] Frank Biocca, Chad Harms, and Judee K Burgoon. 2003. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & virtual environments* 12, 5 (2003), 456–480.

[12] 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.

[13] Dario Bombari, Marianne Schmid Mast, Elena Canadas, and Manuel Bachmann. 2015. Studying social interactions through immersive virtual environment technology: virtues, pitfalls, and future challenges. *Frontiers in psychology* 6 (2015), 869.

[14] Doug A Bowman and Ryan P McMahan. 2007. Virtual reality: how much immersion is enough? *Computer* 40, 7 (2007), 36–43.
[15] Cynthia Breazeal. 2004. *Designing sociable robots*. MIT press.
[16] David C Brogan, Ronald A Metoyer, and Jessica K Hodgins. 1998. Dynamically simulated characters in virtual environments. *IEEE computer graphics and applications* 18, 5 (1998), 58–69.
[17] Stéphanie Buisine, Sarkis Abrilian, and Jean-Claude Martin. 2004. Evaluation of Multimodal Behaviour of Embodied Agents: Cooperation between Speech and Gestures. *From brows to trust: evaluating embodied conversational agents* (2004), 217–238.
[18] 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.
[19] Angelo Cafaro, Brian Ravenet, Magalie Ochs, Hannes Högni Vilhjálmsson, and Catherine Pelachaud. 2016. The effects of interpersonal attitude of a group of agents on user's presence and proxemics behavior. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 6, 2 (2016), 1–33.
[20] Christer Carlsson and Olof Hagsand. 1993. DIVE A multi-user virtual reality system. In *Proceedings of IEEE virtual reality annual international symposium*. IEEE, 394–400.
[21] Justine Cassell, Timothy Bickmore, Mark Billinghurst, Lee Campbell, Kenny Chang, Hannes Vilhjálmsson, and Hao Yan. 1999. Embodiment in conversational interfaces: Rea. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 520–527.
[22] Wesley P Chan, Geoffrey Hanks, Maram Sakr, Haomiao Zhang, Tiger Zuo, HF Machiel Van der Loos, and Elizabeth Croft. 2022. Design and evaluation of an augmented reality head-mounted display interface for human robot teams collaborating in physically shared manufacturing tasks. *ACM Transactions on Human-Robot Interaction (THRI)* 11, 3 (2022), 1–19.
[23] Minsoo Choi, Yeling Jiang, Farid Breidi, Christos Mousas, and Mesut Akdere. 2022. A Mixed Reality Platform for Collaborative Technical Assembly Training. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*. 1–2.
[24] Minsoo Choi, Alexandros Koilias, Matias Volonte, Dominic Kao, and Christos Mousas. 2023. Exploring the Appearance and Voice Mismatch of Virtual Characters. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 555–560.
[25] Mathieu Chollet, Magalie Ochs, and Catherine Pelachaud. 2017. A methodology for the automatic extraction and generation of non-verbal signals sequences conveying interpersonal attitudes. *IEEE Transactions on Affective Computing* 10, 4 (2017), 585–598.
