# Out of Control: Effects of Multimodal Self-similarity on Embodiment During Autonomous Avatar Demonstrations in Virtual Reality

Siqi Guo
Purdue University
West Lafayette, Indiana, USA
guo477@purdue.edu

Fengze Zhang
Purdue University
West Lafayette, Indiana, USA
zhan5455@purdue.edu

Claudia Krogmeier
University of Wisconsin-Madison
Madison, Wisconsin, USA
ckrogmeier@wisc.edu

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

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

**Figure 1: A user controlling her self-similar virtual avatar in our virtual reality application.**

## Abstract

Virtual reality (VR) training often requires autonomous avatar demonstrations, yet embodiment is strongest under direct control. We examine whether multimodal self-similarity (i.e., in appearance and voice) can preserve embodiment when control is constrained. In a 2 (self-similarity: self-similar vs. non-self-similar) × 2 (autonomy: autonomous vs. non-autonomous) within-group study, 24 participants performed a block-assembling task with self-avatars. Autonomous self-avatars increased emotional reactivity and frustration; non-autonomous self-avatars improved presence, agency, and self-attribution. Self-similarity was maintained, and self-attribution persisted during autonomous demonstrations. Tracking of head-direction (as a proxy for gaze) showed autonomy, and self-similarity increased head-based dwell on the mirror, whereas non-autonomous avatars redirected head orientation toward the body, environment, and task; an interaction effect revealed greater task-focused head-direction for non-self-similar autonomous avatars. These results indicate that autonomy and self-similarity appear to have potential additive influences on user perception in this study. We conclude that multimodal self-similarity can buffer embodiment loss during non-controllable phases and offer evidence-based guidance for designing mixed-control VR experiences.

---

*CHI '26, Barcelona, Spain*

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3790629

## CCS Concepts

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

## Keywords

Virtual Reality, Self-avatar, Self-similarity, Motor Learning, Rehabilitation, Voice Cloning, Movement Autonomy, Embodiment

**ACM Reference Format:**
Siqi Guo, Fengze Zhang, Claudia Krogmeier, Dominic Kao, and Christos Mousas. 2026. Out of Control: Effects of Multimodal Self-similarity on Embodiment During Autonomous Avatar Demonstrations in Virtual Reality. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3772318.3790629

## 1 Introduction

Motor learning and rehabilitation applications increasingly rely on virtual reality to provide patients with repeated demonstrations of correct movements [46, 63, 106]. However, these applications face a design challenge: patients benefit from embodying self-similar avatars that enhance motor learning and engagement [35, 58], yet they also need to observe autonomous avatar demonstrations without direct control. This challenge stems from virtual embodiment principles, where users perceive a virtual body as their own through three foundational components: body ownership, agency, and sense of self-location [70]. When users feel that the virtual body both belongs to them and responds to their commands, the system supports motor learning, coordination, and engagement [12, 86]. However, autonomous self-similar avatars may paradoxically increase user frustration by allowing them to observe themselves acting without control.

While embodiment enhances learning when users experience agency and ownership simultaneously [34], emerging rehabilitation and training protocols explore autonomous avatar demonstrations to show correct techniques [15, 31]. Since agency—the sense of control over avatar actions—represents a fundamental component of embodiment [44], autonomous demonstrations that remove direct user control raise questions about whether embodiment benefits can be preserved. Prior work has not systematically examined how self-similarity might preserve embodiment during these necessary autonomous phases, particularly when combined with realistic voice synthesis. Jung et al. [49] found that personalized avatars created via 3D face scans could maintain illusory agency even in non-controllable conditions. Inspired by this, our study investigates whether more accessible personalization methods can similarly support embodiment when paired with different levels of avatar autonomy.

To address the aforementioned gap, we investigated how multimodal self-similarity (i.e., visual appearance and voice synthesis) interacts with avatar autonomy to influence embodiment, user experience, and behavioral responses during task performance. We employed a 2×2 within-group design manipulating both self-similarity (i.e., self-similar vs. non-self-similar appearance and voice) and autonomy (i.e., autonomous vs. non-autonomous behavior). We included both subjective measures and objective head-direction tracking to understand how these factors influence not only self-reported embodiment but also behavioral patterns. Participants ($N = 24$) engaged in a block-assembling task while embodying avatars that varied in self-similarity and autonomy (see Figure 1). To capture attentional dynamics, we incorporated a virtual mirror and measured head-direction distributions.

