Seeing is not Thinking: Testing Capabilities of VR to Promote Perspective-Taking

Eugene Kukshinov, Federica Gini, Anchit Mishra, Nicholas Bowman, Brendan Rooney, and Lennart E. Nacke.

Abstract—Virtual Reality (VR) technologies offer compelling experiences by allowing users to immerse themselves in simulated environments interacting through avatars. However, despite its ability to evoke emotional responses, and seeing 'through the eyes' of the displayed other, it remains unclear to what extent VR actually fosters perspective-taking (PT) or thinking about others' thoughts and feelings. It might be that the common belief that one can "become someone else" through VR is misleading, and that engaging situations through a different viewpoint does not produce a different cognitive standpoint. To test this, we conducted a 2 (perspective, first-person or third-person) by 2 (perspective-taking task or no task) to examine effects on perspective taking, measured via audio-recordings afforded by the think-aloud protocol. Our data demonstrate that while first-person perspective (1PP) facilitates perceived embodiment, it has no appreciable influence on perspective-taking. Regardless of 1PP or third-person perspective (3PP), perspective-taking was substantially and significantly increased when users were given a specific task prompting them to actively consider a character's perspective. Without such tasks, it seems that participants default to their own viewpoints. These data highlight the need for intentional design in VR experiences to consider content rather than simply viewpoint as key to authentic perspective-taking. To truly harness VR's potential as an "empathy machine," developers must integrate targeted perspective-taking tasks or story prompts, ensuring that cognitive engagement is an active component of the experience.

Index Terms—Virtual Reality, Perspective-Taking, Point of View, Empathy, Sense of Embodiment.

1 Introduction

VR technologies are undeniably impactful, offering unique capabilities to simulate situations and provide vivid, interactive content [1, 22, 34]. Among the celebrated features of VR, developers, researchers, and scholars alike have claimed that VR enhances users' empathy with others and leads to pro-social behaviours (e.g. [5]. Specifically, it is proposed that virtual exposure to situations or stories of other people will allow users to more easily to understand and mentally represent others' thoughts and feelings in unusual situations [19] by using avatars, that represent an ethnic minority [7], an elderly person [38], or a person in a military crossfire [26]. In general, avatars are virtual representations that are controlled by human users [37].

In many cases, the use of a first-person perspective (1PP), where the user *sees* the experience from the *visual* perspective of the avatar, is said to help the user take the social or emotional perspective of others in a similar situation. Yet there is "little empirical evidence of a correlation between VR exposure and an increase in empathy that motivates pro-social behavior" [47]. A key criticism is that (conscious) perspective-taking (PT) requires cognitive effort to make inferences about others' intentions, goals, and motives, and that it cannot be automatically achieved by viewing a scenario from 1PP.

While some researchers argue that VR helps to make perspective-taking tasks less cognitively taxing [20, 29], mental effort is crucial for making meaningful inferences about others [48]. The discourse raises the question of the relative impact of simply viewing from a 1PP in VR versus engaging in effortful mental PT in VR. By understanding this and avoiding assumptions, we can ensure that VR's potential is harnessed effectively.

- *Eugene Kukshinov, Anchit Mishra and Lennart E. Nacke are with University of Waterloo. E-mail: ekukshin@uwaterloo.ca; amishra@uwaterloo.ca; len@uwaterloo.ca*
- *Federica Gini is with University of Trento. E-mail: federica.gini@unitn.it.*
- *Nicholas Bowman is with Syracuse University. E-mail: nbowman@syr.edu.*
- *Brendan Rooney is with University College Dublin. E-mail: brendan.rooney@ucd.ie*

Received 18 September 2024; revised 13 January 2025; accepted 13 January 2025. Date of publication 7 March 2025; date of current version 31 March 2025. Digital Object Identifier no. 10.1109/TVCG.2025.3549137

To understand and test the capabilities of VR technologies, our study employed a two-factorial design to investigate VR perspective-taking capabilities. Our participants assumed the point of view of an avatar of a safety officer in a virtual factory environment, with either a first-person perspective (1PP) or a third-person perspective (3PP), and they were randomly assigned to either a perspective-taking task [10] or a standard think-aloud task [23], while their verbalizations were recorded for analysis.

Our study reveals that the visual point of view in VR alone does not automatically foster PT. Instead, specific PT tasks significantly influence these mental processes, which prompt users to consider a character's thoughts and feelings, regardless of whether they are in a 1PP or 3PP. These findings highlight the need for VR experiences to intentionally incorporate tasks that promote PT and empathy. The implications of these findings suggest that VR developers should design experiences that actively encourage PT and empathy through targeted tasks and exercises, rather than relying on VR's engaging qualities alone.

2 Background

(Conscious) Perspective-taking (PT) is the cognitive process that enables individuals to think about the world from another person's viewpoint. It involves imagining (or mentalizing) what someone else might be thinking, feeling, or experiencing in a given situation. There several ways to distinguish broader types of PT. There is a cognitive and affective distinction [18]. Cognitive PT refers to effortfully understanding others' thoughts and beliefs, while affective PT involves understanding and sharing someone else's emotions. Research indicates that cognitive and affective perspective-taking differ and can even be distinguished on a neural level [18]. PT can be further divided into Imagine-other PT, where one thinks about another person's experience, and Imagine-self PT, where one pictures themselves in another's shoes. These have distinct effects [3]: Imagine-self can evoke empathy but may also lead to self-centered biases, while Imagine-other tends to foster a more accurate understanding and cooperative behavior [2, 15].

Closely related to PT is the concept of empathy, which is the ability to understand and share another person's feelings, thoughts, and experiences [4, 17, 42]. Empathy is a complex process, which components can overlap with the construct of PT [11, 25, 28, 46, 52]. In the end, to avoid conceptual confusions, the research community recently suggested to

refer to various forms of empathy and perceptive-taking as mentalizing about cognitive states or mentalizing about affective states [41]. In this paper, to remain consistent with the HCI and psychological fields, we will use the term perspective-taking (PT) to refer to effortful (conscious) perspective-taking or mentalizing about cognitive or affective mental states, rather than the multi-sensory illusion of being embodied within a virtual body.

3 Perspective-taking in VR

Methodologically, perspective-taking is usually not measured in VR studies [32, 38], instead, researchers use questionnaires [27], such as IRI [10], to measure empathy as a trait. It is usually assumed that VR promotes PT naturally by placing users in avatars. The process is often called VR Perspective-Taking (VRPT) [43] that involves 'first person perspective-taking' presented in head-mounted displays [19].

3.1 Origin of perspective-taking in VR

Researchers have assumed that adopting the visual perspective of an avatar can automatically prompt mentalizing and PT. Yet there are mixed results underpinning this assumption. For instance, some studies claim that presence, as defined in the study [20], the ability to vividly experience a situation, facilitates PT. Herrera et al. [20] suggested that presence enables viewers to feel as though they are in a different location, which supposedly allows them to experience the perspective of a character depicted in a story by being in the same space. Similarly, Shin [45] argues that the sense of being in a different location provides viewers the opportunity to feel the character's perspective. Other research claims that immersion, realism, and embodiment (different from presence) in VR allow users to experience situations from another's viewpoint [19]. Particularly, the use of 1PP is often highlighted as a key factor in enabling this experience [19].

Overall, embodiment is a complex multisensory illusion that one is operating in a different body [21]. It sometimes refers to the concept of placing the user "in another's shoes," which is believed to encourage adopting another person's perspective [49]. Researchers argue such embodiment in VR, usually equated with the mere use of an avatar [29, 50], prompts the perspective taking of the avatar [29]. In some studies, the concept of embodiment is referred to "body transfer," where participants undergo an embodiment phase such as waving their hands in front of a virtual mirror to synchronize the felt and visual movements with the virtual body [20, 50]. However, it remains unclear how this leads to PT as thinking of another person's social or emotional experience.

