The Effects of Virtual Character's Intelligence and Task's Complexity during an Immersive Jigsaw Puzzle Co-solving Task

Minsoo Choi
Department of Computer Graphics Technology, Purdue University
USA
choi714@purdue.edu

Dixuan Cui
Department of Mass Communication, Sam Houston State University
USA
dxc164@shsu.edu

Alexandros Koilias
Department of Performing and Digital Arts, University of the Peloponnese
Greece
alex.kilias@go.uop.gr

Christos Mousas
Department of Computer Graphics Technology, Purdue University
USA
cmousas@purdue.edu

Figure 1: A study participant immersed in the living room environment co-solving the jigsaw puzzle with a virtual character.

Abstract

Virtual character's intelligence and task complexity are essential components for the design of collaboration between humans and virtual characters. However, studies have yet to explore the impact of a virtual character's intelligence and task complexity during a collaboration task. To explore these impacts, we implemented a jigsaw puzzle co-solving experience and conducted a within-group study ($N = 27$) following a 2 (intelligence: low vs. high) × 2 (complexity: low vs. high) study design. During the puzzle co-solving process, we collected participants' gaze data for each experimental condition. After our participants completed each condition (co-solved the jigsaw puzzle), they answered a survey that captured several variables. We found several significant main and interaction effect results indicating the impact of the virtual character's intelligence and the task's complexity. Specifically, the virtual character's intelligence impacted most survey ratings, including the virtual character's intelligence, perceived collaboration, self-confidence, and perceived contribution. Also, the task complexity affected perceived collaboration, the virtual character's public awareness, and self-confidence. Furthermore, we found that the virtual character's intelligence and complexity dominantly impacted our participants' task load and perception of the virtual character's confidence. Moreover, we found that our participants gazed at the virtual character and the puzzle piece more often than the puzzle goal across all experimental conditions. Our findings can inform the design of intelligent virtual characters that co-solve tasks with users.

CCS Concepts

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

Keywords

virtual reality, virtual character, intelligence, task complexity, jigsaw puzzle, puzzle co-solving

MIG '24, November 21–23, 2024, Arlington, VA, USA

ACM ISBN 979-8-4007-1090-2/24/11
https://doi.org/10.1145/3677388.3696324

ACM Reference Format:
Minsoo Choi, Dixuan Cui, Alexandros Koilias, and Christos Mousas. 2024. The Effects of Virtual Character's Intelligence and Task's Complexity during an Immersive Jigsaw Puzzle Co-solving Task. In *The 17th ACM SIGGRAPH Conference on Motion, Interaction, and Games (MIG '24), November 21–23, 2024, Arlington, VA, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3677388.3696324

1 Introduction

Utilizing virtual reality to provide immersive interaction with intelligent virtual characters has opened new avenues in human-computer interaction by providing more engaging user experiences [Choi et al. 2024; Gamage and Ennis 2018; Yang et al. 2024]. This trend has triggered the use of intelligent virtual characters in applications for various purposes, such as training [Choi et al. 2022; Guimarães et al. 2020; Liu et al. 2022] or games [Gamage and Ennis 2018; Liu et al. 2023]. Several researchers have focused on utilizing intelligent virtual characters in human-agent interaction within virtual environments since intelligent virtual characters can express emotions [Qu et al. 2014; Volonte et al. 2020], complete tasks [Shoulson et al. 2013], provide instructions [Choi et al. 2022] or communicate with users through either verbal or nonverbal feedback [Zhao and Ma 2020]. Furthermore, several researchers have investigated how features of virtual characters, such as appearance and voice [Choi et al. 2023; Lam et al. 2023; Mousas et al. 2018, 2021; Nelson et al. 2022], impact user experiences like immersion [Sierra Rativa et al. 2020] or social presence [Cui et al. 2021; Fox et al. 2015; Yoon et al. 2019]. These studies have expanded exploration into the impact of intelligent virtual characters on human perception [Hanna et al. 2015] and user experiences [Kim et al. 2020] during collaboration. However, researchers have yet to investigate the effect of task complexity during co-solving tasks that humans and intelligent virtual characters perform. Furthermore, to the best of our knowledge, there has yet to be any study about the impacts of a virtual character's intelligence, task complexity, and their interaction effect on human perceptions and user experiences during collaborative tasks.

Thus, to explore how these factors impact study participants, we implemented a virtual reality jigsaw puzzle co-solving experience with an intelligent virtual character (see Figure 1). We base our decision to use a jigsaw puzzle co-solving activity on several factors. First, as noted by Scoular et al. [Scoular et al. 2020], collaboration necessitates shared understanding and acknowledgment of contribution. Completing a jigsaw puzzle provides a shared goal for our participants and the virtual character, fostering collaboration. Additionally, observing how the virtual character approaches the puzzle enhances shared understanding. Second, since the virtual character's intelligence was essential to our study, we considered the jigsaw puzzle since solving it requires cognitive and perceptual abilities [Burns et al. 2006]. Fissler et al. [Fissler et al. 2018] described jigsaw puzzle solving as a task demanding cognitive skills, perceptual reasoning, and working memory, making it suitable for our purposes. Last, researchers [Burghart et al. 2006; Häring et al. 2012] have used jigsaw puzzle solving as a collaborative task to study human-agent interaction, supporting our choice.

We created four experimental conditions for our study following a 2 (intelligence: low vs. high) × 2 (complexity: low vs. high) study design. Regarding the intelligence factor, the virtual character with 0% intelligence (i.e., unintelligent virtual character) always placed puzzle pieces in the wrong spots, and the virtual character with 100% intelligence always placed puzzle pieces in the correct spot. In the case of the complexity factor, we defined it by changing the number and size of puzzle pieces. The low complexity task consisted of 25 puzzle pieces, and the high complexity task consisted of 100 puzzle pieces, smaller in size. We conducted a within-group study ($N = 27$) to collect quantitative and qualitative data. Specifically, the quantitative data includes participants' ratings on their perceptions and user experiences and application logs of participants' dwell gazing. The qualitative data consisted of participants' feedback regarding their experiences in co-solving the jigsaw puzzle.

We organized this paper into the following sections. In Section 2, we discuss related work. In Section 3, we present our methodology and materials. In Section 4, we report our results. In Section 5, we discuss our findings, limitations, and implications of our study. Last, in Section 6, we conclude and discuss future works.

2 Related Work

2.1 Human-Agent Interaction

Several researchers have focused on automation as one of the critical elements in defining an agent and expanded this definition to include human-computer interaction [Lewis 1998; Norman 1994]. Based on this, agents can take various forms (i.e., virtual characters, robots) with the capacity to perform specific tasks automatically. Researchers have explored human-agent interaction with multiple dimensions. Sauppé and Mutlu [Sauppé and Mutlu 2015] presented an instructional robot for training machine assembly and its instructional style to enhance training outcomes in terms of performance and user experience. Goetz et al. [Goetz et al. 2003] investigated the impact of matching the appearance and behavior of robots on humans and robot collaboration. They found that participants followed the robot more when its attitude aligned with the seriousness of the given task. Salem et al. [Salem et al. 2011] delved into the non-verbal behaviors of humanoid robots and reported that gestures positively impacted participants' evaluation of the robot.

Furthermore, Yang et al. [Yang et al. 2022] focused on a virtual character that translated a lecture to sign language in a mixed reality environment and investigated how this virtual character framing and manipulation affected users' preferences and accessibility. They reported that their participants preferred a virtual character, the size of which is similar to a real human. Zibrek and McDonnell [Zibrek and McDonnell 2014] integrated different rendering styles into a virtual character to investigate how it affected the human perception of the virtual character's personality. They found that the cartoon style made the virtual character appear to have a more agreeable personality. Moreover, Wang et al. [Wang et al. 2019] compared different types of virtual agents, such as voice-only or different-sized virtual agents, in augmented reality platforms and found that participants preferred miniature embodied agents.