[26] Michael Chu, Jane Matthews, and Peter ED Love. 2018. Integrating mobile building information modelling and augmented reality systems: an experimental study. *Automation in Construction* 85 (2018), 305–316.
[27] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. Academic press.
[28] Dixuan Cui, Dominic Kao, and Christos Mousas. 2021. Toward understanding embodied human-virtual character interaction through virtual and tactile hugging. *Computer Animation and Virtual Worlds* 32, 3-4 (2021), e2009.
[29] Salam Daher, Kangsoo Kim, Myungho Lee, Ryan Schubert, Gerd Bruder, Jeremy Bailenson, and Greg Welch. 2017. Effects of social priming on social presence with intelligent virtual agents. In *Intelligent Virtual Agents: 17th International Conference, IVA 2017, Stockholm, Sweden, August 27-30, 2017, Proceedings 17*. Springer, 87–100.
[30] Amol Deshmukh, Bart Craenen, Mary Ellen Foster, and Alessandro Vinciarelli. 2018. The more I understand it, the less I like it: The relationship between understandability and Godspeed scores for robotic gestures. In *2018 27th IEEE International symposium on robot and human interactive communication (RO-MAN)*. IEEE, 216–221.
[31] Nat Durlach and Mel Slater. 2000. Presence in shared virtual environments and virtual togetherness. *Presence: Teleoperators & Virtual Environments* 9, 2 (2000), 214–217.
[32] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods* 41, 4 (2009), 1149–1160.
[33] Sara Feldman, Ozge Nilay Yalcin, and Steve DiPaola. 2017. Engagement with artificial intelligence through natural interaction models. *Electronic Visualisation and the Arts (EVA 2017)* (2017), 296–303.
[34] 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), 299.
[35] Jesse Fox, Sun Joo Ahn, Joris H Janssen, Leo Yeykelis, Kathryn Y Segovia, and Jeremy N Bailenson. 2015. Avatars versus agents: a meta-analysis quantifying the effect of agency on social influence. *Human–Computer Interaction* 30, 5 (2015), 401–432.
[36] Rebecca Fribourg, Ferran Argelaguet, Ludovic Hoyet, and Anatole Lécuyer. 2018. Studying the sense of embodiment in VR shared experiences. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 273–280.
[37] Manuel Giuliani and Alois Knoll. 2011. Evaluating supportive and instructive robot roles in human-robot interaction. In *Social Robotics: Third International Conference, ICSR 2011, Amsterdam, The Netherlands, November 24-25, 2011. Proceedings 3*. Springer, 193–203.
[38] Scott W Greenwald, Zhangyuan Wang, Markus Funk, and Pattie Maes. 2017. Investigating social presence and communication with embodied avatars in room-scale virtual reality. In *Immersive Learning Research Network: Third International Conference, iLRN 2017,*