3.2 Impact of perspective-taking in VR

Previous studies on PT have used these embodiment effects to explore the use of VR to promote prosocial behavior, particularly by 'embodying' avatars from minority groups in virtual environments, in the absence of specific social tasks. Despite the simplicity of these setups, research has shown that such experiences can reduce biases. For example, Peck et al. [40] found that light-skinned participants who embodied a Black avatar in a virtual setting exhibited a reduction in implicit bias, measured by the Implicit Association Test (IAT). Similar results were found by Maister et al. [32], who used the Rubber Hand Illusion paradigm to show that participants embodying a Black hand experienced a positive shift in racial attitudes. Most studies in this field rely on the (IAT) as a primary measure.

Despite the promising findings, the role of VR in promoting PT is far from conclusive. While some studies, like those by Herrera et al. [20], show that VR can foster long-term empathy, others report null or even negative outcomes. Nikolaou et al. [36] found that in some cases, embodying a minority avatar can activate harmful stereotypes, increasing prejudice rather than reducing it [51]. Schulze et al. [44] also found mixed results when using VR to reduce gender bias, increasing it among men.

A few studies employed PT activities within VR environments [20]. PT activities are exercises designed to help individuals understand and appreciate the viewpoints and experiences of others [10]. These activities are often used to build empathy, improve communication, and foster better relationships by encouraging people to see situations from different angles. These tasks, or other techniques, like live-action role-playing [12], can promote PT. There is evidence that VR-based perspective-taking tasks can have more enduring effects compared to traditional PT tasks. Herrera et al. [20] compared traditional PT activities with VR-based tasks and found that while both conditions led to short-term increases in empathy, participants who engaged in VR PT tasks showed more lasting behavioral changes, such as signing petitions in support of homeless people. Crone and Kallen [8] similarly found that participants who embodied an avatar of a different gender in a VR job interview setting exhibited shifts in hiring preferences, which were not observed in an online PT task. These studies, along with meta-analyses by Nikolaou et al. [36], confirm that VR is generally more effective at fostering prosocial behaviors compared to other media.

However, recent meta-analysis of VR perspective-taking studies [33] also found that while VR technologies effectively promote compassionate emotional reactions, or affective empathy, they do not enhance cognitive empathy or PT. The authors concluded that VR limits the need for imagination, reducing opportunities for mental practice in unfamiliar situations.

In summary, while VR has shown promise in reducing biases and promoting empathy, the outcomes are not universally positive. The scholarship on the VR perspective-taking is minimal, but it already shows that it may depend on the specific context and the cognitive and social tasks involved in the experience. Further research is needed to better understand how to optimize VR environments for empathy-building and PT promotion.

4 Hypotheses and Research Questions

While VR shows promise as a tool for promoting pro-social behaviour and reducing bias, there is still much to uncover about the roles of embodiment, POVs, perspective-taking (PT), and social interaction in these processes. Further research is needed to clarify these mechanisms and to determine how VR can be most effectively used to foster cognitive empathy and improve intergroup relations. In this study, we will test the following:

H1: Perspective-taking exercise will increase perspective-taking (PT) expressions.

H2: 1PP will increase perspective-taking (PT) expressions.

We will also test the effect of the POV on the sense of embodiment, as 1PP is supposed to promote it [14].

H3: Sense of embodiment will be rated higher in the 1PP vs 3PP

Finally, we want to learn about different forms of PT, such as cognitive or affective types, or the ones focused on either self- or Other-images.

RQ: How do different forms of PT vary in PT expressions across conditions?

5 Methods

The in-lab experiment with two-factorial study design (Task Type x Point of view) was based on a virtual factory with 1PP and 3PP developed for this study in Unity, a think-aloud protocol [23] during the session to measure PT via audio recordings, and the post-experimental questionnaire to gather demographics and control variables. See the overall design in the Fig. 1. The study received ethics approval from the University of Waterloo Ethics Research Office (Protocol number 45902).

5.1 Sampling

Students from a large public university in North America were recruited through email lists and social media posts within the university community. Out of an initial sample of 97 participants, the final sample for analysis consisted of N=96 participants. The discrepancy resulted from adjustments made to the virtual environment after the first participant. Participants were compensated with a gift card for their involvement. Ethical approval for the study was obtained.

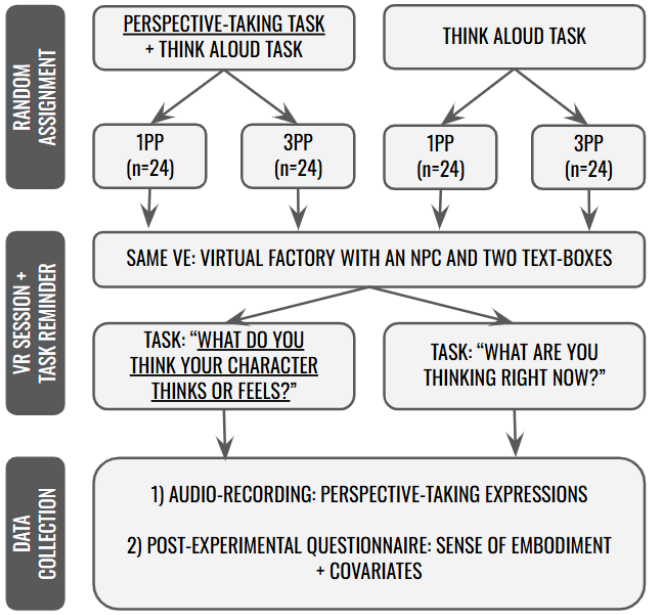

Fig. 1: Study Design Scheme

There were 36 South Asian, 22 East Asian, 12 White, 8 Middle Eastern, 6 Southeast Asian, five Black, three Hispanic, 3 Mixed race/ethnicity participants, and one participant preferred not to disclose this information. Among them, 48 were Female, 47 were Male, and one were Non-binary. All participants received an Amazon gift card for their participation. Informed consent was received from all participants.

Through random assignment, $n = 24$ were assigned to each of the 4 conditions. There were no differences between the groups on the demographic variables such as age, $F(3,92) = 1.527$, $p = .21$, or Gender, $\text{Chi}^2 = 7.42$, $p = .28$.

5.2 Virtual Environment

Our virtual environment (VE) was designed using the Unity game engine. A rigged virtual avatar was mapped to the Meta Quest 3 VR headset and its accompanying Touch controllers to track the head and hands respectively, using Inverse Kinematics (IK). Additionally, a two-bone constraint was introduced for the arms of the virtual avatar, to enable more realistic movement for the elbow joint as the avatar's hands moved. The idea behind IK is to derive the motion of an object, typically one with joints, given changes in the position of its endpoints. The avatar was fully covered in a safety suit so it was not possible to assume their race and/or gender.

5.2.1 The virtual scene and scenario

The virtual environment depicted a factory where participants assumed the role of a safety officer responsible for making decisions. This scenario was intentionally selected as an *unfamiliar* or unusual setting for students to mirror studies that simulate various situations and identities [19, 33]. However, it is also crucial to investigate empathy and PT is important for the work environments [31].

The scene took place in a room where the participant was positioned alongside a non-playable character (NPC) responsible for displaying two key messages. Through a large glass window, the entire factory was visible in front of them. The virtual environment was enhanced with two distinct audio tracks: the sound of factory machinery, which gradually diminished from the start of the first message, and the sound of protesters outside the factory, which began to increase in volume after the first message.

At the beginning of the scene, participants had 30 seconds to observe the virtual environment without any interaction. After this initial observation period, the NPC turned towards the participant and displayed the first message:

> "Good afternoon! As you ordered, the new safety system was purchased and installed over the weekend. Thirty workers, as was planned, were let go to cut costs due to the updates. They received their monthly allowance and bonuses. However, it seems some workers are gathering outside to protest your decision. In any case, you can review the new safety system in 10 minutes."