Our findings revealed that autonomous self-similar avatars create a unique profile of responses, characterized by increased emotional reactivity and frustration; yet, our participants maintained their presence and enhanced mirror-focused attention, which could support observational learning. Importantly, we discovered distinct head-direction patterns where autonomous conditions directed head orientation toward mirrors, potentially supporting motor learning through observation. In contrast, non-autonomous conditions focused attention on task elements and body monitoring. Additionally, we observed an interaction effect, showing that participants allocated significantly more dwell time to task-related objects when a non-self-similar avatar was paired with autonomous behavior. These findings provide systematic evidence for designing avatar systems that balance demonstration needs with embodiment preservation, offering concrete guidance for motor learning applications where both autonomous demonstration and user agency are essential.

The remainder of this paper is structured as follows. In Section 2, we review related work on self-similarity, agency, and embodiment in VR. In Section 3, we detail the experimental methodology, including the study design and data collection process. In Section 4, we present our results, followed by an in-depth discussion in Section 5. Finally, we conclude in Section 6 and discuss directions for future research.

## 2 Literature Review

### 2.1 Embodiment

The sense of embodiment, which is the feeling that a virtual body belongs to oneself, emerges from the integration of body ownership, self-location, and agency [53]. This foundational concept shapes how users experience virtual environments, particularly when avatar control is disrupted. The first-person perspective (1PP) is crucial for enhancing embodiment, aligning with the user's natural viewpoint, and increasing presence. Studies show that 1PP enhances body ownership more than the third-person perspective (3PP), where users observe their avatar externally [25]. While early VR research used virtual mirrors to reinforce embodiment, later work found that avatar visual attributes may be more influential. Krogmeier and Mousas [61] demonstrated that body type had a greater impact on ownership than mirror exposure, emphasizing the importance of avatar realism and congruence. Sensory feedback is also critical as congruent tactile input in 1PP enhances ownership, while 3PP reduces it [25].

Advancements in realism and motion tracking further refine embodiment. Wu et al. [113] found that avatars with full-body and facial tracking improved social presence and task performance in collaborative VR, emphasizing behavioral realism. VR can now simulate social interaction effectively. In Llobera et al. [68], users playing a mirror game with an autonomous agent exhibited synchronization levels comparable to those with human partners. These studies show how self-embodiment and interaction contribute to immersive VR experiences. Perspective also affects behavior and psychological responses. Won and Zhou [112] found that users, even without controlling their avatar, subtly adjusted their movements to match it, supporting the self-follower effect—an unconscious motor alignment with avatar motion—observed in both 1PP and 3PP.

However, the influence of perspective on higher-order behaviors remains a complex issue. Dupraz et al. [27] investigated whether embodiment amplifies the Proteus effect, a phenomenon in which behavioral shifts occur based on avatar identity, by having participants embody an elderly avatar under high and low embodiment conditions. Increased embodiment did not enhance stereotype-driven behavior, suggesting that while perspective influences motor alignment, it may not significantly affect identity-linked behavioral changes. The avatar customization perspective also shapes embodiment. Gonzalez-Franco et al. [39] found that customizing from 3PP reduced embodiment, especially when body sizes were altered. In contrast, 1PP customization increased body ownership and reduced implicit bias toward larger bodies, suggesting that direct 1PP experience supports psychological integration, while external observation weakens self-identification.

Last, the personalization of self-avatar also enhances embodiment further. Fielder et al. [32] showed that avatar personalization and active control independently increased embodiment and self-identification. Participants' weight judgments of virtual bodies were influenced by their own body mass index (BMI) and self-esteem, revealing a perceptual bias under high-embodiment. This demonstrates that embodiment in VR affects motor alignment, social interactions, self-perception, and implicit bias. While these studies establish how avatar personalization and control enhance embodiment, they raise critical questions about maintaining these benefits when direct control is disrupted, as occurs during autonomous demonstration phases essential for motor learning applications.

## 2.2 Body Image Perception

Building on embodiment principles, research demonstrates how multisensory integration reshapes body perception. Botvinick and Cohen [11] demonstrated the rubber hand illusion: when a fake hand is stroked in sync with a hidden real hand, people often feel it is "their" hand. This illusion extends to VR, where synchronous visuotactile or visuomotor cues induce a virtual hand illusion. Lin and Jörg [67] found that more realistic, anthropomorphic virtual hands elicited stronger ownership than non-human-like models such as wooden blocks. Armel and Ramachandran [3] showed that threatening a synchronized rubber hand triggered stress responses, indicating its acceptance as part of the body. Other studies have emphasized the dominance in recalibrating body awareness. Pavani et al. [83] found that visual cues could distort perceived tactile location. In contrast, Farne et al. [29] reported that simply seeing a rubber hand in a plausible position induced tactile neglect in brain-damaged patients. These findings demonstrate that the brain rapidly integrates visual, tactile, and proprioceptive inputs to modify the perception of body image.