Coimbra, Portugal, June 26–29, 2017. Proceedings 3. Springer, 75–90.
[39] Rosanna E Guadagno, Jim Blascovich, Jeremy N Bailenson, and Cade McCall. 2007. Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology* 10, 1 (2007), 1–22.
[40] Rosanna E Guadagno, Kimberly R Swinth, and Jim Blascovich. 2011. Social evaluations of embodied agents and avatars. *Computers in Human Behavior* 27, 6 (2011), 2380–2385.
[41] Andrea L Guzman and Seth C Lewis. 2020. Artificial intelligence and communication: A Human–Machine Communication research agenda. *New Media & Society* 22, 1 (2020), 70–86.
[42] Parian Haghighat, Toan Nguyen, Mina Valizadeh, Mohammad Arvan, Natalie Parde, Myunghee Kim, and Heejin Jeong. 2023. Effects of an intelligent virtual assistant on office task performance and workload in a noisy environment. *Applied Ergonomics* 109 (2023), 103969.
[43] Reinhard Haller. 2015. *The Narcissist in the Mirror: A Field Guide to Our Selves and Other People*. Ecowin.
[44] David Hanson, Andrew Olney, Steve Prilliman, Eric Mathews, Marge Zielke, Derek Hammons, Raul Fernandez, and Harry Stephanou. 2005. Upending the uncanny valley. In *AAAI*, Vol. 5. 1728–1729.
[45] Markus Häring, Jessica Eichberg, and Elisabeth André. 2012. Studies on grounding with gaze and pointing gestures in human-robot-interaction. In *Social Robotics: 4th International Conference, ICSR 2012, Chengdu, China, October 29-31, 2012. Proceedings 4*. Springer, 378–387.
[46] Frank Hegel, Soren Krach, Tilo Kircher, Britta Wrede, and Gerhard Sagerer. 2008. Understanding social robots: A user study on anthropomorphism. In *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 574–579.
[47] Jon-Chao Hong, Ming-Yueh Hwang, Ker-Ping Tam, Yi-Hsuan Lai, and Li-Chun Liu. 2012. Effects of cognitive style on digital jigsaw puzzle performance: A GridWare analysis. *Computers in Human Behavior* 28, 3 (2012), 920–928.
[48] Xing-Da Jhan, Sai-Keung Wong, Elham Ebrahimi, Yuwen Lai, Wei-Chia Huang, and Sabarish V Babu. 2022. Effects of Small Talk With a Crowd of Virtual Humans on Users' Emotional and Behavioral Responses. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3767–3777.
[49] Matthew Johnson, Jeffrey M Bradshaw, Paul J Feltovich, Catholijn M Jonker, M Birna Van Riemsdijk, and Maarten Sierhuis. 2014. Coactive design: Designing support for interdependence in joint activity. *Journal of Human-Robot Interaction* 3, 1 (2014), 43–69.
[50] 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.
[51] Kangsoo Kim, Luke Boelling, Steffen Haesler, Jeremy Bailenson, Gerd Bruder, and Greg F Welch. 2018. Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 105–114.
[52] Kangsoo Kim, Celso M de Melo, Nahal Norouzi, Gerd Bruder, and Gregory F Welch. 2020. Reducing task load with an embodied intelligent virtual assistant for improved performance in collaborative decision making. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 529–538.
[53] Jérémy Lacoche, Nico Pallamin, Thomas Boggini, and Jérôme Royan. 2017. Collaborators awareness for user cohabitation in co-located collaborative virtual environments. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. 1–9.
[54] Khanh-Duy Le, Morten Fjeld, Ali Alavi, and Andreas Kunz. 2017. Immersive environment for distributed creative collaboration. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. 1–4.
[55] Jae-Gil Lee, Ki Joon Kim, Sangwon Lee, and Dong-Hee Shin. 2015. Can autonomous vehicles be safe and trustworthy? Effects of appearance and autonomy of unmanned driving systems. *International Journal of Human-Computer Interaction* 31, 10 (2015), 682–691.
[56] Kwan Min Lee, Wei Peng, Seung-A Jin, and Chang Yan. 2006. Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *Journal of communication* 56, 4 (2006), 754–772.
[57] Huimin Liu, Minsoo Choi, Dominic Kao, and Christos Mousas. 2023. Synthesizing game levels for collaborative gameplay in a shared virtual environment. *ACM Transactions on Interactive Intelligent Systems* 13, 1 (2023), 1–36.
[58] Huimin Liu, Minsoo Choi, Liuchuan Yu, Alexandros Koilias, Lap-Fai Yu, and Christos Mousas. 2022. Synthesizing Shared Space Virtual Reality Fire Evacuation Training Drills. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 459–464.
[59] Kuan-Yu Liu, Matias Volonte, Yu-Chun Hsu, Sabarish V Babu, and Sai-Keung Wong. 2019. Interaction with proactive and reactive agents in box manipulation tasks in virtual environments. *Computer Animation and Virtual Worlds* 30, 3-4 (2019), e1881.
[60] Kuan-Yu Liu, Sai-Keung Wong, Matias Volonte, Elham Ebrahimi, and Sabarish V Babu. 2022. Investigating the Effects of Leading and Following Behaviors of Virtual Humans in Collaborative Fine Motor Tasks in Virtual Reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 330–339.
[61] Birgit Lugrin, Catherine Pelachaud, and David Traum. 2022. The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics, Volume 2: Interactivity, Platforms, Application. (2022).