This message remained visible for one minute. During this time, the sound of the protest outside gradually grew louder. After one minute of silence, the NPC turned back to display the second message:

> "It seems like the protest outside the factory gates is getting stronger. Looks like they're not backing down anytime soon. We need to figure out our next move. Should we try talking to them directly or maybe consider getting the police involved? It's not an easy call, but we need to do something to keep things under control."

The second message also remained visible for one minute. Following another minute of inactivity, the scene concluded. Overall, the virtual scene lasted four and a half minutes.

5.2.2 The points of view

The first-person perspective (1PP) conditions utilized the IK tracking described earlier, where the player's viewpoint originated from the position of the virtual avatar's head, closely mimicking real-life perspective. For the third-person perspective (3PP) conditions, the approach was similar to the allocentric third-person perspective used by Bhandari and O'Neill [6], which has been shown to enhance dynamic task performance and spatial perception compared to egocentric third-person cameras. In this method, the camera was positioned at a fixed distance behind the virtual avatar, but its rotation was not centered around the avatar's position, unlike a conventional egocentric third-person perspective where the camera always rotates around the avatar's position. This setup meant that if the player turned 180 degrees, they would see the part of the virtual environment (VE) that was initially behind them, and while the virtual avatar would also turn similarly due to IK, the player would no longer be able to see their avatar.

The allocentric third-person perspective was implemented by creating an additional camera that followed the rotation of the VR headset but was placed at a fixed offset from the virtual avatar's head. The view of the headset was toggled to use this camera instead of the standard XR origin camera used in Unity's XR interaction toolkit. This format of the third-person perspective (3PP) was chosen due to its association with reduced cybersickness [6] and the need for participants to observe the situation in front of them in a static mode.

5.3 Procedure

Upon arrival, participants were first provided with a consent form. After obtaining informed consent, participants were guided to the center of the lab, where the experimental setup was located. A wireless microphone was carefully set up and tested to ensure clear audio recording of the participant's speech throughout the session.

Participants were given detailed instructions about the upcoming virtual environment. They were informed that the virtual environment would be set in a factory, where they would assume the role of a safety officer responsible for making decisions within this setting. The primary task for participants was to engage in a "think-aloud" protocol, where they were asked to verbalize their thoughts and opinions continuously as they navigated the environment [23]. They were instructed to express any thoughts, regardless of content, out loud. Participants were informed that if they forgot to speak up, the experimenter would remind them to continue verbalizing their thoughts. They were also instructed to read any text boxes that appeared on the screen out loud. For participants in the Perspective-Taking (P.T.) condition, additional (perspective-taking) instruction [10] were provided: they were asked to try to take the perspective of their assigned character and to imagine how this character might feel or think in the situation.

Fig. 2: Screenshot From the Virtual Scene Depicting 3PP Point of View on the NPC with a Text Box

Fig. 3: Screenshot From the Virtual Scene Depicting 1PP Point of View on the Rest of Virtual Room and Factory

Participants were then equipped with a VR headset. They were informed that the buttons on the headset would not be functional and that they could move within the virtual environment by physically walking, although they were advised not to walk excessively. The participants were reminded of their task throughout the process to ensure they remained focused on verbalizing their thoughts, i.e., depending on the condition, it was either "what are thinking right now?" or "what do you think your character feels or thinks about the situation."

Once the participants completed the virtual environment task, they removed the microphone and VR headset. They were then asked to fill out a post-experiment questionnaire. Finally, participants were thoroughly debriefed about the purposes of the study and any specific hypotheses being tested. They were thanked for their participation before being dismissed.

5.4 Measures

5.4.1 Perspective-taking (PT)

To measure perspective-taking, we employed think-aloud protocol and recorded with a wireless microphone what people think. Four different forms of PT were measured by the length of time participants expressed it during the session in seconds (through a 4-minute period) via content analysis of the audio recordings. It was done via (quantitative) content analysis, which is a technique used to convert qualitative data, such as interview transcripts, videos, or images, into a quantitative format [24, 30]. Two independent coders underwent training using a codebook they developed during the analysis process. After achieving reliable coding on selected sub-samples, they proceeded to independently code the remaining data. The complete codebook can be found in Appendix Sec. 8.

Codebook and Coder Training In this study, the coding focused on two primary variables of the perspective-taking expressions: "content" and "identity." "Content" referred to the cognitive or emotional aspects of the participants' expressions, while "identity" referred to the agent in the situation that participants thought about and described. For the "content" variable, coders distinguished between cognitive perspective-taking (CPT) and affective perspective-taking (APT). CPT involved participants inferring thoughts or beliefs *of the character*, while APT involved inferring emotions or feelings *of the character*. For the "identity" variable, coders identified whether the participants' PT was oriented towards imagining themselves in the situation (imagine-self) or imagining the other person (imagine-other). These orientations have been shown to have distinct effects [2, 3, 15]. Each expression of PT was coded based on these two dimensions, resulting in four possible combinations of perspective-taking: CPT x Imagine-other (CO), CPT x Imagine-self (CS), APT x Imagine-other (AO), and APT x Imagine-self (AS). For example, if participants said "I'd feel terrible about this," this was considered as Self-oriented affective PT; if participant said "He is probably thinking about the solution," this was considered other-oriented cognitive PT. Any uncertain expressions were not coded to avoid assumptions.

Coders marked the start and end times of each instance of PT and AE, following specific criteria for distinguishing and identifying these expressions. A new instance was marked when participants shifted from one type of expression to another (e.g., from cognitive to affective perspective-taking), introduced unrelated content (e.g., describing the room), or after a significant pause or task reminder from the experimenter. Coders ensured that each instance was logically complete, meaning that the expressed idea or thought was fully articulated before being marked as the end of an instance (this included reasoning for a feeling or a thought that participants described). The coding process excluded expressions unrelated to thoughts, feelings, emotions, or beliefs, as well as suggestions on how to resolve the virtual situation that lacked a contextualized thinking process, e.g., "we should call the police."

Reliability Calculations and Retraining The initial training lasted one hour, followed by a week for the first round of coding using a random 20% subsample of the data. After this round, two independent coders achieved an acceptable level of inter-coder reliability, with Krippendorff's $\alpha = 0.87$ [24, 30], regarding the count of instances of PT. However, reliability concerning the length of the PT expressions was not sufficient, with Krippendorff's $\alpha = 0.71$.

During the first retraining, coders identified and resolved inconsistencies. The main issue was that coders used different criteria to determine the end of PT instances, sometimes neglecting the reasoning behind the PT or making technical miscalculations. Consequently, after the first retraining, the level of inter-coder reliability for the length of PT expressions remained similar, with Krippendorff's $\alpha = 0.70$.

During the second and final retraining, coders pinpointed the primary reasons for the remaining disagreements and decided to skip uncertain cases, as well as those unclear due to participants being difficult to hear on the recordings. After this final retraining, we achieved an appropriate level of reliability, with Krippendorff's $\alpha = 0.90$. The coded subsample was reconciled, and the remaining data was divided among the coders.

Descriptives of PT measures Overall, on average, each participant expressed PT 2.2 times ($SD = 2.37$), which on average for each participant took 31 seconds ($SD = 36.47$), or, more specifically, CO (M =8, $SD = 17.59$), CS ($M = 8.76$, $SD = 17.59$), AO ($M = 9.89$, $SD =$

Table 1: The Differences Between the Conditions in Terms of the Total PT Expressions, in Seconds

Condition	N	*M*	*SD*
1PP x regular	24	1.71	5.05
3PP x regular	24	4.54	9.93
1PP x PT task	24	65.21	30.91
3PP x PT task	24	52.5	32.95

17.62), and AS (M = 4.34, SD = 8.90). Notably, focal analysis of H1 and H2 considers total time spent engaging in PT, with *post-hoc* analysis of the distinct types of PT; RQ1 focused on discovering differences between different types of PT.