2.2 Collaboration between Human and Agent

Researchers have explored collaboration between humans and virtual agents. Bellamy et al. [Bellamy et al. 2017] emphasized the necessity for mutual understanding of goals and tasks in human-agent interaction and highlighted the significance of trust in virtual agents during collaboration. Andrist et al. [Andrist et al. 2017] developed a collaborative virtual agent that guided participants in making sandwiches and introduced a bidirectional gaze model. This model enabled the virtual agent to comprehend the participant's gaze and reciprocate with its gaze to facilitate communication. Their findings revealed that this gaze model improved performance and reduced errors during collaborative tasks. Schmidbauer et al. [Schmidbauer et al. 2020] investigated collaborative robots assisting humans in assembly tasks and proposed a methodology for efficient task allocation between humans and robots. Additionally, Moradinezhad and Solovey [Moradinezhad and Solovey 2021] examined how the cooperativeness of virtual characters influenced trust levels among participants. They reported that participants exhibited higher trust in more cooperative virtual characters and stressed the significance of a virtual character's cooperativeness based on prior experiences.

2.3 Collaborative and Intelligent Agent

Researchers have explored the impact of collaborative and intelligent agents. Cavazza et al. [Cavazza et al. 2001] implemented interactive virtual agents for storytelling. They integrated the Hierarchical Task Network with the AO* (And-Or Search) algorithm to decide the virtual agents' behavior. They presented how these virtual agents interacted with users and other virtual agents to synthesize dynamic storytelling. Furthermore, Baker et al. [Baker et al. 2019] presented a methodology to train multiple virtual agents to play the hide-and-seek game using reinforcement learning. They reported that the trained virtual agents cooperated to win the game. Choi et al. [Choi et al. 2023] investigated how the appearance and voice of a collaborative virtual character affected human perception and found that the human-like voice could trigger a positive perception of the virtual character despite the mismatch between its appearance and voice. Walter et al. [Walters et al. 2009] investigated the impact of robots' appearance and embodiment on human perception and reported that their participants rated humanoid robots as having higher perceived intelligence than mechanical robots. Guo et al. [Guo et al. 2024] investigated the impact of self-similarity of appearance and voice between human and virtual characters on human perception. They indicated that the self-similar appearance of the virtual character made participants feel the virtual character was more intelligent. Finally, Krenig and Feigh [Krening and Feigh 2018] investigated agents based on machine learning and indicated that people felt the agents trained by action instruction were more intelligent than others trained by critiques.

2.4 Task Complexity

To design collaboration between humans and agents, it is necessary to consider not only virtual agents but also task complexity [Kiyokawa et al. 2023]. Task complexity has been defined by various components, such as the number of elements [Williams 1999], relationships between tasks [Wood 1986], time pressure [Greitzer 2005], and cognitive demands [Bailey and Scerbo 2007]. Several researchers have investigated the impact of task complexity on human and agent interaction. Stollnberger et al. [Stollnberger et al. 2013] focused on the interdependency of input modality and task complexity. They indicated that the perceived task complexity is highly related to the cognitive workload based on the input modality. Malik and Bilberg [Malik and Bilberg 2019] presented task allocation based on task complexity and indicated that the proposed task allocation enhanced the efficiency of human and robot interaction. Last, Guo et al. [Guo et al. 2020] investigated the impact of the type of conversation and task complexity on human and conversational agent interaction. They reported higher task complexity caused more queries for each task.

2.5 Research Questions

Although previous studies have explored various factors of human and agent interaction, few studies have focused on the intelligence of virtual characters. Also, despite the importance of task complexity in collaboration between humans and agents, the impacts of task complexity on human perceptions and user experiences still lack sufficient understanding. Furthermore, to our knowledge, the interaction effects of virtual characters' intelligence and task complexity have yet to be investigated. Thus, in this paper, we examined five overarching topics, each including sub-questions to understand how the virtual character's intelligence and task complexity during a co-solving task could affect participants' perceptions and user experience:

- **Virtual Character's Intelligence:**
  - **RQ1.1:** How do a virtual character's intelligence and the complexity of the task impact study participants' perceived intelligence ratings?
  - **RQ1.2:** How do a virtual character's intelligence and the complexity of the task impact study participants' intelligence comparison ratings?
- **Perceptual Experience:**
  - **RQ2.1:** How do a virtual character's intelligence and the complexity of the task impact study participants' perceived collaboration rating?
  - **RQ2.2:** How do a virtual character's intelligence and the complexity of the task impact study participants' perceived contribution ratings?
  - **RQ2.3:** How do a virtual character's intelligence and the complexity of the task impact study participants' ratings regarding the virtual character's public awareness?
- **User Experience:**
  - **RQ3.1:** How do a virtual character's intelligence and the complexity of the task impact study participants' attentional allocation ratings?
  - **RQ3.2:** How do a virtual character's intelligence and the complexity of the task impact study participants' task load ratings?
  - **RQ3.3:** How do a virtual character's intelligence and the complexity of the task impact study participants' frustration ratings?
- **Confidence in Performance:**
  - **RQ4.1:** How do a virtual character's intelligence and the complexity of the task impact study participants' self-confidence ratings?

  - **RQ4.2:** How do a virtual character's intelligence and the complexity of the task impact study participants' ratings regarding the virtual character's confidence?
- **Gazing during Co-solving Process:**
  - **RQ5.1:** How do a virtual character's intelligence and the complexity of the task impact study participants' dwell gazing patterns?

2.6 Contribution

So far, several researchers have explored interactions between humans and intelligent virtual characters in virtual reality. However, to our knowledge, this is the first study investigating the impact of a virtual character's intelligence and task complexity on human perceptions and user experience. Thus, in this paper, we contribute to expanding the current knowledge on combining two levels of a virtual character's intelligence and task complexity. We also provide insights to researchers and developers interested in human perceptions and user experience in virtual reality applications where users collaborate to solve tasks with a virtual character.

3 Materials and Methods

3.1 Participants

We performed an *a priori* power analysis using the G*Power software [Faul et al. 2009] with the following settings: a medium effect size [Cohen 2013] of $f = .30$, an $\alpha = .05$ error probability, one group with four repeated measurements, an $r = .50$ correlation among repeated measures, and an $\epsilon = .70$ for non-sphericity correction. The power analysis suggested an $N = 21$ sample size to achieve a .80 power ($1 - \beta$ error probability). We recruited 27 participants through class announcements and emails forwarded to our university's listservs. Our participants (age range: $18 - 52$) were composed of ten males (age: $M = 22.70$, $SD = 10.81$) and 17 females (age: $M = 22.59$, $SD = 5.8$). Most participants reported they had some prior experience with virtual reality applications.

3.2 Implementation

We implemented our virtual reality jigsaw puzzle co-solving application in the Unity (version 2020.3.20) game engine using the Oculus Integration Toolkit. For implementation and experimentation, we used a Dell Alienware computer (Intel i7, NVIDIA GeForce RTX 2080, and 32GB RAM) and Meta Quest 2 as a virtual reality head-mounted display.

We integrated a 3D living room model into our application to provide immersive experiences to our participants (see Figure 2). When we immersed our participants in the virtual environment, they found themselves sitting on a chair in front of a table in the living room. On the table, we placed puzzle pieces to be solved, a puzzle board where these pieces would be placed, and a puzzle goal to provide participants with a clear image of the completed puzzle. Specifically, the puzzle board has a semitransparent texture to let participants place a puzzle piece in the target spot. The puzzle pieces were automatically attached when they were at a certain distance from their target spots on the puzzle board.