Research also examined how the sensorimotor system adjusts to discrepancies between vision and proprioception. Welch [108] found that belief in a visual distortion affects proprioceptive adaptation, highlighting cognitive influences. Groen and Werkhoven [42] showed people can adjust to visuo-motor discrepancies when controlling a virtual hand, reflecting body representation plasticity. At the neural level, Graziano et al. [40, 41] identified premotor cortex neurons that integrate proprioception and visual input, even from fake limbs, to encode arm position. This supports the brain's ability to merge senses flexibly and prior beliefs to maintain bodily coherence. These neural mechanisms provide the foundation for virtual embodiment, where visual and proprioceptive cues must be carefully coordinated to achieve a seamless experience.

Translating these insights to virtual environments, researchers have shown how avatar characteristics systematically influence self-perception. IJsselsteijn et al. [47] showed that a first-person view of a virtual body can induce ownership, though typically weaker than in physical setups. Later work examined how embodying different avatars affects perception and behavior. Banakou et al. [6] found that embodying a child avatar led adults to overestimate object sizes and show shifts toward child-like self-identification. Osimo et al. [80] employed a virtual Sigmund Freud scenario in which participants counseled themselves. This altered self-perception and improved mood, especially when the avatar's movements matched their own.

## 2.3 Self-Similarity

When avatars resemble users themselves, they create unique psychological effects that extend beyond basic embodiment. Among other social and cognitive factors, self-similarity influences both trust and identification processes [5, 104]. Bailenson et al. [5] found that users responded differently to virtual agents wearing their own faces versus those of others. Verberne et al. [104] observed greater trust in assistants that mirrored users' behavior, appearance, and goals. In social VR, Shih et al. [95] found that self-resembling avatars were seen as more persuasive and easier to communicate with. These studies suggest users are more likely to trust and engage with virtual entities that reflect their identity, across both conversational and task-based settings.

Self-similarity also shapes embodiment and self-identification. Users tend to focus more on avatars that resemble themselves. Krogmeier and Mousas [62] found that increased visual fixation on self-avatars was linked to stronger perceived resemblance, reinforcing embodiment through self-recognition. Realism also enhances these effects. Specifically, Kim et al. [57] showed that self-similar avatars increased presence, and Salagean et al. [91] found that photorealistic, self-modeled avatars strengthened body ownership compared to generic ones. Fiedler et al. [32] demonstrated that facial texture and real-time motion tracking enhanced embodiment and self-identification, supporting the notion that multimodal self-resemblance deepens the user-avatar connection—even subtle behavioral congruence matters. Verberne et al. [104] found that agents mimicking user movements and goals fostered greater trust. However, these benefits of self-similarity must be balanced against the challenges of maintaining embodiment when users lose direct control.

Extending beyond visual similarity, voice personalization creates powerful bonds between users and avatars. Research consistently shows that self-similar voices enhance social presence [57], educational engagement [52, 76], and agent believability [43], suggesting that multimodal self-similarity may be particularly effective in maintaining connections during autonomous demonstrations.

However, extreme realism in avatars can backfire. Weisman and Peña [107] found that hyper-realistic avatars using users' faces caused uncanny valley effects, reducing trust. Aymerich-Franch et al. [4] reported that users with social anxiety felt increased stress when embodying self-similar avatars during public speaking. Likewise, Wolf et al. [111] found that users with low self-esteem rated photorealistic avatars of themselves as less attractive, struggling with appearance-based customization. These findings suggest that while personalization enhances engagement and embodiment, excessive realism can trigger discomfort, particularly in users who are sensitive to self-image or social evaluation.

The psychological benefits of personalization have made it popular in virtual health, training, and self-reflection. Feijóo-García et al. [30] found that users engaged more with counseling agents who resembled them in terms of personality or demographics, supporting the similarity-attraction principle in building trust. Alves da Silva et al. [2] explored avatar customization for self-reflection, demonstrating that users who adjusted their avatars to match their self-image experienced heightened self-awareness in a training simulation.

## 2.4 Agency

While self-similarity influences identification with avatars, the sense of agency, the feeling of being in control of avatar actions [102], represents another critical component of embodiment that becomes disrupted during autonomous demonstrations. Early research on bodily self-consciousness showed how agency and body ownership interact. Tsakiris and Haggard [101] revisited the rubber hand illusion, finding that synchronous visuotactile feedback induces body ownership, while top-down representations modulate its strength. These results highlighted the role of both sensory integration and cognitive processes in shaping bodily self, laying the groundwork for understanding agency. Building on this, Tsakiris et al. [102] identified distinct neural signatures for agency versus ownership, suggesting that active control engages different brain mechanisms than simply feeling ownership.