5.4.2 Sense of embodiment

While the current study is focused on the impact of viewpoint on perspective-taking, POV may also affect the sense of embodiment [14]. For this reason we measured the sense of embodiment based on the items suggested by Peck and Gonzalez-Franco [39], Cronbach's α=0.86. The scale included 13 items, including the questions on the tactile feedback, even though there was no interaction with the VE. As authors noted, in situations where there are no active touching situations, the participant will likely still experience some form of passive haptics such as their feet touch the ground [39].

5.4.3 Covariates

In the questionnaire, along with the demographics, we measured the participants' empathy trait via Interpersonal Reactivity Index [9]. Specifically, we used the 7-item subscale of perspective-taking, Cronbach's α = 0.54. Finally, in the content analysis, we also counted how many times participants were reminded about the task to control for the effect of prompting. After the initial training, there was perfect agreement on the count of task reminders. The remaining data was divided among the coders. The data-set was not normally distributed, W = 0.959, p < .001. There was a significant difference between the groups, specifically, those in the PT task groups were reminded to speak up more frequently, $H(3, 92) = 21.54$, p < .001, as per non-parametric ANOVA.

6 Results

6.1 PT as a function of experimental conditions (H1 & H2)

Our core prediction was focused on variance in perspective-taking as a function of experimental condition. Notably, none of our four coded PT variables was normally distributed (W = 0.805, p < .001), including CO (W = 0.534, p < .001), CS (W = 0.530, p < .001), AO (W = 0.635, p < .001), AS (W = 0.568, p < .001), as per Shapiro-Wilk test, requiring non-parametric statistical tests below.

A Scheirer-Ray-Hare test (a non-parametric alternative to ANOVA) was conducted to examine the effects of POV and Task on overall time spend on PT. The results indicated that perspective-taking task had a significant effect on PT, $H(1,92) = 70.79$, $p < .001$, $\varepsilon^2 = 0.75$, supporting H1. However, POV did not significantly influence PT, $H(1,92) = 0.06$, $p = .80$, $\varepsilon^2 < .001$; thus, H2 was not supported. Additionally, there was no significant interaction between the Task and POV, $H(1,92) = 0.72$, $p = .395$, $\varepsilon^2 = .02$. Also see Tab. 1 and Fig. 4.

As a follow-up to the primary tests of H1 and H2, we also used a multiple linear regression to investigate the influence of covariates on our findings reported above, as the non-parametric Scheirer-Ray-Hare test cannot analyze covariates. The overall regression model was statistically significant, $F(6,86) = 24.11$, $p < .001$, adjusted $R2 = .601$. Closer inspection shows that the main effect of PT task remains sizable and statistically significant($\beta = .97$, $p < .001$) on PT, and POV is non-significant ($\beta = 0.04$, $p = .607$). However, task reminders ($\beta = -.17$, $p = .025$) had a slight negative influence, wheres trait perspective taking and sense of embodiment did not significantly predict PT. The interaction between Task and POV was not statistically significant ($\beta = -.22$, $p = .056$).

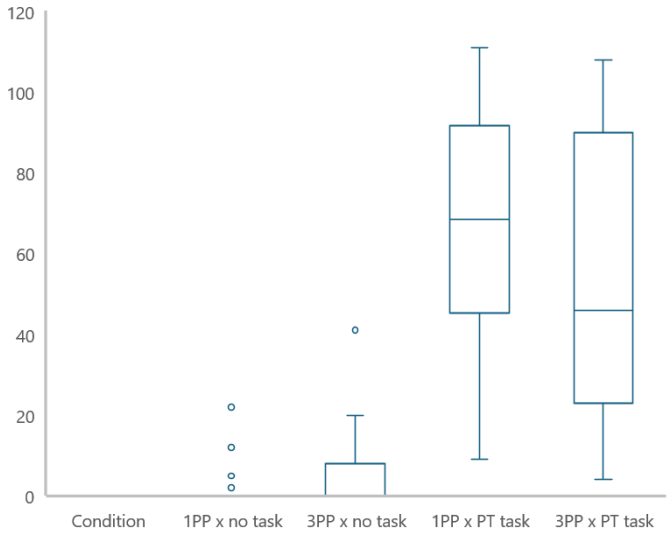

Fig. 4: Distribution of PT Across Conditions (in Seconds)

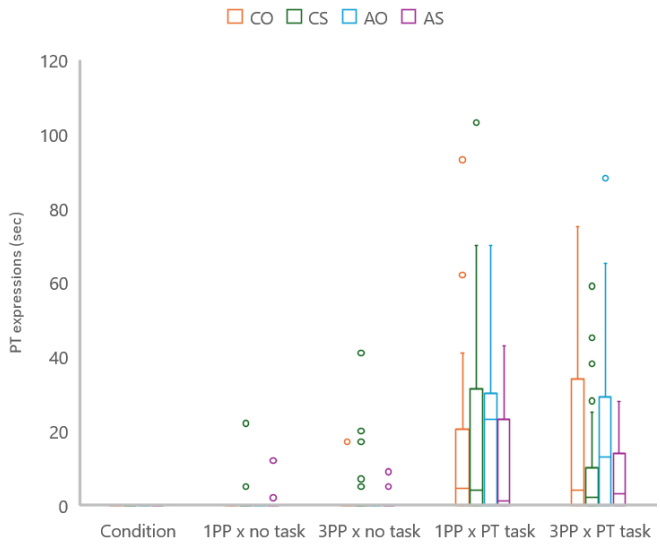

Fig. 5: Distribution of Separate Forms of PT

6.2 Sense of Embodiment (H3)

Sense of embodiment was normally distributed (W = 0.987, $p = .512$) allowing for a factorial ANOVA. Indeed, the factorial ANOVA revealed significant main effects of both POV ($F(1,89) = 7.16$, $p = 0.009$, partial $\eta^2 = 0.07$) and PT Task ($F(1,89) = 7.44$, $p = 0.008$, partial $\eta^2 = 0.08$) on the sense of embodiment. However, the interaction between POV and Task was not significant ($F(1,89) = 0.21$, $p = 0.65$, partial $\eta^2 < .001$). The statistically significant influence of POV supported H3, although the influence of PT task on sense of embodiment was not expected.

6.3 Variance in types of PT by experimental condition (RQ1)

The results of the Scheirer-Ray-Hare test were the same (as for the total PT) for each individual form of PT: Task significantly influenced CO (H(1, 92) = 29.40, $p < .001$, $\eta^2 = .31$), CS (H(1, 92) = 18.28, p < .001, $\eta^2 = .20$), AO (H(1, 92) = 46.95, p < .001, $\eta^2 = .49$), and AS (H(1, 92)=24.74, p < .001, $\eta^2 = .26$). However, POV and the interaction between the POV and Task did not significantly affect any of the dependent variables. See Fig. 5.

7 Discussion

Results of our study show that using an avatar by itself does not promote empathy or perspective-taking. Results indicate that it is the perspective-taking task or exercise that promoted VR users in the unfamiliar virtual situation to think about the character's thoughts or feelings. It did not matter whether it was 1PP or 3PP, or whether it was higher or lower sense of embodiment. PT task was the only significant variable, explaining most of the variance in PT. Without the task, only a few participants inferred from the perspective of the character, but it is likely that it was due to their natural tendency to do so. However, as the PT trait scale was not reliable, so we cannot be sure of the replicability of this finding.

With few exceptions, participants in the conditions with the regular task either described what they saw, what they felt or what they thought about the situation from their own perspective, like what the solution could be to the situation, what they should do next, or how they do not know what to do in this situation. They were not explicitly thinking about the other or 'becoming the character' just because they physically saw the situation from their visual point of view. A visual point of view is not the same as a mental/cognitive point of view or perspective. In the same way, seeing is not the same as thinking.