On the right side of the participants, we placed a virtual character that solved the puzzles with them (see Figure 3). The virtual

Figure 2: We immersed our participants in a 3D living room environment.

character is a human female 3D model (Female_Adult_01) from Microsoft's Rockebox Avatar library.[1] To enable the virtual character to solve the jigsaw puzzle, we implemented a script that decided the behavior of the virtual character. The behavior was composed of two actions: one is picking up a puzzle piece, and the other is placing it in a specific spot. The implemented script randomly chose an unsolved puzzle piece on the table and decided whether the virtual character placed it in the correct or wrong spot based on the assigned intelligence. Specifically, the virtual character's intelligence is the probability of the virtual character placing a puzzle piece in the correct spot. For example, a virtual character with 100% intelligence always places puzzle pieces in the correct spot. In contrast, the virtual character always places puzzle pieces in the wrong spots when assigned to 0% intelligence (i.e., unintelligent virtual character). The virtual character repeated its routine during the co-solving process with the participant until the puzzle was fully solved. To animate the virtual character according to decisions from the script, we integrated forward and backward reaching inverse kinematics solver [Aristidou and Lasenby 2011] into the virtual character to make it pick up a puzzle piece and place it on a specific spot.

To achieve engaging and realistic user experiences [Feldman et al. 2017], we implemented a short talk that evolved during the co-solving process' beginning, middle, and end. The virtual character asked questions, and our application rendered a set of predefined answers using a graphical user interface (GUI). The participants could answer by clicking one of the answers through the ray from the virtual reality controller. Also, we integrated other predefined speech based on the progress of the co-solving process (e.g., when the co-solving process started, the virtual character said *"This puzzle looks so hard to solve,"* and during the co-solving process, the virtual character said *"It looks easy now," "Let me think where this puzzle piece goes,"* or *"I enjoy solving the puzzle."*). For the dialogs, we generated the speech using a female voice model on PlayHT.[2] Note that we used the same dialogue in all experimental conditions to eliminate introducing additional variables to our experiment and, therefore, to standardize our experimental conditions. We also

[1] https://github.com/microsoft/Microsoft-Rocketbox
[2] https://play.ht/

Figure 3: When we immersed our participants in the virtual environment, they could see the puzzle pieces, the puzzle board on the table, and the virtual character on their right side.

integrated the Salsa LipSync Suite[3] from Unity Asset Store into the virtual character to synthesize its lip sync animation. Finally, we implemented eye blinks and head movements to enhance the realism of the virtual character.

3.3 Experimental Conditions

Following our 2 (intelligence: low intelligence [LI] vs. high intelligence [HI]) × 2 (complexity: low complexity [LC] vs. high complexity [HC]) study design, we created the following experimental conditions:

- **LILC:** low intelligence (0%) and low complexity (25 puzzle pieces);
- **LIHC:** low intelligence (0%) and high complexity (100 puzzle pieces);
- **HILC:** high intelligence (100%) and low complexity (25 puzzle pieces); and
- **HIHC:** high intelligence (100%) and high complexity (100 puzzle pieces).

Based on the assigned intelligence, the virtual character places puzzle pieces in each condition on the wrong spots (intelligence: 0%) or the correct spot (intelligence: 100%). Regarding task complexity, we followed the factors outlined by Kiyokawa et al. [Kiyokawa et al. 2023], specifically defining it based on the number of puzzle pieces and their size. For this study, we assigned 25 puzzle pieces to the low complexity and 100 to the high complexity (see Figure 4 and Figure 5). The size of the puzzle and puzzle board was the same for all conditions. Thus, the puzzle piece size in the low complexity was four times larger than the puzzle piece size in the high complexity conditions.

3.4 Ratings and Measurements

After each condition, we provided the developed survey to explore how the virtual character's intelligence and task complexity impacted our participants. We also recorded application logs during the co-solving process.

[3]https://assetstore.unity.com/packages/tools/animation/salsa-lipsync-suite-148442

3.4.1 Survey. We developed a survey to understand how our participants perceived the virtual character and their user experience during the jigsaw puzzle co-solving process. The survey was composed of six items for perceived intelligence from Moussawi and Koufaris [Moussawi and Koufaris 2019], six items for perceived collaboration from Liu et al. [Liu et al. 2023], two items for public awareness of the virtual character from Govern and Marsch [Govern and Marsch 2001], six items for attentional allocation from Biocca et al. [Biocca et al. 2001], and six items for task load from NASA TLX [Hart 2006]. We also included one item for intelligence comparison, one for perceived contribution, one for frustration, one for self-confidence, and one for the confidence of the virtual character we developed. We used a 7-point Likert scale for all items except NASA TLX. For the NASA TLX, we used its original scale, 21 gradations on the scales. We provide our survey we developed in our supplementary materials.

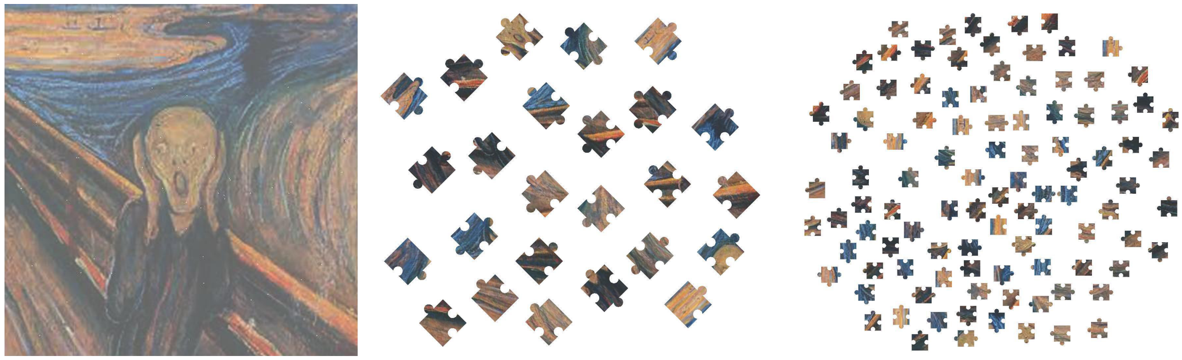
3.4.2 Application Logs. Gaze has been used to interpret human attention [Ajanki et al. 2009; Knight and Simmons 2013; Krogmeier and Mousas 2020, 2021; Li et al. 2017] in various contexts. Thus, we collected application logs of our participants' dwell gazing on the virtual character and objects to understand their interest and attention during co-solving puzzles with the virtual character:

- **Virtual character dwell gazing:** We measured how long (in seconds) participants gazed at the virtual character (face, arm, and torso).
- **Puzzle goal dwell gazing:** We measured how long (in seconds) participants gazed at the puzzle goal.
- **Puzzle pieces dwell gazing:** We measured how long (in seconds) participants gazed at puzzle pieces.

To collect these logs, we integrated a pseudo-gazing methodology into our application. Specifically, we projected a ray from the center of the participant's camera in its forward direction. When the ray collided with one of the targeted virtual objects, it returned the duration of this collision with the name of the collided virtual object. The application returned accumulated dwell gazing durations when participants completed the co-solving process. For our statistical analyses, we normalized these durations by using the total time in each condition.