Active control over a virtual avatar strengthens agency and improves VR task performance. Salomon et al. [92] found that active self-movement enabled faster recognition of one's avatar among others, suggesting that agency aids in distinguishing self from others. Similarly, Kong et al. [60] showed that active control enhanced agency through a stronger intentional binding effect. Finally, Yun et al. [116] found that increased avatar motion fidelity improved both task performance and embodiment, reinforcing the link between agency and immersive task execution.

Various design and feedback factors can modulate agency. Banakou and Slater [7] demonstrated that body ownership over a virtual body can create illusory agency, even without actual control, although it does not affect real-world behavior. Koilias et al. [59] found that sensitivity to motion artifacts in avatars depended on the observation task. Last, Kim et al. [55] demonstrated that avatar appearance and motion synchrony significantly impacted agency and other embodiment components, thereby shaping users' sense of presence.

Recent work has also examined how agency interacts with other components of embodiment and explored methods for measuring them. Guy et al. [44] manipulated agency, ownership, and self-location in VR, demonstrating that these dimensions are interrelated and that aligning them enhances immersion. Building on this, Guy et al. [45] reviewed embodiment measures and advocated for dynamic, real-time assessments.

Last, personalization of avatars can influence users' sense of agency, even in the absence of motion synchrony. Jung et al. [49] explored personalized avatars under different motion conditions. Using face scanning and body-size matching, they found that motion synchrony increased agency across avatars. Even with asynchronous motion, personalized avatars elicited higher agency than generic ones, suggesting that appearance similarity can create an illusory sense of control, extending Banakou and Slater's [7] findings on illusory agency.

## 2.5 Motor Learning and Therapeutic VR Applications

Motor learning theory emphasizes that skill acquisition occurs through demonstration, observation, and practice phases [64]. Recent VR research has demonstrated that observational learning can outperform active learning in transferring skills to real-world tasks [33]. In therapeutic VR applications, the tension between embodiment and demonstration becomes particularly acute. While observational learning activates mirror neuron systems essential for motor skill acquisition [13, 93], effective rehabilitation also requires users to feel embodied in their avatars for optimal engagement and neuroplasticity [18, 85].

The efficacy of VR-based motor rehabilitation is significantly enhanced through embodied avatar experiences. These findings extend beyond rehabilitation to motor skill training in sports and professional contexts [78, 82]. Supporting this need for embodied experiences, Jung et al. [49] demonstrated that personalized avatars, combined with motion synchronization, create a sense of agency toward movements that participants did not actually perform, a finding with profound implications for patients with limited motor capabilities. This work demonstrates that self-similar avatars enhance both body ownership and agency compared to generic representations. A meta-analysis also confirms that avatar characteristics induce behavioral conformity through the Proteus effect [89]. The therapeutic impact extends beyond mere visual similarity; embodied first-person perspective avatars have been shown to increase cortical excitability and strengthen interhemispheric connections, directly supporting neuroplasticity mechanisms essential for motor recovery [14].

However, current therapeutic systems struggle to maintain engagement during autonomous demonstrations, where passive observation results in a suboptimal cognitive load [18, 93]. This challenge motivates our investigation of whether multimodal self-similarity can preserve embodiment benefits even when users lose direct avatar control.

Building on these embodiment findings, recent approaches recognize that effective motor learning interventions require the integration of multimodal feedback, with research demonstrating that combined visual-auditory-haptic systems outperform unimodal approaches [88, 96]. Voice feedback serves as a critical mechanism for motor control, enabling real-time monitoring and correction through auditory feedback loops [79]. Studies demonstrate that multimodal systems, which combine visual feedback with audio cues, enhance motor learning compared to single-modality approaches, with audio-visual integration producing higher satisfaction and engagement in rehabilitation contexts [26]. However, current implementations focus primarily on user-controlled interactions, leaving the challenge of maintaining engagement during autonomous demonstrations largely unaddressed.

## 2.6 Research Questions

In this study, we aimed to explore the effects of avatar self-similarity (i.e., in appearance and voice) as well as autonomy on various aspects of embodiment, user experience, and behavioral responses while also examining their potential interaction effects. Specifically, we seek to address the following research questions, categorized into three key areas: embodiment (**RQ1**), user experience (**RQ2**), and behavioral responses (**RQ3**).