[62] Sebastian Marwecki, Maximilian Brehm, Lukas Wagner, Lung-Pan Cheng, Florian'Floyd' Mueller, and Patrick Baudisch. 2018. Virtualspace-overloading physical space with multiple virtual reality users. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–10.

[63] Elias Matsas and George-Christopher Vosniakos. 2017. Design of a virtual reality training system for human–robot collaboration in manufacturing tasks. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 11 (2017), 139–153.

[64] Rachel McDonnell, Martin Breidt, and Heinrich H Bülthoff. 2012. Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics (TOG)* 31, 4 (2012), 1–11.

[65] David McNeill. 2006. Gesture, gaze, and ground. In *Machine Learning for Multimodal Interaction: Second International Workshop, MLMI 2005, Edinburgh, UK, July 11-13, 2005, Revised Selected Papers 2*. Springer, 1–14.

[66] Tim Merritt, Kevin McGee, Teong Leong Chuah, and Christopher Ong. 2011. Choosing human team-mates: perceived identity as a moderator of player preference and enjoyment. In *Proceedings of the 6th International Conference on Foundations of Digital Games*. 196–203.

[67] Tim Merritt, Christopher Ong, Teong Leong Chuah, and Kevin McGee. 2011. Did you notice? artificial team-mates take risks for players. In *Intelligent Virtual Agents: 10th International Conference, IVA 2011, Reykjavik, Iceland, September 15-17, 2011. Proceedings 11*. Springer, 338–349.

[68] Timothy Robert Merritt. 2012. A failure of imagination: How and why people respond differently to human and computer team-mates. (2012).

[69] Tim R Merritt, Kian Boon Tan, Christopher Ong, Aswin Thomas, Teong Leong Chuah, and Kevin McGee. 2011. Are artificial team-mates scapegoats in computer games. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. 685–688.

[70] Christos Mousas, Dimitris Anastasiou, and Ourania Spantidi. 2018. The effects of appearance and motion of virtual characters on emotional reactivity. *Computers in Human Behavior* 86 (2018), 99–108.

[71] Christos Mousas, Alexandros Koilias, Banafsheh Rekabdar, Dominic Kao, and Dimitris Anastaslou. 2021. Toward understanding the effects of virtual character appearance on avoidance movement behavior. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 40–49.

[72] Sara Moussawi and Marios Koufaris. 2019. Perceived intelligence and perceived anthropomorphism of personal intelligent agents: Scale development and validation. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*. 115–124.

[73] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 72–78.

[74] Michael G Nelson, Alexandros Koilias, Christos-Nikolaos Anagnostopoulos, and Christos Mousas. 2022. Effects of Rendering Styles of a Virtual Character on Avoidance Movement Behavior. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 594–599.

[75] Anton Nijholt. 2002. Embodied agents: A new impetus to humor research. In *The April Fools Day Workshop on Computational Humour*, Vol. 20. In: Proc. Twente Workshop on Language Technology, 101–111.

[76] 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.

[77] Kristine L Nowak and Frank Biocca. 2003. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments* 12, 5 (2003), 481–494.

[78] Ye Pan and Anthony Steed. 2017. The impact of self-avatars on trust and collaboration in shared virtual environments. *PloS one* 12, 12 (2017), e0189078.

[79] 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.

[80] Catherine Pelachaud and Isabella Poggi. 2002. Subtleties of facial expressions in embodied agents. *The Journal of Visualization and Computer Animation* 13, 5 (2002), 301–312.

[81] David-Paul Pertaub, Mel Slater, and Chris Barker. 2002. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence* 11, 1 (2002), 68–78.

[82] Thammathip Piumsomboon, Youngho Lee, Gun Lee, and Mark Billinghurst. 2017. CoVAR: a collaborative virtual and augmented reality system for remote collaboration. In *SIGGRAPH Asia 2017 Emerging Technologies*. 1–2.

[83] Lauriane Pouliquen-Lardy, Isabelle Milleville-Pennel, Francois Guillaume, and Franck Mars. 2016. Remote collaboration in virtual reality: asymmetrical effects of task distribution on spatial processing and mental workload. *Virtual Reality* 20 (2016), 213–220.

[84] Chao Qu, Willem-Paul Brinkman, Yun Ling, Pascal Wiggers, and Ingrid Heynderickx. 2014. Conversations with a virtual human: Synthetic emotions and human responses. *Computers in Human Behavior* 34 (2014), 58–68.