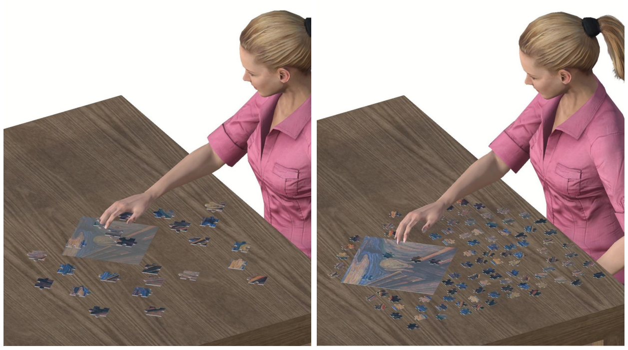
3.5 Procedure

We first invited our participants into our lab space and introduced them to the experiment and the purpose of the study. After they signed the consent form, researchers helped them set up and adjust the virtual reality equipment. Once the participants were ready, we started a tutorial scene to familiarize themselves with the interaction mechanism and the controller. We did so because a prior study showed that tutorials about controllers improve participants' performance and user experience [Kao et al. 2021]. Once our participants indicated they were ready, we started the experiment with the first condition. The order in which our participants experienced each condition was based on the Latin square method [Williams 1949], which balances the conditions to eliminate first-order carryover (residual) effects. Participants solved the jigsaw puzzle task in each condition with our virtual character. Once our participants had completed each condition, they removed the headsets and provided their ratings on the survey. Participants needed to repeat the



(a) (b) (c)

Figure 4: The puzzle we used in our study: (a) completed puzzle, (b) 25 pieces for the low complexity task, and (c) 100 pieces for the high complexity task.



(a) (b)

Figure 5: The virtual character solves the jigsaw puzzle during the (a) low complexity and (b) high complexity task.

process until all four conditions were complete. When the participants completed all conditions, we asked them to leave feedback about their user experiences with the four conditions. Then, the researchers expressed their appreciation and excused the participants. There was at least a 10-minute gap between each experiment session. Participants spent at most 60 minutes in our lab to finish the study.

4 Results

To analyze self-reported ratings, we used a 2×2 factorial experiment design. We used two independent variables, intelligence and complexity, and ten dependent variables from self-reported ratings. Regarding application logs, we used 2×2×3 factorial experiment design to investigate differences among dwell gazing. Specifically, we used three independent variables (intelligence, complexity, and gazes [virtual character vs. puzzle goal vs. puzzle pieces]) and three dependent variables. We examined the normality of our data graphically using Q-Q plots of the residuals and the Shapiro-Wilk test at the 5% level. Our data was normally distributed; thus, we used two-way and three-way repeated measures analysis of variance (ANOVA), respectively, for the mentioned measurement. We provide detailed results for the self-reported data and application logs in Table 1 and Table 2, respectively.

4.1 Self-reported Ratings

Perceived Intelligence. Our simple main effect analysis on the intelligence factor indicated that our participants rated the perceived intelligence lower when we exposed them to the low intelligence ($M = 1.70$, $SE = .27$) than the high intelligence ($M = 5.23$, $SE = .27$) conditions (Wilk's $\Lambda = .329$, $F[1, 26] = 53.132$, $p < .001$, $\eta_p^2 = .671$). However, we did not find a statistically significant result for the complexity factor (Wilk's $\Lambda = .908$, $F[1, 26] = 2.646$, $p = .116$, $\eta_p^2 = .092$) and for the intelligence×complexity interaction (Wilk's $\Lambda = .936$, $F[1, 26] = 1.771$, $p = .195$, $\eta_p^2 = .064$).

Intelligence Comparison. Our simple main effect analysis on the intelligence factor indicated that our participants rated the intelligence comparison lower when we exposed them to the low intelligence ($M = 1.70$, $SE = .17$) than the high intelligence ($M = 3.47$, $SE = .31$) conditions (Wilk's $\Lambda = .390$, $F[1, 26] = 40.647$, $p < .001$, $\eta_p^2 = .610$). However, we did not find a statistically significant result for the complexity factor (Wilk's $\Lambda = .994$, $F[1, 26] = .146$, $p = .705$, $\eta_p^2 = .006$) and the intelligence×complexity interaction (Wilk's $\Lambda = .917$, $F[1, 26] = 2.359$, $p = .137$, $\eta_p^2 = .083$).

Perceived Collaboration. Our simple main effect analysis on the intelligence factor indicated that our participants rated the perceived collaboration lower when we exposed them to the low intelligence ($M = 3.97$, $SE = .26$) than the high intelligence ($M = 5.57$, $SE = .23$) conditions (Wilk's $\Lambda = .264$, $F[1, 26] = 72.669$, $p < .001$, $\eta_p^2 = .736$). Moreover, our main effect analysis on the complexity factor showed that our participants rated the perceived collaboration higher when we exposed them to the low complexity ($M = 5.03$, $SE = .22$) than the high complexity ($M = 4.52$, $SE = .26$) conditions (Wilk's $\Lambda = .616$, $F[1, 26] = 16.206$, $p < .001$, $\eta_p^2 = .384$). However, we did not find a statistically significant result for the intelligence×complexity interaction (Wilk's $\Lambda = .983$, $F[1, 26] = .444$, $p = .511$, $\eta_p^2 = .017$).

Table 1: Detailed results of our study for the self-reported ratings (we present significant results with bold font).

	(1)		**(2)**		**(3)**		**(4)**		**(5)**		**(6)**		**(7)**		**(8)**		**(9)**		**(10)**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
LILC	3.40	1.43	1.59	.97	4.27	1.24	2.48	1.22	2.31	1.32	3.00	.71	10.31	4.15	4.19	1.92	5.44	1.87	3.85	1.88
LIHC	3.35	1.57	1.81	1.11	3.67	1.67	2.81	2.02	1.69	1.12	3.11	.85	9.84	4.87	4.00	2.25	3.70	2.18	4.70	2.00
HILC	5.46	1.56	3.59	1.99	5.79	1.27	5.93	1.49	3.72	2.16	3.27	1.16	7.09	2.47	1.74	1.13	6.00	1.73	5.41	1.36
HIHC	4.99	1.49	3.22	1.63	5.36	1.28	5.52	1.55	2.31	2.03	3.25	.95	9.42	4.10	2.22	1.48	4.78	2.01	5.19	1.76
Main Effect (Intelligence)																				
F		**53.132**		**40.647**		**72.669**		**82.049**		**15.822**		.912		**10.667**		**40.026**		**6.699**		**8.543**
p		**<.001**		**<.001**		**<.001**		**<.001**		**<.001**		.348		**.003**		**<.001**		**.016**		**.007**
η_p^2		**.671**		**.610**		**.736**		**.759**		**.378**		.034		**.291**		**.606**		**.205**		**.247**
Main Effect (Complexity)																				
F		2.646		.146		**16.206**		.026		**12.564**		.085		3.587		.217		**33.32**		2.276
p		.116		.705		**<.001**		.873		**.002**		.772		.069		.645		**<.001**		.143
η_p^2		.092		.006		**.384**		.001		**.326**		.003		.121		.008		**.562**		.08
Interaction Effect (Intelligence×Complexity)																				
F		1.771		2.359		.444		3.178		.202		.384		**25.911**		2.4		1.16		**5.709**
p		.195		.137		.511		.086		.657		.541		**<.001**		.133		.291		**.024**
η_p^2		.064		.083		.017		.109		.008		.015		**.499**		.085		.043		**.18**
Intelligence df = 1, Complexity df = 1, Interaction df = 1, Error df = 26																				
(1) Perceived Intelligence, (2) Intelligence Comparison, (3) Perceived Collaboration, (4) Perceived Contribution, (5) Public Awareness of the Virtual Character, (6) Attentional Allocation, (7) Task Load, (8) Frustration, (9) Self-Confidence, (10) Confidence of the Virtual Character.																				

Perceived Contribution. Our simple main effect analysis on the intelligence factor indicated that our participants rated the perceived contribution lower when we exposed them to the low intelligence ($M = 2.65$, $SE = .28$) than the high intelligence ($M = 5.72$, $SE = .26$) conditions (Wilk's $\Lambda = .241$, $F[1, 26] = 82.049$, $p < .001$, $\eta_p^2 = .759$). However, we did not find a statistically significant result for the complexity factor (Wilk's $\Lambda = .999$, $F[1, 26] = .026$, $p = .873$, $\eta_p^2 = .001$) and the intelligence×complexity interaction (Wilk's $\Lambda = .891$, $F[1, 26] = 3.178$, $p = .086$, $\eta_p^2 = .109$).