- **Embodiment:** How does a self-similar avatar with autonomous behavior impact body ownership (**RQ1.1**), agency and motor control ratings (**RQ1.2**), perception of body location (**RQ1.3**), and perception of external appearance (**RQ1.4**)?
- **User Experience:** How does a self-similar avatar with autonomous behavior impact presence (**RQ2.1**), emotional reactivity (**RQ2.2**), frustration (**RQ2.3**), and self-attribution (**RQ2.4**)?
- **Behavioral Responses:** How do head-direction tracking, as a proxy for gaze, (**RQ3.1**) and task completion time (**RQ3.2**) reveal differences in participant behavior between self-similar and non-self-similar avatars across autonomous and non-autonomous conditions?

## 2.7 Contributions

This work significantly extends prior research by synthesizing insights from multimodal embodiment studies. Building on Richard et al.'s [90] demonstration of how haptic feedback enhances embodiment, we extend our multimodal investigation to include synchronized voice alongside visual appearance. While Kao et al. [52] demonstrated that voice similarity enhances performance in interactive contexts, our work investigates how combined voice and visual similarity affect embodiment during passive observation. Moreover, we build upon Fitton et al.'s [33] work on observational learning by investigating how self-similarity affects user experience during autonomous demonstrations, which is a critical yet understudied scenario. Also, unlike Fribourg et al.'s [35] virtual co-embodiment, where users shared control with another entity and could anticipate motions, we examine pre-programmed autonomous demonstrations where users observe predetermined movements of their self-similar avatars.

Overall, our study builds upon Jung et al.'s [49] findings on personalized avatars that maintain illusory agency by incorporating realistic voice cloning and examining the specific context of autonomous demonstrations, a scenario essential for motor learning and rehabilitation but previously unexplored in embodiment research. More importantly, we reveal that autonomy and self-similarity operate as independent factors, influencing embodiment in parallel, providing VR designers with the flexibility to adjust these parameters separately without complex interdependencies. More importantly, across our self-report measures, we did not detect a statistically significant autonomy × self-similarity interaction, suggesting largely additive influences on embodiment in this study. This pattern gives VR designers flexibility to tune autonomy and multimodal personalization as separate levers in similar mixed-control experiences. By integrating objective behavioral measures (i.e., head-direction tracking) with comprehensive subjective assessments, the research establishes evidence-based design principles for VR systems that balance expert demonstration requirements with the preservation of embodiment necessary for effective motor learning and engagement.

# 3 Methodology

## 3.1 Participants

To determine the necessary sample size for this study, we conducted an *a priori* power analysis using G*Power v3.1. This analysis was based on four conditions following a $2 \times 2$ factorial within-group design, examining the effects of two main factors: self-similarity of avatar appearance and voice (self-similarity: self-similar vs. non-self-similar) and whether or not the avatar demonstrates autonomous behaviors (autonomy: autonomous vs. non-autonomous). Based on one group with four repeated measures, a small effect size ($f = .25$) [24] and an alpha level of .05 to achieve 80% power ($1 - \beta$ error probability), the analysis recommended a sample size of 24 participants. We recruited the recommended sample size ($N = 24$) of participants through department-wide email distributions and in-class announcements. Of the sample (age: $M = 24.38$, $SD = 3.50$), 12 were female, 11 were male, and one preferred not to disclose. Our participants volunteered without receiving monetary compensation.

## 3.2 Virtual Reality Application

In this study, we developed a VR application utilizing the Unity game engine with the Final IK tool[1] for inverse kinematics animations and self-avatar control. We designed a semi-realistic living room as the virtual environment (see Figure 2), lit with a combination of directional and point lighting sources to produce subtle shadows and visual depth. At the center of the room, we placed a workstation featuring nine wooden blocks, each distinguished by a unique combination of textures and shapes, alongside an empty game board for the block assembly task. A completed wooden block castle served as the reference model. We selected this simple spatial reasoning task for its clarity and consistent pacing across participants. Additionally, we placed a virtual mirror in front of the workstation, allowing participants to observe their avatars and upper body movements, thus enhancing self-awareness and engagement with the virtual experience. Throughout all experimental conditions, participants maintained a first-person perspective, embodying their avatar and viewing the virtual environment through the avatar's eyes. Moreover, a virtual mirror provided a reflected

---

[1] http://root-motion.com/

**Figure 2: The virtual environment and task within our VR application. (a) Perspective view of the virtual environment. (b) Top-down view of the entire virtual room layout. (c) Close-up view of the block-assembling task area. (d) Top-down view of the block-assembling workspace.**

view of their embodied avatar while participants retained this first-person viewpoint.