It is often implied that consciously inferring the mental states of others and adopting the same stance or attitude as another person can be *automatically* prompted by merely seeing a VR scenario from a person's visual perspective. Our results suggest that effortfully and actively engaging with the task of adopting someone's perspective is a functional route to mentalizing. In this context, the term "perspective-taking" is fraught with inter-disciplinary difficulties. At face value, the term "perspective" is not sensorily specific (mental, visual, haptic) and the term "taking" might be read as "taken", depending on the degree of effort imbued in the process.

Our findings do not speak to the way in which prompting a full body-transfer illusion, using an embodiment procedure, may impact more unconscious bodily resonance with the virtual body (unless using an avatar in 1PP is somehow equated to it). However, the self-reported sense of embodiment was lower in the perspective-taking (PT) conditions, possibly because participants were preoccupied with thinking about and adopting the character's perspective, rather than focusing on the sensations from their own virtual body and environment.

We saw, while not statistically significant at the 0.05 level (p = .0561), the interaction between Task and POV in predicting PT was close to significance and likewise with a small sample we could be committing Type II error here, especially as the observed effect was β = .224. There is a possibility that the effect of Task on the PT variable is influenced by the POV. Considering that regressions assume data normality, this requires further investigation as this finding is not robust - our data was both non-normal and based on smaller samples. That said, what we provide evidence for is that effortful perspective taking tasks had a larger effect on mentalizing than just the visual viewpoint.

Here we demonstrate that seeing from visual viewpoint of a virtual body does not equate to thinking about the character and situation they are in; an avatar is not the same as a character with their represented thoughts and feelings [13]. Even if there is a developed character with the story, it does not mean we automatically adopt their thoughts(as in [29]); we may relate to them, which is still a mental process and effort [16]. Although we do not have direct measures of mentalizing and cognitive load, our data (and existing research) do not show that VR contributes to perspective-taking, as it may leave little room for imagination, limiting the mentalizing practice of recreating what it is like to be in an unfamiliar situation; it can stimulate vivid scenes [33]. Nevertheless, this remains an unexplored area that requires more meticulous investigation to better understand the reality of the findings in the current scholarship.

7.1 Implications and future research

VR researchers and designers should recognize that perspective-taking is supported by active engagement from users and thus, creating immersive experiences that intentionally encourage reflection and empathy is powerful. This highlights the importance of designing VR experiences with specific tasks, exercises, or stories [27] that explicitly prompt perspective-taking.

Our study challenges the assumption that first-person visual perspective automatically leads to (conscious) perspective-taking. VR researchers should explore more nuanced definitions and measurement methods for this construct (which involves measuring thinking process during VR experiences). Alternatively, VR research can focus on augmenting the PT based on what VR provides, which is vivid sensorial cues and interactive features. We already know that VR is more effective in promoting prosocial behaviour [20]. VR can also enhance learning about others' thoughts and feelings by providing vivid sensory cues by engaging multiple senses, such as visual, auditory, and even tactile feedback. The high fidelity of VR environments may help users grasp the situational factors that influence how others feel. Interactivity in VR may potentially enhance perspective-taking by actively engaging users in decision-making and reflective processes that require them to consider others' thoughts and feelings. Through role-playing, scenario-based learning, or immediate feedback, users may be more inclined to empathize and understand different perspectives. Interactive experiences can create immersive environments where users must respond to the emotions and behaviours of other characters, making the cognitive process of perspective-taking more natural.

Finally, we need to compare the difference between VR and other technologies based on the exact measures of perspective-taking, not just its potential outcomes, such as bias reduction. As well as, we might explore how individual differences such as gender, culture or socio-economic status might impact the results. In the end, future VR research should be developed with a deeper understanding of the cognitive processes involved in perspective-taking and should include mechanisms that explicitly foster these processes. This could lead to more effective use of VR in therapeutic settings, social skills training, and educational programs designed to build empathy and social understanding. At the same time, this understanding will promote more critical approach to the research of empathy and perspective-taking in VR, as it would be possible to avoid issues like identity tourism [35], which happen when researchers try condense a complex life and experiences of a (traditionally marginalized) person into a minuscule virtual scene that has no real-life consequences to users.

7.2 Limitations

A major limitation of this study is the task reminders used. Specifically, the length of the reminders differed between conditions. The perspective-taking task reminder ("What do you think your character thinks or feels about the situation?") was longer than the regular think-aloud task reminder ("What are you thinking right now?"), which may have taken more time from the experimental group for the think-aloud task. Additionally, there were a few instances where the perspective-taking task reminder varied from the standard one. For example, it may have been stated without mentioning feelings or thoughts, or it might have included the phrase "What else do you think?" However, while there was significantly more task reminders in the PT task conditions, it was a negative predictor of the PT: in other words, those participants in the PT task conditions who thought aloud less were more often reminded to speak up. However, in the regular task conditions, participants often did not want to say much and when they did, they would simply say that they do not have much to say, so reminding them to speak up was, at times, fruitless.

Another aspect of the task that we need to consider is the direct nature of the instructions. Participants were explicitly told to adopt another's viewpoint, which may have influenced their behavior by making perspective-taking feel more like a task or mandate rather than a natural, spontaneous process. While this approach ensures compliance and clarity, it raises questions about the dosage and salience of the intervention. In real-world scenarios, perspective-taking may emerge more organically, prompted by narrative cues or character dialogue rather than strict instructions. A potential future direction could involve replicating the study with softer, more seamless prompts embedded within the narrative or delivered exclusively through interactions with non-playable characters (NPCs).

In regard to the measurements, the speech rate may have affected the length of the PT expressions. However, in our study, participants in the regular task conditions almost never spoke up regarding character's feelings or thoughts. Also, we did not calculate inter-coder reliability statistics for each coded type or form of the PT, as it was not our focus. While the difference between the conditions was strong, it still could have affected the accuracy of the presented differences between the forms of PT, as it is possible that one specific form of PT was not coded reliably. However, we did not notice any significant deviations during the coding of the data.

In regard to the (sense of) embodiment, participants in the first-person perspective (1PP) did not have access to a mirror to see their virtual body. In contrast, participants in the third-person perspective (3PP) could always view their body from the back. Despite this, the embodiment scale indicated that participants in the 1PP condition reported a higher sense of embodiment. It is also important to note that the "out of body experience" question in the embodiment questionnaire was not designed for the 3PP condition and thus may be biased against it. It is one of the questions that likely excluded 3PP participants from fully reporting a sense of embodiment.

Furthermore, the perspective-taking (PT) trait scale demonstrated low reliability, which may explain why it did not function effectively in the study. In the PT task conditions, the scale may not have been sensitive enough to accurately measure the intended effects.

8 Conclusion

This study demonstrates that VR technology alone does not inherently promote empathy or perspective-taking; instead, these processes are significantly influenced by perspective-taking tasks. Our findings reveal that it is the explicit task prompting users to consider a character's thoughts or feelings that drives perspective-taking, regardless of whether users are in a first-person or third-person (visual) perspective. Without such tasks, few participants naturally mentalised with the character's perspective, i.e., they mainly described their own perceptions and solutions rather than describing the character's viewpoint, suggesting that other factors, rather than VR's inherent features, drives PT. Additionally, the physical viewpoint in VR and the sense of embodiment did not translate into a process of thinking about others' thoughts or feelings, highlighting that the process of seeing is not the same as thinking. These results suggest that VR experiences aimed at promoting empathy should be designed with specific tasks that actively encourage reflection and mentalizing. Future research should focus on refining perspective-taking measures and exploring how VR's sensory and interactive features can enhance cognitive engagement with others' perspectives and prompt users to consider others' thoughts and feelings deeply.