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

[86] Daniel Roth, Constantin Klelnbeck, Tobias Feigl, Christopher Mutschler, and Marc Erich Latoschik. 2018. Beyond replication: Augmenting social behaviors in multi-user virtual realities. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 215–222.
[87] Daniel Roth, Kristoffer Waldow, Marc Erich Latoschik, Arnulph Fuhrmann, and Gary Bente. 2017. Socially immersive avatar-based communication. In *2017 IEEE Virtual Reality (VR)*. IEEE, 259–260.
[88] Maha Salem, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. 2011. A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction. In *2011 ro-man*. IEEE, 247–252.
[89] Franziska Schrammel, Sebastian Pannasch, Sven-Thomas Graupner, Andreas Mojzisch, and Boris M Velichkovsky. 2009. Virtual friend or threat? The effects of facial expression and gaze interaction on psychophysiological responses and emotional experience. *Psychophysiology* 46, 5 (2009), 922–931.
[90] Ralph Schroeder. 2002. Social interaction in virtual environments: Key issues, common themes, and a framework for research. *The social life of avatars: Presence and interaction in shared virtual environments* (2002), 1–18.
[91] Elaine Short, Justin Hart, Michelle Vu, and Brian Scassellati. 2010. No fair‼ an interaction with a cheating robot. In *2010 5th acm/ieee international conference on human-robot interaction (hri)*. IEEE, 219–226.
[92] Alexander Shoulson, Max L Gilbert, Mubbasir Kapadia, and Norman I Badler. 2013. An event-centric planning approach for dynamic real-time narrative. In *Proceedings of Motion on Games*. 121–130.
[93] Keng Siau and Weiyu Wang. 2018. Building trust in artificial intelligence, machine learning, and robotics. *Cutter business technology journal* 31, 2 (2018), 47–53.
[94] Barry G Silverman. 2004. Toward realism in human performance simulation. In *The science and simulation of human performance*. Vol. 5. Emerald Group Publishing Limited, 469–498.
[95] Marita Skjuve, Ida Maria Haugstveit, Asbjørn Følstad, and Petter Brandtzaeg. 2019. Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human–chatbot interaction. *Human Technology* 15, 1 (2019), 30.
[96] James Skorupski, Josh McCoy, Catherine Zanbaka, Kathy Ryall, and Michael Mateas. 2012. Three avatars walk into a bar: Defining and evaluating realism for virtual supporting characters. In *Proc. International Conference on the Foundations of Digital Games*.
[97] Harrison Jesse Smith and Michael Neff. 2018. Communication behavior in embodied virtual reality. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.
[98] Misha Sra, Aske Mottelson, and Pattie Maes. 2018. Your place and mine: Designing a shared VR experience for remotely located users. In *Proceedings of the 2018 designing interactive systems conference*. 85–97.
[99] Jan-Philipp Stein and Peter Ohler. 2017. Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition* 160 (2017), 43–50.
[100] William R Swartout, Jonathan Gratch, Randall W Hill Jr, Eduard Hovy, Stacy Marsella, Jeff Rickel, and David Traum. 2006. Toward virtual humans. *AI Magazine* 27, 2 (2006), 96–96.
[101] Kazuaki Tanaka, Satoshi Onoue, Hideyuki Nakanishi, and Hiroshi Ishiguro. 2013. Motion is enough: How real-time avatars improve distant communication. In *2013 International Conference on Collaboration Technologies and Systems (CTS)*. IEEE, 465–472.
[102] Marcus Thiebaux, Stacy Marsella, Andrew N Marshall, and Marcelo Kallmann. 2008. Smartbody: Behavior realization for embodied conversational agents. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*. 151–158.
[103] Serge Thill, Maria Riveiro, and Maria Nilsson. 2015. Perceived intelligence as a factor in (semi-) autonomous vehicle UX. In *"Experiencing Autonomous Vehicles: Crossing the Boundaries between a Drive and a Ride" workshop in conjunction with CHI2015*.
[104] David Traum and Jeff Rickel. 2002. Embodied agents for multi-party dialogue in immersive virtual worlds. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2*. 766–773.
[105] Daniel Ullman, Lolanda Leite, Jonathan Phillips, Julia Kim-Cohen, and Brian Scassellati. 2014. Smart human, smarter robot: How cheating affects perceptions of social agency. In *Proceedings of the annual meeting of the cognitive science society*, Vol. 36.
[106] Matias Volante, Sabarish V Babu, Himanshu Chaturvedi, Nathan Newsome, Elham Ebrahimi, Tania Roy, Shaundra B Daily, and Tracy Fasolino. 2016. Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. *IEEE transactions on visualization and computer graphics* 22, 4 (2016), 1326–1335.
[107] Matias Volonte, Yu-Chun Hsu, Kuan-Yu Liu, Joe P Mazer, Sai-Keung Wong, and Sabarish V Babu. 2020. Effects of interacting with a crowd of emotional virtual humans on users' affective and non-verbal behaviors. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 293–302.
[108] 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.
[109] Astrid M Von der Pütten, Nicole C Krämer, Jonathan Gratch, and Sin-Hwa Kang. 2010. "It doesn't matter what you are!" explaining social effects of agents and avatars. *Computers in Human Behavior* (2010).