The main task in this experiment was to assemble the wooden blocks. Block assembly tasks have been widely used in rehabilitation research, showing benefits for improving both motor skills and cognitive functions, including visuospatial working memory, executive functions, and fine motor coordination [65, 72, 75]. These three-dimensional construction paradigms have become popular methodologies in VR research for investigating spatial cognition and motor learning, offering ecological validity with experimental control [17, 71]. We incorporated intuitive guidance mechanisms for participants in designing the application. At the onset, the first block designated for placement is highlighted with an emissive material, encouraging participants to engage with it and place it on the game board. After the block was accurately placed, it snapped into its predetermined location. Then, the highlighting transitioned to the next sequential block, and the previously completed block reverted to its original texture. This iterative process continued until the construction of the block castle was complete.

In conditions featuring non-autonomous avatars, participants directly controlled their avatars to complete the assembly task, resembling a traditional VR gameplay scenario. Conversely, avatars independently completed the block puzzle task in the autonomous avatar scenarios, requiring no input from the participants. Here, the avatars' movements were animated using inverse kinematics. During these instances, participants could observe their avatars in action via the virtual mirror, which serves as a layer of self-observation and reflection on the experience. We illustrate the four conditions in Figure 3 and the accompanying video.

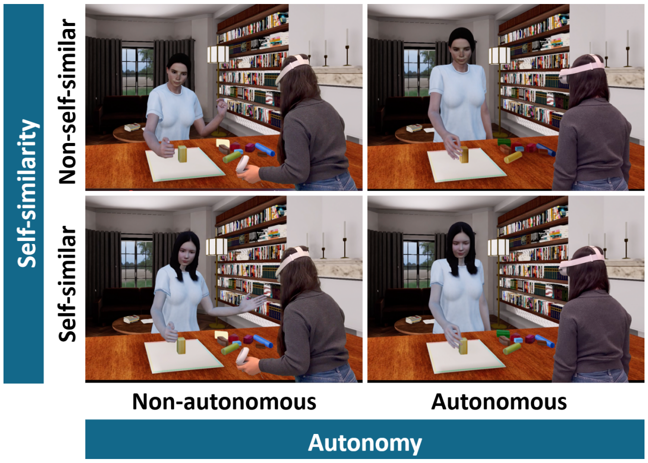

**Figure 3: A two-by-two matrix illustrating the four experimental conditions of the user study.**

*3.2.1 Virtual Characters and Voice Generating.* For this study, we selected two types of virtual characters. For the non-self-similar conditions, we used prompts tailored to match the demographics of university students from the Midwest, United States, to generate headshots of male and female individuals in MidJourney[2] and transformed these headshots into 3D models using Character Creator[3] v4.31 with the Headshot v2 plugin. In the experiments, we paired these models with the participants' self-reported sex assigned at birth. Consequently, participants who identified as non-binary were matched with an avatar corresponding to their biological sex characteristics to maintain the fidelity of the limb-length and joint-center mapping for the Final IK solver (see Figure 4). To create avatars closely resembling the participants, we utilized Character Creator v4.31 and the Headshot v2 plugin again, maintaining consistency across all conditions. We based these models on high-resolution photographs (6240×4160 pixels) taken with a Fujifilm XT-4 camera and a 35mm F2 lens, ensuring uniform lighting and background for consistency. When automated processes could not replicate specific features, such as beards, a 3D modeler manually added these details, enhancing self-similarity. The 3D modeler (27 years old with five years of 3D modeling experience) also manually adjusted the avatars' body shapes in Character Creator to approximate participants' heights and weights, thereby minimizing the impact of body type discrepancies during the VR interaction (see Figure 5).

We employed ElevenLabs' Text to Speech[4] service for the avatars' voices, selecting sex-matched, predefined voice models for non-self-similar conditions. For avatars meant to mirror participants' voices, we used ElevenLabs' voice cloning service, starting with participant recordings of "*The Tale of Peter Rabbit*" by Beatrix Potter [87]. We conducted these recordings in a noise-minimized lab environment and used Descript[5] v79.1.2 for any additional noise reduction. We then generated monologues that closely matched the scripted

---

[2] https://www.midjourney.com/imagine
[3] https://www.reallusion.com/character-creator/
[4] https://elevenlabs.io/text-to-speech
[5] https://www.descript.com/