Supplemental Materials

The virtual scene is available via GitHub: https://github.com/EugeneK-PhD/VR-studies/tree/13bdbc4386b17099b05cad16db6d0d565d646daa/Perspective%20Project

Acknowledgment

This study was supported by the SSHRC INSIGHT Grant (grant number:435-2022-0476), NSERC Discovery Grant (grant number: RGPIN-2023-03705), CFI John R. Evans Leaders Fund (grant number: 41844), Meta Research Award, and Lupina Foundation. Special thanks to the members of the Games Institute at the University of Waterloo for all their support.

References

[1] J. Bailenson. *Experience on demand: what virtual reality is, how it works, and what it can do*. W.W. Norton & Company, New York, NY, 2018. OCLC: 1017735353. 1

[2] C. D. Batson. *The altruism question: Toward a social-psychological answer*. Lawrence Erlbaum Associates, Inc, 1991. 1, 4

[3] C. D. Batson, S. Early, and G. Salvarani. Perspective taking: Imagining how another feels versus imagining how you would feel. *Personality and Social Psychology Bulletin*, 23(7):751–758, 1997. doi: 10.1177/0146167297237008 1, 4

[4] P. S. Bellet. The Importance of Empathy as an Interviewing Skill in Medicine. *JAMA: The Journal of the American Medical Association*, 266(13):1831, Oct. 1991. doi: 10.1001/jama.1991.03470130111039 1

[5] P. Bertrand, J. Guegan, L. Robieux, C. A. McCall, and F. Zenasni. Learning Empathy Through Virtual Reality: Multiple Strategies for Training Empathy-Related Abilities Using Body Ownership Illusions in Embodied Virtual Reality. *Frontiers in Robotics and AI*, 5:26, Mar. 2018. doi: 10.3389/frobt.2018.00026 1

[6] N. Bhandari and E. O'Neill. Influence of Perspective on Dynamic Tasks in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 939–948. IEEE, Atlanta, GA, USA, Mar. 2020. doi: 10.1109/VR46266.2020.00114 3

[7] V. H. H. Chen, S. H. M. Chan, and Y. C. Tan. Perspective-Taking in Virtual Reality and Reduction of Biases against Minorities. *Multimodal Technologies and Interaction*, 5(8):42, Aug. 2021. doi: 10.3390/mti5080042 1

[8] C. L. Crone and R. W. Kallen. Interview with an avatar: Comparing online and virtual reality perspective taking for gender bias in STEM hiring decisions. *PLOS ONE*, 17(6):e0269430, June 2022. doi: 10.1371/journal.pone.0269430 2

[9] M. H. Davis. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1):113–126, Jan. 1983. doi: 10.1037/0022-3514.44.1.113 5

[10] M. H. Davis, L. Conklin, A. Smith, and C. Luce. Effect of perspective taking on the cognitive representation of persons: A merging of self and other. *Journal of Personality and Social Psychology*, 70(4):713–726, 1996. doi: 10.1037/0022-3514.70.4.713 1, 2, 3

[11] J. Decety and P. L. Jackson. The Functional Architecture of Human Empathy. *Behavioral and Cognitive Neuroscience Reviews*, 3(2):71–100, June 2004. doi: 10.1177/1534582304267187 1

[12] T. Dyrks, L. Ramirez, S. Denef, B. Penkert, and D. Meyer. Designing for firefighters - building empathy through live action role-playing. In *Proceedings*, 2009. 2

[13] J. Fox and S. J. Ahn. Avatars: Portraying, Exploring, and Changing Online and Offline Identities. In R. Luppicini, ed., *Handbook of Research on Technoself: Identity in a Technological Society*, pp. 255–271. IGI Global, Hershey, PA, USA, 2013. doi: 10.4018/978-1-4666-2211-1.ch014 6

[14] R. Fribourg, F. Argelaguet, A. Lecuyer, and L. Hoyet. Avatar and Sense of Embodiment: Studying the Relative Preference Between Appearance, Control and Point of View. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):2062–2072, May 2020. doi: 10.1109/TVCG.2020.2973077 2, 5

[15] A. D. Galinsky, G. Ku, and C. S. Wang. Perspective-Taking and Self-Other Overlap: Fostering Social Bonds and Facilitating Social Coordination. *Group Processes & Intergroup Relations*, 8(2):109–124, Apr. 2005. doi: 10.1177/1368430205051060 1, 4

[16] M. Green, H. Bilandzic, K. Fitzgerald, and E. Paravati. Narrative Effects. In *Media Effects*, pp. 130–145. Routledge, Oxfordshire, England, UK, 4 ed., 2019. 6

[17] J. A. Hall and R. Schwartz. Empathy present and future. *The Journal of Social Psychology*, 159(3):225–243, May 2019. doi: 10.1080/00224545.2018.1477442 1

[18] M. L. Healey and M. Grossman. Cognitive and Affective Perspective-Taking: Evidence for Shared and Dissociable Anatomical Substrates. *Frontiers in Neurology*, 9, June 2018. doi: 10.3389/fneur.2018.00491 1

[19] D. Hekiert, M. Igras-Cybulska, and A. Cybulski. Designing VRPT experience for empathy toward out-groups using critical incidents and cultural explanations. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 506–507, Oct. 2021. doi: 10.1109/ISMAR-Adjunct54149.2021.00122 1, 2, 3

[20] F. Herrera, J. Bailenson, E. Weisz, E. Ogle, and J. Zaki. Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. *PLOS ONE*, 13(10):e0204494, Oct. 2018. doi: 10.1371/journal.pone.0204494 1, 2, 6

[21] K. Kilteni, R. Groten, and M. Slater. The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, Nov. 2012. doi: 10.1162/PRES_a_00124 2

[22] A. Kitson, M. Prpa, and B. E. Riecke. Immersive Interactive Technologies for Positive Change: A Scoping Review and Design Considerations.

Frontiers in Psychology, 9:1354, Aug. 2018. doi: 10.3389/fpsyg.2018.01354 1

[23] Krahmer. Thinking About Thinking Aloud: A Comparison of TwoVerbal Protocols for Usability Testing. *IEEE Transactions on Professional Communication*, 47(2):105–117, June 2004. doi: 10.1109/TPC.2004.828205 1, 2, 3

[24] K. Krippendorff. *Content Analysis: An Introduction to Its Methodology*. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320, 2019. doi: 10.4135/9781071878781 4

[25] C. Lamm, C. D. Batson, and J. Decety. The Neural Substrate of Human Empathy: Effects of Perspective-taking and Cognitive Appraisal. *Journal of Cognitive Neuroscience*, 19(1):42–58, Jan. 2007. doi: 10.1162/jocn.2007.19.1.42 1

[26] H. M. Lee and B. J. Li. So far yet so near: Exploring the effects of immersion, presence, and psychological distance on empathy and prosocial behavior. *International Journal of Human-Computer Studies*, 176:103042, Aug. 2023. doi: 10.1016/j.ijhcs.2023.103042 1

[27] Y. Lee, H. Shin, and Y.-H. Gil. Measurement of Empathy in Virtual Reality with Head-Mounted Displays: A Systematic Review. *IEEE Transactions on Visualization and Computer Graphics*, 30(5):2485–2495, May 2024. doi: 10.1109/TVCG.2024.3372076 2, 6

[28] S. Leiberg and S. Anders. The multiple facets of empathy: a survey of theory and evidence. In *Progress in Brain Research*, vol. 156, pp. 419–440. Elsevier, 2006. doi: 10.1016/S0079-6123(06)56023-6 1

[29] B. J. Li and H. K. Kim. Experiencing organ failure in virtual reality: effects of self- versus other-embodied perspective taking on empathy and prosocial outcomes. *New Media and Society*, 3(8):2144–2166, 2021. doi: 10.1177/1461444821993122 1, 2, 6

[30] M. Lombard, J. Snyder-Duch, and C. C. Bracken. Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. *Human Communication Research*, 28(4):587–604, Oct. 2002. doi: 10.1111/j.1468-2958.2002.tb00826.x 4