Public Awareness of the Virtual Character. Our simple main effect analysis on the intelligence factor indicated that our participants rated the public awareness of the virtual character lower when we exposed them to the low intelligence ($M = 2.00$, $SE = .22$) than the high intelligence ($M = 3.47$, $SE = .38$) conditions (Wilk's $\Lambda = .622$, $F[1, 26] = 15.822$, $p < .001$, $\eta_p^2 = .378$). Moreover, our main effect analysis on the complexity factor showed that our participants rated the public awareness of the virtual character higher when we exposed them to the low complexity ($M = 3.02$, $SE = .29$) than the high complexity ($M = 2.45$, $SE = .24$) conditions (Wilk's $\Lambda = .674$, $F[1, 26] = 12.564$, $p = .002$, $\eta_p^2 = .326$). However, we did not find a statistically significant result for the intelligence×complexity interaction (Wilk's $\Lambda = .992$, $F[1, 26] = .202$, $p = .657$, $\eta_p^2 = .008$).

Attentional Allocation. We did not find a statistically significant result on our participants' attentional allocation rating for the intelligence (Wilk's $\Lambda = .966$, $F[1, 26] = .912$, $p = .348$, $\eta_p^2 = .034$) and complexity (Wilk's $\Lambda = .997$, $F[1, 26] = .085$, $p = .772$, $\eta_p^2 = .003$) factors, and intelligence×complexity interaction (Wilk's $\Lambda = .985$, $F[1, 26] = .384$, $p = .541$, $\eta_p^2 = .015$).

Task Load. Our simple main effect analysis on the intelligence factor indicated that our participants rated the task load higher when we exposed them to the low intelligence ($M = 10.08$, $SE = .84$) than the high intelligence ($M = 8.26$, $SE = .57$) conditions (Wilk's $\Lambda = .709$, $F[1, 26] = 10.667$, $p = .003$, $\eta_p^2 = .291$). We did not find a statistically significant result for the complexity factor (Wilk's $\Lambda = .879$, $F[1, 26] = 3.587$, $p = .069$, $\eta_p^2 = .121$). However, we found a statistically significant intelligence×complexity interaction effect (Wilk's $\Lambda = .501$, $F[1, 26] = 25.911$, $p < .001$, $\eta_p^2 = .499$), indicating that, in the presence of the low intelligence, participants rated the task load higher.

Frustration. Our simple main effect analysis on the intelligence factor indicated that our participants rated their frustration higher when we exposed them to the low intelligence ($M = 4.09$, $SE = .33$) than the high intelligence ($M = 1.98$, $SE = .21$) conditions (Wilk's $\Lambda = .394$, $F[1, 26] = 40.026$, $p < .001$, $\eta_p^2 = .606$). However, we did not find a statistically significant result for the complexity factor (Wilk's $\Lambda = .992$, $F[1, 26] = .217$, $p = .645$, $\eta_p^2 = .008$) and intelligence×complexity interaction (Wilk's $\Lambda = .915$, $F[1, 26] = 2.400$, $p = .133$, $\eta_p^2 = .085$).

Self-confidence. Our simple main effect analysis on the intelligence factor indicated that our participants rated their self-confidence lower when we exposed them to the low intelligence ($M = 4.57$, $SE = .34$) than the high intelligence ($M = 5.39$, $SE = .33$) conditions (Wilk's $\Lambda = .795$, $F[1, 26] = 6.699$, $p = .016$, $\eta_p^2 = .205$). Moreover, our main effect analysis on the complexity factor showed that our participants rated their self-confidence higher when we exposed them to the low complexity ($M = 5.72$, $SE = .31$) than the high complexity (M = 4.24, SE = .33) conditions (Wilk's Λ = .438, $F[1, 26]$ = 33.320, p < .001, η_p^2 = .562). However, we did not find a statistically significant result for the intelligence×complexity interaction (Wilk's Λ = .957, $F[1, 26]$ = 1.160, p = .291, η_p^2 = .043).

Confidence of the Virtual Character. Our simple main effect analysis on the intelligence factor indicated that our participants rated the confidence of the virtual character lower when we exposed them to the low intelligence (M = 4.28, SE = .33) than the high intelligence (M = 5.30, SE = .27) conditions (Wilk's Λ = .753, $F[1, 26]$ = 8.543, p = .007, η_p^2 = .247). We did not find a statistically significant result for the complexity factor (Wilk's Λ = .920, $F[1, 26]$ = 2.276, p = .143, η_p^2 = .080). However, we found a statistically significant intelligence×complexity interaction effect (Wilk's Λ = .820, $F[1, 26]$ = 5.709, p = .024, η_p^2 = .180), indicating that, in the presence of high intelligence, participants rated the confidence of the virtual character higher.

4.2 Application Logs

Dwell Gazing. We did not find a statistically significant result on collected dwell gazing for the intelligence (Wilk's Λ = .939, $F[1, 26]$ = 1.691, p = .205, η_p^2 = .061) and complexity (Wilk's Λ = .982, $F[1, 26]$ = .473, p = .498, η_p^2 = .018) factors. However, the simple main effect analysis on the gazes factor indicated a statistically significant result (Wilk's Λ = .149, $F[1, 25]$ = 71.150, p < .001, η_p^2 = .851). The post hoc pairwise comparison showed that participants gazed at the puzzle goal less time (M = .01, SE = .00) than the virtual character (M = .11, SE = .01; p < .001) and puzzle pieces (M = .10, SE = .01; p < .001). We did not find statistically significant interaction effects for intelligence×complexity (Wilk's Λ = .866, $F[1, 26]$ = 4.018, p = .056, η_p^2 = .134), intelligence×gaze (Wilk's Λ = .931, $F[1, 25]$ = .924, p = .410, η_p^2 = .069), and complexity×gaze (Wilk's Λ = .956, $F[1, 25]$ = .573, p = .571, η_p^2 = .044). However, we found a significant intelligence×complexity×gaze interaction effect (Wilk's Λ = .700, $F[1, 25]$ = 5.370, p = .011, η_p^2 = .300), indicating that participants gazed at either the virtual character or puzzle pieces rather than the puzzle goal, regardless of the intelligence and complexity factors.

4.3 Qualitative Data

We asked our participants to leave feedback about their experience in our study. We identified two main themes: the virtual character's intelligence and how they enjoyed the virtual reality experience.

Most participants mentioned the differences in the virtual character's intelligence. P6 wrote: *"...I noticed that in the first two rounds* (with low intelligence)*, the other character was eager to complete the puzzle but would put the pieces in the wrong places. She started putting them in the correct places the last two rounds* (with high intelligence)*, which made us reach our goal easier."* P17 mentioned: *"In the beginning* (with low intelligence and low task complexity)*, it did not seem like the AI* (virtual character) *was functioning very well. But then, as we went through the different levels, it became easier to work with."* P24 reported: *"1 and 2* (low intelligence conditions) *were very annoying to work with, but 3 and 4* (high intelligence conditions) *were extremely better."* P27 wrote: *"When the virtual person was functioning properly, I felt more of a joint effort. When the virtual person was not helpful, it was so distracting, and I spent more time fixing her issue than working on the task."*

Also, some participants expressed how they enjoyed our study. P11 reported: *"I really liked the experience. It was very interactive and fun."* P6 mentioned: *"This was very fun to do and a very interesting experience..."* P16 wrote: *"This was interesting..."* P20 reported: *"I liked this research experiment."* Finally, P7 wrote: *"Super fun."*

5 Discussion

Our participants responded to the provided questionnaires, which asked about their perceptions of virtual characters and user experiences. The statistical analyses uncovered several interesting findings, which we discuss in the following subsections.