[110] Henriette C Van Vugt, Jeremy N Bailenson, Johan F Hoorn, and Elly A Konijn. 2008. Effects of facial similarity on user responses to embodied agents. *ACM Transactions on Computer-Human Interaction (TOCHI)* 17, 2 (2008), 1–27.

[111] Katharina Weitz, Dominik Schiller, Ruben Schlagowski, Tobias Huber, and Elisabeth André. 2019. " Do you trust me?" Increasing user-trust by integrating virtual agents in explainable AI interaction design. In *Proceedings of the 19th ACM international conference on intelligent virtual agents*. 7–9.

[112] Carolin Wienrich, Kristina Schindler, Nina Döllinqer, Simon Kock, and Ole Traupe. 2018. Social presence and cooperation in large-scale multi-user virtual reality-the relevance of social interdependence for location-based environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 207–214.

[113] Evan James Williams. 1949. Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Chemistry* 2, 2 (1949), 149–168.

[114] Aya Yadlin-Segal and Yael Oppenheim. 2021. Whose dystopia is it anyway? Deepfakes and social media regulation. *Convergence* 27, 1 (2021), 36–51.

[115] Qinyu Zhang, Shan Liu, Jie Xu, Xiang Ji, Yubo Zhang, and Yanfang Liu. 2022. Designing for Perceived Intelligence in Human-Agent Interaction: A Systematic Review. In *International Conference on Human-Computer Interaction*. Springer, 713–720.

[116] Katja Zibrek, Elena Kokkinara, and Rachel McDonnell. 2018. The effect of realistic appearance of virtual characters in immersive environments-does the character's personality play a role? *IEEE transactions on visualization and computer graphics* 24, 4 (2018), 1681–1690.

[117] Katja Zibrek, Benjamin Niay, Anne-Hélène Olivier, Ludovic Hoyet, Julien Pettre, and Rachel McDonnell. 2020. The effect of gender and attractiveness of motion on proximity in virtual reality. *ACM Transactions on Applied Perception (TAP)* 17, 4 (2020), 1–15.

A ALGORITHM

Algorithm 1 Virtual Character State Decision Algorithm

Input:

$R \in \{R_1, \cdots, R_l\}$ ▷ R is a list of puzzle targets not yet solved
$A \in \{A_1, \cdots, A_m\}$ ▷ A is a list of puzzle pieces able to be picked up
$P \in \{(x_1, x_1^a), \cdots, (x_n, x_n^a)\}$ ▷ P is a list of pairs of a puzzle piece and its answer
I ▷ I is virtual character intelligence
V_c ▷ V_c is current puzzle piece interacted by virtual character
S_c ▷ S_c is current virtual character state $\{PickUp, Place, \text{or } Wait\}$

Output:

V_n ▷ V_n is next puzzle piece interacted by virtual character
S_n ▷ S_n is next virtual character state $\{PickUp, Place, \text{or } Wait\}$

function CharacterAI(R, A, P, I, V_c, S_c)
  if $S_c = PickUp$ **then**
    if $A > 0$ **then**
      Choose T from A Randomly
      Pick up T
      $V_n \leftarrow T$
      $S_n \leftarrow Place$
    else
      $S_n \leftarrow Wait$
    end if
  else if $S_c = Place$ **then**
    Choose D from 0 to 100 Randomly
    if $D \le I$ **then**
      Choose T as V_c^a from P
    else
      while $T = V_c^a$ **do**
        Choose T from R Randomly
      end while
    end if
    Place V_c on the T
    $V_n \leftarrow NULL$
    $S_n \leftarrow PickUp$
  else if $S_c = Wait$ **then**
    if $A > 0$ **then**
      $S_n \leftarrow PickUp$
    end if
  end if
  return V_n, S_n
end function

B QUESTIONNAIRE

Table 2. The questionnaire we used in our study.

#	Question/Statement	Anchors of the Scale
	Co-Presence	
1	I noticed another individual.	1: Never; 7: Always
2	The other individual noticed me.	1: Never; 7: Always
3	The other individual's presence was obvious to me.	1: Never; 7: Always
4	My presence was obvious to the other individual.	1: Never; 7: Always
5	The other individual caught my attention.	1: Never; 7: Always
6	I caught the other individual's attention.	1: Never; 7: Always
	Attentional Allocation	
7	I was easily distracted from the other individual when other things were going on.	1: Never; 7: Always
8	The other individual was easily distracted from me when other things were going on.	1: Never; 7: Always
9	I remained focused on the other individual throughout our interaction.	1: Never; 7: Always
10	The other individual remained focused on me throughout our interaction.	1: Never; 7: Always
11	The other individual did not receive my full attention.	1: Never; 7: Always
12	I did not receive the other individual's full attention.	1: Never; 7: Always
	Perceived Behavioral Independence	
13	My behavior was often in direct response to the other individual's behavior.	1: Never; 7: Always
14	The behavior of the other individual was often in direct response to my behavior.	1: Never; 7: Always
15	I reciprocated the other individual's actions.	1: Never; 7: Always
16	The other individual reciprocated my actions.	1: Never; 7: Always
17	The other individual's behavior was closely tied to my behavior.	1: Never; 7: Always
18	My behavior was closely tied to the other individual's behavior.	1: Never; 7: Always
	Perceived Intelligence	
19	The other individual was able to operate without my intervention.	1: Never; 7: Always
20	The other individual was aware of the virtual environment.	1: Never; 7: Always
21	The other individual was able to set and pursue tasks by herself in anticipation of future needs.	1: Never; 7: Always
22	The other individual was able to complete tasks quickly.	1: Never; 7: Always
23	The other individual was able to find and process the necessary information for completing the task.	1: Never; 7: Always
24	The other individual was able to adapt/adjust her behavior based on prior events.	1: Never; 7: Always
	Intelligence Comparison	
25	Do you think the other individual was more intelligent than you?	1: Not more intelligent than me; 7: More intelligent than me
	Character's Interaction Realism	
26	How realistic were your interactions with the other individual?	1: Not realistic at all; 7: Highly realistic
	Character's Behavior Realism	
27	I found the other individual's behavior realistic.	1: Not realistic at all; 7: Highly realistic
	Character's Appearance Realism	
28	I found the other individual's appearance realistic.	1: Not realistic at all; 7: Highly realistic
	Character's Movement Realism	
29	I found the other individual's movements realistic.	1: Not realistic at all; 7: Highly realistic
	Puzzle Piece Manipulation	
30	Was it simple to pick up and manipulate the puzzle pieces?	1: Very hard; 7: Very simple

Received 14 May 2023; revised 17 July 2024; accepted 10 October 2024