[31] N. H. Longmire and D. A. Harrison. Seeing their side versus feeling their pain: Differential consequences of perspective-taking and empathy at work. *Journal of Applied Psychology*, 103(8):894–915, 2018. doi: 10.1037/apl0000307 3

[32] L. Maister, N. Sebanz, G. Knoblich, and M. Tsakiris. Experiencing ownership over a dark-skinned body reduces implicit racial bias. *Cognition*, 128(2):170–178, Aug. 2013. doi: 10.1016/j.cognition.2013.04.002 2

[33] A. J. Martingano, F. Hererra, and S. Konrath. Virtual reality improves emotional but not cognitive empathy: A meta-analysis. *Technology, Mind, and Behavior*, 2(1), June 2021. doi: 10.1037/tmb0000034 2, 3, 6

[34] F. Moustafa and A. Steed. A longitudinal study of small group interaction in social virtual reality. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–10. ACM, Tokyo Japan, Nov. 2018. doi: 10.1145/3281505.3281527 1

[35] L. Nakamura. Feeling good about feeling bad: virtuous virtual reality and the automation of racial empathy. *Journal of Visual Culture*, 19(1):47–64, Apr. 2020. doi: 10.1177/1470412920906259 6

[36] A. Nikolaou, A. Schwabe, and H. Boomgaarden. Changing social attitudes with virtual reality: a systematic review and meta-analysis. *Annals of the International Communication Association*, 46(1):30–61, Jan. 2022. doi: 10.1080/23808985.2022.2064324 2

[37] K. Nowak and J. Fox. Avatars and computer-mediated communication: A review of the definitions, uses, and effects of digital representations. *Review of Communication Research*, 6:30–53, Jan. 2018. doi: 10.12840/issn.2255-4165.2018.06.01.015 1

[38] S. Y. Oh, J. Bailenson, E. Weisz, and J. Zaki. Virtually old: Embodied perspective taking and the reduction of ageism under threat. *Computers in Human Behavior*, 60:398–410, July 2016. doi: 10.1016/j.chb.2016.02.007 1, 2

[39] T. C. Peck and M. Gonzalez-Franco. Avatar Embodiment. A Standardized Questionnaire. *Frontiers in Virtual Reality*, 1:575943, Feb. 2021. doi: 10.3389/frvir.2020.575943 5

[40] T. C. Peck, S. Seinfeld, S. M. Aglioti, and M. Slater. Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and Cognition*, 22(3):779–787, Sept. 2013. doi: 10.1016/j.concog.2013.04.016 2

[41] F. Quesque, I. Apperly, R. Baillargeon, S. Baron-Cohen, C. Becchio, H. Bekkering, D. Bernstein, M. Bertoux, G. Bird, H. Bukowski, P. Burgmer, P. Carruthers, C. Catmur, I. Dziobek, N. Epley, T. M. Erle, C. Frith, U. Frith, C. M. Galang, V. Gallese, D. Grynberg, F. Happé, M. Hirai, S. D. Hodges, P. Kanske, M. Kret, C. Lamm, J. L. Nandrino, S. Obhi, S. Olderbak, J. Perner, Y. Rossetti, D. Schneider, M. Schurz, T. Schuwerk, N. Sebanz, S. Shamay-Tsoory, G. Silani, S. Spaulding, A. R. Todd, E. Westra, D. Zahavi, and M. Brass. Defining key concepts for mental state attribution. *Communications Psychology*, 2(1), Apr. 2024. doi: 10.1038/s44271-024-00077-6 2

[42] H. Read. A typology of empathy and its many moral forms. *Philosophy Compass*, 14(10):e12623, Oct. 2019. doi: 10.1111/phc3.12623 1

[43] M. Romak, M. Igras-Cybulska, D. Hemmerling, and D. Hekiert. EthnoVR 3.0: Fostering Ethnocultural Empathy and Mitigating Ethnocentrism through Immersive Virtual Environments and Virtual Reality Perspective Taking. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 297–299. IEEE, Orlando, FL, USA, Mar. 2024. doi: 10.1109/vrw62533.2024.00059 2

[44] S. Schulze, T. Pence, N. Irvine, and C. Guinn. The Effects of Embodiment in Virtual Reality on Implicit Gender Bias. In J. Y. Chen and G. Fragomeni, eds., *Virtual, Augmented and Mixed Reality. Multimodal Interaction*, pp. 361–374. Springer International Publishing, Cham, 2019. doi: 10.1007/978-3-030-21607-8_28 2

[45] D. Shin. Empathy and embodied experience in virtual environment: To what extent can virtual reality stimulate empathy and embodied experience? *Computers in Human Behavior*, 78:64–73, Jan. 2018. doi: 10.1016/j.chb.2017.09.012 2

[46] T. Singer. The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience & Biobehavioral Reviews*, 30(6):855–863, Jan. 2006. doi: 10.1016/j.neubiorev.2006.06.011 1

[47] C. Sora-Domenjó. Disrupting the "empathy machine": The power and perils of virtual reality in addressing social issues. *Frontiers in Psychology*, 13, Sept. 2022. doi: 10.3389/fpsyg.2022.814565 1

[48] J. Stietz, E. Jauk, S. Krach, and P. Kanske. Dissociating Empathy From Perspective-Taking: Evidence From Intra- and Inter-Individual Differences Research. *Frontiers in Psychiatry*, 10, Mar. 2019. doi: 10.3389/fpsyt.2019.00126 1

[49] M. C. C. Tan, S. Y. L. Chye, and K. S. M. Teng. "In the shoes of another": immersive technology for social and emotional learning. *Education and Information Technologies*, 27(6):8165–8188, July 2022. doi: 10.1007/s10639-022-10938-4 2

[50] A. Van Loon, J. Bailenson, J. Zaki, J. Bostick, and R. Willer. Virtual reality perspective-taking increases cognitive empathy for specific others. *PLOS ONE*, 13(8):e0202442, Aug. 2018. doi: 10.1371/journal.pone.0202442 2

[51] N. Yee and J. Bailenson. The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Human Communication Research*, 33(3):271–290, July 2007. ERIC Number: EJ799740. doi: 10.1111/j.1468-2958.2007.00299.x 2

[52] J. Zaki and K. N. Ochsner. The neuroscience of empathy: progress, pitfalls and promise. *Nature Neuroscience*, 15(5):675–680, May 2012. doi: 10.1038/nn.3085 1

A Questionnaire Items

1. Age
2. Gender
   - Man
   - Woman
   - Non-binary
   - Prefer not to say
   - Prefer to self-describe:
3. Race/Ethnicity
   - Indigenous (e.g., First Nations, Metis, or Inuit)
   - Black / African
   - East Asian (e.g., Chinese, Japanese, Korean)
   - Southeast Asian (e.g, Filipino, Vietnamese, Korean)
   - South Asian (e.g., Pakistani, Indian)
   - Hispanic / Latine
   - Middle Eastern
   - West Indian / Caribbean

- White / European
- Mixed
- Another Group, not listed above
- Prefer not to say

4. Rate this from 1 – it does not describe me well to 5 – It describes me well
   - I daydream and fantasize, with some regularity, about things that might happen to me.
   - I sometimes find it difficult to see things from the "other guy's" point of view.
   - I really get involved with the feelings of the characters in a novel.
   - I am usually objective when I watch a movie or play, and I don't often get completely caught up in it.
   - I try to look at everybody's side of a disagreement before I make a decision.
   - I sometimes try to understand my friends better by imagining how things look from their perspective.
   - Becoming extremely involved in a good book or movie is somewhat rare for me.
   - If I'm sure I'm right about something, I don't waste much time listening to other people's arguments.
   - After seeing a play or movie, I have felt as though I were one of the characters.
   - I believe that there are two sides to every question and try to look at them both.
   - When I watch a good movie, I can very easily put myself in the place of a leading character.
   - When I'm upset at someone, I usually try to "put myself in his shoes" for a while.
   - Before criticizing somebody, I try to imagine how I would feel if I were in their place.