5.1 Virtual Character's Intelligence

We aimed our study to understand how the virtual character's intelligence and the task complexity impacted our participants' perceptions of the virtual character's intelligence. We found significant differences in the results of both perceived intelligence (**RQ1.1**) and intelligence comparison (**RQ1.2**). These findings indicated that perceived intelligence increased when the virtual character's intelligence possessed high value. These results align with our participants' feedback, who mentioned that they noticed a difference in the virtual characters' intelligence in the experimental conditions. Researchers have reported several factors of perceived intelligence, such as anthropomorphism [Lee et al. 2015] or animacy [Bartneck et al. 2009], and we think that the virtual character's intelligence can be another factor. However, we did not find significant results from the complexity factor or intelligence×complexity interaction. Thus, we cannot argue that these factors impact users' perceptions of the virtual character's intelligence.

5.2 Perceptual Experience

To explore our participants' perceptual experience from combinations of the virtual character's intelligence and task complexity, we asked questions about perceived collaboration and contribution, as well as public awareness of the virtual character. We found significant results from perceived collaboration (**RQ2.1**). These findings indicated that our participants felt they collaborated more when either the virtual character was more intelligent or the task was less complex. Similarly, based on the qualitative data, participants reported a joint effort when the virtual character was assigned high intelligence. According to Thomson et al. [Thomson et al. 2009], "trust is a central component of collaboration." Based on this premise, these findings agree with Huang et al. [Huang et al. 2021], reporting their participants rated higher trust from interacting with the high-ability robot and lower trust from complex tasks. We think our participants felt the more intelligent virtual characters had a higher capability to solve the puzzles, triggering higher ratings of perceived collaboration.

In the case of perceived contribution (**RQ2.2**), we found a significant result reporting that our participants felt the virtual character contributed more in the co-solving process when we exposed them to higher intelligence conditions. This finding expands Lampe's and Chatila's study [Lampe and Chatila 2006], stating mission success

Table 2: Detailed results of our study for application logs (we present significant results with bold font).

	Virtual Character				Puzzle Goal				Puzzle Pieces			
	M	*SD*	*Min*	*Max*	*M*	*SD*	*Min*	*Max*	*M*	*SD*	*Min*	*Max*
LILC	.13	.08	.02	.41	.01	.02	.00	.10	.11	.11	.00	.53
LIHC	.11	.06	.04	.36	.02	.03	.00	.16	.10	.08	.01	.25
HILC	.09	.04	.00	.20	.02	.03	.00	.14	.10	.08	.00	.26
HIHC	.12	.05	.02	.33	.01	.02	.00	.09	.11	.09	.00	.35

Main Effect

	Intelligence	Complexity	Gaze
F	1.691	.982	**.149**
p	.205	.498	**<.001**
η_p^2	.061	.018	**.851**

Interaction Effect

	Intelligence×Complexity	Intelligence×Gaze	Complexity×Gaze	Intelligence×Complexity×Gaze
F	.866	.931	.956	**.700**
p	.056	.410	.571	**.011**
η_p^2	.134	.069	.044	**.300**

Intelligence df = 1 (Error df = 26), Complexity df = 1 (Error df = 26), Gaze df = 2 (Error df = 25)
Intelligence×Complexity df = 1 (Error df = 26), Intelligence×Gaze df = 1 (Error df = 26)
Complexity×Gaze df = 1 (Error df = 26), Intelligence×Complexity×Gaze df = 2 (Error df = 25)

rate as a metric to evaluate the performance of robots by indicating this metric can be applied to virtual characters.

We found significant results from public awareness of the virtual character (**RQ2.3**). Based on our findings, we argue our participants felt the virtual character was more aware when either the virtual character was more intelligent or the task was less complex. These findings agree with Hayes et al. [Hayes et al. 2016], reporting that their participants assumed that a robot understood their instruction until it made the first mistake. We think our participants thought a more intelligent virtual character became more aware because it did not make errors during co-solving puzzles. Additionally, these findings build upon Ijaz et al.'s [Ijaz et al. 2011] study, which highlights environmental understanding as crucial for the virtual character's awareness. Our results suggest that the virtual character's intelligence and task complexity can also enhance public awareness of virtual characters.

Last, we did not find significant intelligence×complexity results in all perceptual experience ratings. This indicates that the impact of the virtual character's intelligence on our participants' perceptual experiences was consistent across all conditions. It also shows that task complexity did not necessarily moderate how the virtual character's intelligence affected our participants' perceptual experiences.

5.3 User Experiences

We asked questions concerning attentional allocation, task load, and frustration to explore how the virtual character's intelligence and task complexity impact our participants' user experience. We did not find significant results regarding attentional allocation (**RQ3.1**). We argue that unlike threats, such as angry faces [Schrammel et al. 2009], the intelligence and task complexity did not trigger participants' attentional allocation.

As for task load (**RQ3.2**), we found a significant result from intelligence, and this finding showed that our participants rated their task load lower when they co-solved puzzles with more intelligent virtual characters. This finding aligns with Rabby et al. [Rabby et al. 2019], which reported that lower robot performance increased participants' cognitive workload. Perhaps our participants felt there was a need for more mental and physical demands when interacting with a less intelligent virtual character because they needed to fix all errors made by the virtual character. Furthermore, we found an interaction effect on task load, reporting that our participants experienced higher task load when interacting with a less intelligent virtual character than a more intelligent one, even though they co-solved the same puzzles. This finding shows that the virtual character's intelligence moderated how task complexity impacted task load.

We also found a significant result from frustration (**RQ3.3**), and this finding showed us that our participants became more frustrated when they co-solved puzzles with less intelligent virtual characters. This finding is consistent with Myers et al. [Myers et al. 2018] reporting that system errors in the voice user interface frustrated their participants. We think our participants felt more frustrated when they observed that the virtual character placed puzzle pieces in the wrong spot. However, we could not find significant results from either task complexity or the interaction effect of intelligence and task complexity, so we cannot argue that task complexity impacted our participants' frustration during co-solving puzzles with the virtual character.

5.4 Confidence in Performance

We explored how our participants perceived confidence in themselves and the virtual character. We found significant results from perceived self-confidence, indicating our participants felt more confident (**RQ4.1**) from either their exposure to a more intelligent virtual character or due to effortless task complexity. These findings align with those of Rienovita et al. [Rienovita et al. 2017], who stated that participants' self-esteem was encouraged when agents helped them. Our participants felt more confident when co-solved the puzzle with a more intelligent virtual character because it helped them solve the puzzle, so our participants needed less time to complete the task. This finding expands Huang et al. [Huang et al. 2021], which reported that the study participants' self-construal was related to trust by indicating that task complexity also discourages our participants' self-confidence.

We also found significant results regarding the confidence of the virtual character (**RQ4.2**). Our findings showed that participants rated the virtual character's confidence lower when they co-solved puzzles with less intelligent virtual characters. This expands on previous research by Thaler et al. [Thaler et al. 2020], which focused on the impact of walking motion on the virtual character's confidence, indicating that intelligence can also play a role. Additionally, we observed an interaction effect on the virtual character's confidence, revealing that participants felt more confident in the virtual character when it was more intelligent, even when solving the same puzzles. This suggests that the virtual character's intelligence had a more substantial influence on participants' perception of its confidence than task complexity. This finding is relatively novel and contributes to the existing literature on the topic.

5.5 Gazing during Co-solving Process

We did not find significant results for the impact of virtual characters' intelligence and task complexity on our participants' dwell gazing patterns (**RQ5.1**). However, we found that our participants gazed at the virtual character and puzzle pieces more than the puzzle goal. This finding aligns with Sidenmark and Lundström's study [Sidenmark and Lundström 2019] investigating the relationship between hand and gaze behavior during hand interactions in a virtual environment. They reported that interactions demanding precision induced gaze fixation on the interacted object. Our participants considered the puzzle pieces interactable objects, whereas the puzzle goal was not. Furthermore, we expand Sidenmark and Ludström's study by indicating that interaction partners (the virtual character) can be another factor inducing gaze fixation during hand interactions.