**Figure 4: The non-self-similar virtual characters we used in our study.**

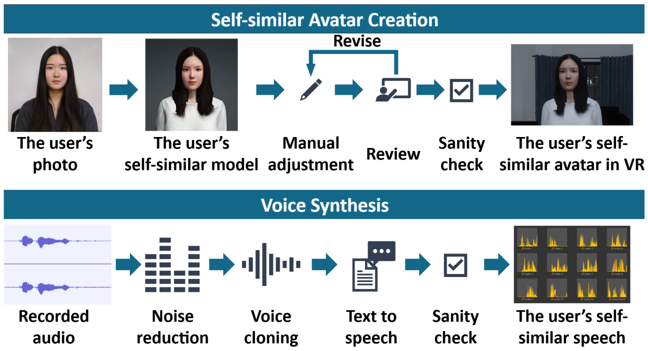

**Figure 5: Flowchart illustrating the generation pipeline for self-similar avatars' appearance (upper) and voice synthesis (lower).**

interactions and accurately reflected the participants' voices (see Table 1). We standardized voice behavior across all conditions through scripted monologues to isolate the effect of motor autonomy while maintaining consistent instructional feedback. This design decision ensures that differences between conditions stem from movement control rather than varying verbal content or interaction styles.

*3.2.2 Sanity Checks.* To validate the perceived self-similarity of the avatars and their voices, we conducted a two-phase sanity check involving both the research team and the study participants. In the initial phase, two computer graphics experts (age: $M = 33.00$, $SD = 8.49$; years of experience: $M = 8.00$, $SD = 5.66$) assessed the resemblance between the generated avatars and the participants. Each researcher voted on the fidelity of the avatars' appearance and voice. Disagreements prompted revisions, with the 3D modeler making adjustments based on the feedback. This iterative process continued until we achieved approval from the team, ensuring that both the visual and auditory representations met our criteria for self-similarity.

Following the expert evaluation, we moved to the second phase, where participants themselves played a critical role in affirming the avatars' plausible self-similarity. We presented each participant with a video showcasing their virtual counterpart, after which they completed a short survey. We utilized Qualtrics, an online survey tool, to distribute the survey and invited participants to rate, on a 7-point Likert scale, the degree to which they believed their avatar mirrored their appearance and voice. Evaluation statements were as follows: Q1: "*The virtual character looks like me.*" (1 = Not at all; 7 = Totally); Q2: "*The virtual character sounds like me.*" (1 = Not at all; 7 = Totally); Q3: "*The combined appearance and voice of the virtual character resemble me.*" (1 = Not at all; 7 = Totally). The results indicated that participants generally perceived their avatars as plausibly self-similar in both appearance and voice. Mean ratings were above the midpoint of the scale, suggesting a strong alignment between participants' expectations and virtual representations. Specifically, the appearance of the avatar (Q1: $M = 5.25$, $SD = .94$) and its voice (Q2: $M = 5.71$, $SD = 1.20$) were rated favorably, with the combined resemblance of appearance and voice (Q3: $M = 5.46$, $SD = 1.02$) also receiving a high score.

To validate dissimilarity, we applied the same procedure to the non-self-similar avatars, which were sex-matched but intentionally generic. Participants rated these avatars substantially lower across the same items, confirming that they were not perceived as resembling them. Specifically, the appearance (Q1: $M = 2.78$, $SD = 1.62$), voice (Q2: $M = 2.11$, $SD = 1.29$), and combined resemblance (Q3: $M = 2.56$, $SD = 1.26$) all fell below the scale's midpoint.

## 3.3 Experiment Conditions

We created four distinct experimental conditions to investigate how avatar self-similarity in appearance and voice, as well as the presence or absence of autonomous behaviors in the avatar during the block assembly task, affect participant perceptions and behavioral responses. The specific conditions are outlined as follows:

- **Non-self-similar Appearance & Voice with Non-autonomous Behavior (NSNA):** For this condition, we employed a sex-matched avatar we created from the generated headshots, as detailed in Section 3.2.1, ensuring low visual self-similarity with participants. Then, we synthesized the accompanying voice using a sex-matched voice model from ElevenLabs, which also lacked self-resemblance. The self-avatar demonstrated a non-autonomous behavior, functioning solely based on participant input.
- **Non-self-similar Appearance & Voice with Autonomous Behavior (NSA):** We maintained the use of a sex-matched non-self-similar avatar and introduced autonomous behaviors, meaning that the self-avatar would solve the puzzle on its own rather than mapping the movements of the participant's physical body.
- **Self-similar Appearance & Voice with Non-autonomous Behavior (SNA):** We used an avatar that visually and vocally mirrors the participant and generated it from participants' photographs and voice recordings. Despite this high degree of self-similarity, the avatar remained under the participant's direct control, with no autonomous actions.
- **Self-similar Appearance & Voice with Autonomous Behavior (SA):** We used an avatar with self-similar appearances and voices that exhibited autonomous behaviors.