5. Please rate these from never/strongly disagree to always/strongly agree regarding your today's VR session:
   - "I felt out of my body"
   - "I felt as if my (real) body were drifting toward the virtual body or as if the virtual body were drifting toward my (real) body"
   - "I felt as if the movements of the virtual body were influencing my own movements"
   - "It felt as if my (real) body were turning into an "avatar" body"
   - "At some point it felt as if my real body was starting to take on the posture or shape of the virtual body that I saw"
   - "I felt like I was wearing different clothes from when I came to the laboratory"
   - "I felt as if my body had changed"
   - "I felt a sensation in my body when I saw virtual environment"
   - "I felt that my own body could be affected by the virtual environment"
   - "I felt as if the virtual body was my body"
   - "At some point it felt that the virtual body resembled my own (real) body, in terms of shape, or other visual features."
   - "I felt as if my body was located where I saw the virtual body"
   - "I felt like I could control the virtual body as if it was my own body"

B Content Analysis Codebook

B.1 Introduction

This content analysis aims to identify instances of perspective-taking expressions by the participants who experienced a virtual environment via a VR head-mounted display (HMD).

Materials: The VE is a virtual factory, which includes a room where participants find themselves. In the room, there is also a character, or NPC, who delivers news to the participants via a text box.

VR sessions description:

1. At the start, nothing happens for 30 seconds, participants observe the VE. There is nothing to code.
2. The NPC turns and displays the first message (the NPC turns away after messages goes away) "Good afternoon! As ordered, the new safety system was purchased and installed over the weekend. 30 workers, as planned, were let go to cut costs due to the updates. They received their monthly allowance and bonuses. However, it seems some workers are gathering outside in the form of a protest. In any case, you can review the new safety system in 10 minutes." [the text stays for a minute]
3. Nothing happens for a minute [the sound of a protest outside is getting louder]
4. The NPC turns back and displays the second message: "It seems like the protest outside the factory gates is getting stronger. Looks like they're not backing down anytime soon. We need to figure out our next move. Should we try talking to them directly or maybe consider getting the police involved? It's not an easy call, but we need to do something to keep things under control." [the text stays there for a minute]
5. Nothing happens for a minute

Overall: 4 minutes is the max time of the think-aloud protocol data sample. This time range will be used to standardize the data.

General procedure: For each participant's audio recordings, coders should carefully listen to them while marking the variables in the code sheet. To start coding, coders should find a moment when participants start reading the first message to start coding after participants are done reading it. In the end, coders need to put their data into the provided spreadsheet with IDs.

B.2 Variables and definitions

There are two main variables that we can call "content" and "identity". Content refers to the cognitive or emotional aspects of the expression, while identity refers to the agent in the situation they think about and describe.

Content: cognitive perspective-taking (CPT) or affective perspective-taking (APT)

- Cognitive PT - the ability to infer thoughts or beliefs.
- Affective PT - the ability to infer emotions or feelings.

Identity: imagine-self or imagine-other PT

- Imagine-self - happens when the person doing PT thinks about the person experiencing the situation
- Imagine-other - happens when the person doing PT thinks about themselves in the other person's shoes

Each expression of PT will be a combination of the two variables.

- **Cognitive PT x Imagine-other** - the ability to infer the thoughts or beliefs of another agent. In the experimental setting, this happens when the participant thinks about what the main character might think or believe in the experienced situation. The thinking process may include the decision-making process.

  Examples: HE/SHE/THEY are trying to make a decision and choose between the new safety system and his employees; HE/SHE/THEY thinks that having the safety system updated is beneficial for all his remaining employees, while not firing those 30 employees would only benefit them

- **Cognitive PT x Imagine-self** - the ability to infer the thoughts or beliefs of another agent as if they are your own. The participant thinks about what they would think or believe in that situation.

  Examples: I would prefer not to fire my employees and I would rather cut off part of my salary for some months in order to update the safety system later, unless the system is really really old

- **Affective PT x Imagine-other** - the ability to infer the emotions or feelings of another agent. The participant thinks about what the main character is feeling in the experienced situation.

  Example: I think he might be stressed because it's a tough choice

- **Affective PT x Imagine-self** - The ability to infer your emotions or feelings as if you were in the other person's situation. The participant thinks about what they would feel in that situation.

  Example: I would be scared of having so many people protesting just because I am trying to keep my employees safe

In addition, we are coding number of the task reminders (as a control), or the number of times the task was reminded to the participants.

B.3 Coding steps and criteria

In general, coders need to identify instances of perspective-taking expressions, and the number of times the tasks were reminded to the participants. Coders need to mark when the instance started (the time of the recording) and when it stopped.

How to distinguish instances:

- When some unrelated to the variables expression starts. E.g., when participants start to describe the room.
- Different variables. For example, participants start with a character's feelings, but then they speak about the character's thoughts.
- Substantial pause and/or experimenter's call for the task. "What are you thinking" is a regular reminder. The perspective-taking task was reminded like this: "What do you think your character feels or thinks about the situation".
- **NOTE:** the idea should be finished, i.e. the sentence or group of them that represents an expression should be logically/semantically complete. Then it counts as the end of the expression.

How to identify instances:

- They should indicate or imply a specific pronoun and then the feeling or a thought. E.g. He thinks, I'd feel, etc. If participants respond directly to PT prompting, then
- Context of expressions. For example, several types of P.T. can be expressed under the same indication of the thinking process: "I think…" Context also helps identifying PT when parts of the sentence are left implicit/implied (e.g., [if i were in that situation] I would choose to) especially when replying to the second text box, in which the text says that the character should take a decision (and therefore also "choose to" can be implicit/implied). NOTE this is different from when participants say that "I think it's better/It makes more sense" as it is not enough to infer that they're doing PT.
- Absence of, or an uncertain pronoun would be related to the character if answered after the P.T. task reminder. Otherwise, it should not be coded.

Coding:

- CPT+OTHER: The sentence refers to another character's thoughts or beliefs, which means that the verb will likely be related to thinking, believing, realizing, considering, etc. and the verb will be used in 3rd person.

  Examples: It doesn't seem like they would be very concerned with their own safety while sitting here making decisions.

- CPT+SELF: The sentence refers to another character's thoughts or beliefs, which means that the verb will likely be related to thinking, believing, realizing, considering, etc. and the verb will be declined in 1st person.

  Note: in some cases also, the pronoun "we" might indicate thinking with the NPC. Example: I think that's maybe a good decision (i.e., an evaluation or a cognitive process) probably, It would be nice if we could just talk to the protesters. I don't like the idea of getting the police involved.

  Note: As a reply to the prompt: "I agree with them" - thinking from the point of view of the character

  Example: "I am not sure how I would want to resolve this, as a safety officer"

- APT+OTHER: The sentence refers to another character's emotions and feelings, which means that the verb will likely be related to being + emotion, feeling… and the verb will be declined in 3rd person. Also, 3rd person pronouns will be present.

  Example: I think they feel quite safe given the fact that they're in the some kind of metal structure container

- APT+SELF: The sentence refers to another character's emotions and feelings, which means that the verb will likely be related to being + emotion, feeling… and the verb will be declined in 1st person. Also, 1st person pronouns will be present.

  Example: I would assume I am feeling quite pressured to, make a decision about the protests, and I am appreciative because he's referring to this as not an easy call.

Major exclusions:

- Participants are suggesting solutions to a problem without contextualizing the thinking process. Example: "It would be logical for them to do this" - the participant is not considering what the character might feel or think, but what would be logical from an external point of view.
- Thinking about anything else but feelings, emotions, thoughts, or beliefs.