5.6 Limitations

Designing and implementing a virtual reality experience to co-solve jigsaw puzzles with intelligent virtual characters requires careful consideration of various components. Although our participants did not experience any issues during their virtual reality experiences, we want to report several limitations. Note that these limitations do not invalidate our findings but rather provide clues to improve the virtual reality experience and directions for future studies.

First, we should consider more factors, such as puzzle piece distributions, in the design of task complexity. Although the difference in task complexity between the experimental conditions was apparent, we did not find a significant result concerning our participants' frustration rating between conditions with low and high complexity tasks. This suggests an improvement in our understanding of the factors that are mandatory for designing the task complexity in the co-solving puzzle process.

Second, in our application, we did not implement highly realistic animations to our virtual character. The virtual character could access any unsolved puzzle pieces on the table through an inverse kinematics solver. However, the virtual character's hand animations for picking up and placing puzzle pieces were absent, making the virtual character look less realistic. Integrating these animations into the virtual character would make our participants' virtual reality experience more lifelike.

Third, the virtual character's puzzle-solving behaviors were driven by random variables rather than through a sophisticated puzzle-solving strategy. The script randomly selected one from the unsolved pieces as the targeted puzzle piece when the virtual character accessed and picked up a puzzle piece. Additionally, the virtual character randomly placed the puzzle pieces on the puzzle board when solving it incorrectly. Due to these limitations, the puzzle-solving behaviors of the virtual character might look less human-like, potentially negatively impacting participants' perceptions and user experiences.

Last, as we mentioned previously, our pseudo-gazing methodology returned what was at the center of the participant's perspective. However, the collected data needed to be more precise to track what participants gazed at and focused on during the co-solving process. Thus, using other tools to track participants' gaze, such as a high-end eye-tracker, will help collect more accurate data and understand participants' co-solving activities.

5.7 Implications for Future Studies

Providing a cooperative atmosphere in collaborative interactions between humans and virtual characters is necessary. According to our results, the virtual character's intelligence and task complexity impacted the rating of perceived collaboration. Specifically, the higher intelligence of the virtual character and low-complexity task provided the highest rating of perceived collaboration. Thus, we suggest the combination of more intelligent virtual characters and less complex tasks when the purpose of the interaction between humans and virtual characters is to provide collaborative experiences.

If the interaction's priority is user experiences, it would be necessary to prevent negative experiences. Our study indicated that the lower intelligence of the virtual character triggered a higher task load, more frustration, and less self-confidence. Therefore, to improve user experiences, we recommend implementing highly intelligent virtual characters when they are essential.

Finally, if enhancing the perception of the virtual character's supportiveness is deemed necessary, human perception of the virtual character would be crucial. Our findings suggest that when the virtual character displayed higher intelligence, participants felt more confident in its abilities and perceived it as more intelligent, contributing significantly to the puzzle-solving process. However, task complexity did not affect participants contrary to the impact of the virtual character's intelligence. Therefore, we assert that a more intelligent virtual character can be perceived as more supportive, irrespective of task complexity.

6 Conclusions and Future Work

Collaborating with intelligent virtual characters in virtual reality has been explored from several perspectives concerning the design of interactions and capabilities of such characters. However, although few studies have focused on task complexity, to our knowledge, studies have yet to consider the impact of virtual character intelligence and task complexity on human perceptions and user experiences. Therefore, we explored how the virtual character's intelligence and task complexity affected our participants' perceptions and user experiences. The statistical analyses revealed several interesting main and interaction effects. Our findings indicated the importance of the virtual character's intelligence and task complexity in human-agent interaction. Thus, this paper expands current knowledge in human-agent interaction and provides guidelines for improving collaborative interaction between humans and virtual characters.

We identified several limitations in our study, which offer the potential to enhance future studies. Based on these limitations, we aim to investigate the design of task complexities further and extend our research to related topics, such as task allocation. Moreover, we will examine strategies of puzzle-solving processes and apply them to the virtual character to explore how different strategies affect human perception of the virtual character and user experiences.

References

Antti Ajanki, David R Hardoon, Samuel Kaski, Kai Puolamäki, and John Shawe-Taylor. 2009. Can eyes reveal interest? Implicit queries from gaze patterns. *User Modeling and User-Adapted Interaction* 19 (2009), 307–339.

Sean Andrist, Michael Gleicher, and Bilge Mutlu. 2017. Looking coordinated: Bidirectional gaze mechanisms for collaborative interaction with virtual characters. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 2571–2582.

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

Nathan R Bailey and Mark W Scerbo. 2007. Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science* 8, 4 (2007), 321–348.

Bowen Baker, Ingmar Kanitscheider, Todor Markov, Yi Wu, Glenn Powell, Bob McGrew, and Igor Mordatch. 2019. Emergent tool use from multi-agent autocurricula. *arXiv preprint arXiv:1909.07528* (2019).

Christoph Bartneck, Takayuki Kanda, Omar Mubin, and Abdullah Al Mahmud. 2009. Does the design of a robot influence its animacy and perceived intelligence? *International Journal of Social Robotics* 1 (2009), 195–204.

Rachel KE Bellamy, Sean Andrist, Timothy Bickmore, Elizabeth F Churchill, and Thomas Erickson. 2017. Human-agent collaboration: Can an agent be a partner?. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 1289–1294.

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.

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.

Nicholas R Burns, Michael D Lee, and Douglas Vickers. 2006. Are individual differences in performance on perceptual and cognitive optimization problems determined by general intelligence? *The Journal of Problem Solving* 1, 1 (2006), 3.

Marc Cavazza, Fred Charles, and Steven J Mead. 2001. Agents' interaction in virtual storytelling. In *Intelligent Virtual Agents: Third International Workshop, IVA 2001 Madrid, Spain, September 10–11, 2001 Proceedings 3*. Springer, 156–170.

Minsoo Choi, Siqi Guo, Alexandros Koilias, Matias Volonte, Dominic Kao, and Christos Mousas. 2024. Exploring the Effects of Self-correction Behavior of an Intelligent Virtual Character during a Jigsaw Puzzle Co-solving Task. *ACM Transactions on Interactive Intelligent Systems* (2024).

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.

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.

Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. routledge.

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.

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.

Sara Salevati Feldman, Ozge Nilay Yalcin, and Steve DiPaola. 2017. Engagement with artificial intelligence through natural interaction models. In *Electronic Visualisation and the Arts (EVA 2017)*. BCS Learning & Development.

Patrick Fissler, Olivia Caroline Küster, Daria Laptinskaya, Laura Sophia Loy, Christine AF Von Arnim, and Iris-Tatjana Kolassa. 2018. Jigsaw puzzling taps multiple cognitive abilities and is a potential protective factor for cognitive aging. *Frontiers in aging neuroscience* 10 (2018), 408085.

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.

Vihanga Gamage and Cathy Ennis. 2018. Examining the effects of a virtual character on learning and engagement in serious games. In *Proceedings of the 11th ACM SIGGRAPH Conference on Motion, Interaction and Games*. 1–9.

Jennifer Goetz, Sara Kiesler, and Aaron Powers. 2003. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003*. Ieee, 55–60.

John M Govern and Lisa A Marsch. 2001. Development and validation of the situational self-awareness scale. *Consciousness and cognition* 10, 3 (2001), 366–378.

Frank L Greitzer. 2005. Toward the development of cognitive task difficulty metrics to support intelligence analysis research. In *Fourth IEEE Conference on Cognitive Informatics, 2005.(ICCI 2005)*. IEEE, 315–320.

Manuel Guimarães, Rui Prada, Pedro A Santos, João Dias, Arnav Jhala, and Samuel Mascarenhas. 2020. The impact of virtual reality in the social presence of a virtual agent. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. 1–8.

Jingya Guo, Da Tao, and Chen Yang. 2020. The effects of continuous conversation and task complexity on usability of an AI-based conversational agent in smart home environments. In *Man–Machine–Environment System Engineering: Proceedings of the 19th International Conference on MMESE*. Springer, 695–703.

Siqi Guo, Minsoo Choi, Dominic Kao, and Christos Mousas. 2024. Collaborating with my Doppelgänger: The Effects of Self-similar Appearance and Voice of a Virtual Character during a Jigsaw Puzzle Co-solving Task. *Proceedings of the ACM on computer graphics and interactive techniques* 7, 1 (2024), 1–23.

Nader Hanna, Deborah Richards, et al. 2015. The Impact of Virtual Agent Personality on a Shared Mental Model with Humans during Collaboration.. In *Aamas*. 1777–1778.

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.

Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

Cory J Hayes, Maryam Moosaei, and Laurel D Riek. 2016. Exploring implicit human responses to robot mistakes in a learning from demonstration task. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 246–252.

Hanjing Huang, Pei-Luen Patrick Rau, and Liang Ma. 2021. Will you listen to a robot? Effects of robot ability, task complexity, and risk on human decision-making. *Advanced Robotics* 35, 19 (2021), 1156–1166.

Kiran Ijaz, Anton Bogdanovych, and Simeon Simo. 2011. Enhancing the believability of embodied conversational agents through environment-, self-and interaction-awareness. In *Conferences in Research and Practice in Information Technology Series*.

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.

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.

Takuya Kiyokawa, Naoki Shirakura, Zhenting Wang, Natsuki Yamanobe, Ixchel G Ramirez-Alpizar, Weiwei Wan, and Kensuke Harada. 2023. Difficulty and complexity definitions for assembly task allocation and assignment in human–robot collaborations: A review. *Robotics and Computer-Integrated Manufacturing* 84 (2023), 102598.

Heather Knight and Reid Simmons. 2013. Estimating human interest and attention via gaze analysis. In *2013 IEEE International Conference on Robotics and Automation*. IEEE, 4350–4355.

Samantha Krening and Karen M Feigh. 2018. Characteristics that influence perceived intelligence in AI design. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 62. SAGE Publications Sage CA: Los Angeles, CA, 1637–1641.

Claudia Krogmeier and Christos Mousas. 2020. Eye fixations and electrodermal activity during low-budget virtual reality embodiment. *Computer Animation and Virtual Worlds* 31, 4-5 (2020), e1941.

Claudia Krogmeier and Christos Mousas. 2021. Evidence for a relationship between self-avatar fixations and perceived avatar similarity within low-cost virtual reality embodiment. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 127–134.

Luchcha Lam, Minsoo Choi, Magzhan Mukanova, Klay Hauser, Fangzheng Zhao, Richard Mayer, Christos Mousas, and Nicoletta Adamo-Villani. 2023. Effects of body type and voice pitch on perceived audio-visual correspondence and believability of virtual characters. In *ACM Symposium on Applied Perception 2023*. 1–11.

Alexandre Lampe and Raja Chatila. 2006. Performance measure for the evaluation of mobile robot autonomy. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006*. IEEE, 4057–4062.

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.

Michael Lewis. 1998. Designing for human-agent interaction. *AI Magazine* 19, 2 (1998), 67–67.

Yixuan Li, Pingmei Xu, Dmitry Lagun, and Vidhya Navalpakkam. 2017. Towards measuring and inferring user interest from gaze. In *Proceedings of the 26th International Conference on World Wide Web Companion*. 525–533.

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.

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.

Ali Ahmad Malik and Arne Bilberg. 2019. Complexity-based task allocation in human-robot collaborative assembly. *Industrial Robot: the international journal of robotics research and application* 46, 4 (2019), 471–480.

Reza Moradinezhad and Erin T Solovey. 2021. Investigating trust in interaction with inconsistent embodied virtual agents. *International Journal of Social Robotics* 13, 8 (2021), 2103–2118.

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.

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.

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

Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for how users overcome obstacles in voice user interfaces. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–7.

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.

Donald A Norman. 1994. How might people interact with agents. *Commun. ACM* 37, 7 (1994), 68–71.

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.

Khurram Monir Rabby, Mubbashar Khan, Ali Karimoddini, and Steven Xiaochun Jiang. 2019. An effective model for human cognitive performance within a human-robot collaboration framework. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE, 3872–3877.

Ellina Rienovita, Masashi Taniguchi, Masato Kawahara, Yugo Hayashi, and Yugo Takeuchi. 2017. Effect of Human Agent Interaction Improves Self-Esteem and Students' Motivation. In *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*. IEEE, 673–680.

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.

Allison Sauppé and Bilge Mutlu. 2015. Effective task training strategies for human and robot instructors. *Autonomous Robots* 39 (2015), 313–329.

Christina Schmidbauer, Sebastian Schlund, Tudor B Ionescu, and Bernd Hader. 2020. Adaptive task sharing in human-robot interaction in assembly. In *2020 IEEE international conference on industrial engineering and engineering management (IEEM)*. IEEE, 546–550.

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.

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

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.

Ludwig Sidenmark and Anders Lundström. 2019. Gaze behaviour on interacted objects during hand interaction in virtual reality for eye tracking calibration. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. 1–9.

Alexandra Sierra Rativa, Marie Postma, and Menno Van Zaanen. 2020. The influence of game character appearance on empathy and immersion: Virtual non-robotic versus robotic animals. *Simulation & Gaming* 51, 5 (2020), 685–711.

Gerald Stollnberger, Astrid Weiss, and Manfred Tscheligi. 2013. " The harder it gets" Exploring the interdependency of input modalities and task complexity in human-robot collaboration. In *2013 IEEE RO-MAN*. IEEE, 264–269.

Anne Thaler, Andreas Bieg, Naureen Mahmood, Michael J Black, Betty J Mohler, and Nikolaus F Troje. 2020. Attractiveness and confidence in walking style of male and female virtual characters. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 678–679.

Ann Marie Thomson, James L Perry, and Theodore K Miller. 2009. Conceptualizing and measuring collaboration. *Journal of public administration research and theory* 19, 1 (2009), 23–56.

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.

Michael L Walters, Kheng Lee Koay, Dag Sverre Syrdal, Kerstin Dautenhahn, and René Te Boekhorst. 2009. Preferences and perceptions of robot appearance and embodiment in human-robot interaction trials. *Procs of New Frontiers in Human-Robot Interaction* (2009).

Isaac Wang, Jesse Smith, and Jaime Ruiz. 2019. Exploring virtual agents for augmented reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

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

Terry M Williams. 1999. The need for new paradigms for complex projects. *International journal of project management* 17, 5 (1999), 269–273.

Robert E Wood. 1986. Task complexity: Definition of the construct. *Organizational behavior and human decision processes* 37, 1 (1986), 60–82.

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

Fu-Chia Yang, Christos Mousas, and Nicoletta Adamo. 2022. Holographic sign language avatar interpreter: A user interaction study in a mixed reality classroom. *Computer Animation and Virtual Worlds* 33, 3-4 (2022), e2082.

Boram Yoon, Hyung-il Kim, Gun A Lee, Mark Billinghurst, and Woontack Woo. 2019. The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, 547–556.

Zhenjie Zhao and Xiaojuan Ma. 2020. Situated learning of soft skills with an interactive agent in virtual reality via multimodal feedback. In *Adjunct Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 25–27.

Katja Zibrek and Rachel McDonnell. 2014. Does render style affect perception of personality in virtual humans?. In *Proceedings of the ACM symposium on applied perception*. 111–115.