## 3.4 Ratings and Measurements

*3.4.1 Survey.* We developed a survey (see Table 2) to assess participants' subjective experiences across eight variables: presence from

**Table 1: The monologues we used in our virtual reality experience.**

| Step | Monologue | Action |
|---|---|---|
| 1 | What a beautiful day! Perfect for a block game. Okay, only nine pieces. This should be easy. | Application start |
| 2 | First piece, you go to the corner. | Picking up the first block |
| 3 | Second piece, you're next to the first one! | Picking up the second block |
| 4 | Hmm, where should I put this one? | Picking up the third block |
| 5 | Piece four, let's see where you fit. | Picking up the fourth block |
| 6 | Halfway there, it's going well! | Picking up the fifth block |
| 7 | This is starting to look like a castle now. | Picking up the sixth block |
| 8 | Piece seven, where do you belong? | Picking up the seventh block |
| 9 | Only two left, I'm almost done. | Picking up the eighth block |
| 10 | Last one... | Picking up the ninth block |
| 11 | Done! That felt good. | Placed the last block |

Slater et al. [97], body ownership, agency and motor control, location of the body, and external appearance from Gonzalez-Franco and Peck [38], emotional reactivity from Mousas et al. [73], frustration from Choi et al. [22], and self-attribution from Wolf et al. [111]. Moreover, we also asked our participants to leave comments that they think would be useful for our study in a designated area in our survey at the end of the study.

*3.4.2 Application Logs.* In our application, we included a head-direction tracking system that uses head orientation as a coarse proxy for the user's gaze direction. While not a substitute for true eye-tracking, using head-mounted display orientation as an indicator of attentional focus is a method adopted in prior VR research [1, 20, 21, 23, 98]. However, we acknowledge that this method has limitations. It approximates the center of the user's field of view and cannot capture fine-grained eye movements, such as saccades or specific fixations, which are distinct from head orientation. We implemented a script that tracks the user's head orientation. The script calculates the center of the field of view to determine the approximate location of the user's viewpoint. A ray is cast from the center point of the HMD's view in the direction the user's head is facing, and any objects hit by the ray are logged along with the timestamp. The orientation data is logged continuously at the application's frame rate, which runs at 90 Hz on the Oculus Quest 2, collecting 90 gaze samples per second. Upon application exit, the recorded data is saved in a structured directory for statistical analysis.

For our head-direction data, we categorized it into four groups based on the names of the game objects in the virtual environment. Objects containing "Hand," "Arm," or "Torso" were grouped into the Body category to represent head orientation toward the self. Objects labeled as "Mirror" were assigned to the Mirror group, head-direction associated with self-reflection. Any objects related to tasks, such as "Block," "Game board," or "Example," were categorized under Task to capture task-relevant head orientation. The remaining objects, such as "Walls" or "Desk," were grouped into the Environment. This grouping allows us to analyze head-direction distribution across these four categories and examine how autonomy and self-similarity influence orientation patterns.

We did not include traditional performance comparisons (e.g., task accuracy, execution quality, or movement efficiency) because these metrics were not meaningful or comparable across our conditions. In half of the experiment (i.e., the autonomous avatar conditions), the block-assembly task was performed entirely by our system rather than by the participant. As a result, any performance outcome would reflect the autonomy controller rather than user behavior or user-avatar interaction. Moreover, the autonomous avatar was not optimized to perform the block-assembly task in a maximally efficient or "correct" manner, and, to our knowledge, no clearly defined optimal trajectory exists for this task in VR. Therefore, even if performance differences were measurable, they would not provide interpretable insights about embodiment or user experience. For these reasons, our analysis focuses on perceptual, experiential, and behavioral indicators (e.g., subjective ratings and head-direction patterns), which more accurately reflect the constructs under investigation.

## 3.5 Procedure

*3.5.1 Appointment 1: Initial Setup and Data Collection.* Upon participants' arrival at our research laboratory, we welcomed them and provided a brief introduction to the study's objectives. We initiated a pre-screening process to confirm eligibility, ensuring that participants had no history of severe motion sickness or other common risk factors typically screened for in VR research. These included self-reported issues such as photosensitive epilepsy, recent head injuries, uncorrected vision impairments, or adverse reactions to VR exposure. Eligible participants then reviewed and signed a consent form approved by our university's Institutional Review Board (IRB), which detailed the study's scope and rights. Following consent, we proceeded with data collection for avatar personalization. We captured high-resolution photographs of each participant under consistent lighting and background conditions. We also recorded a one-minute voice sample as participants read a prepared script aloud for voice cloning.

*3.5.2 Between Appointments: Avatar and Environment Preparation.* Using the collected photographs and voice recordings, we crafted self-similar avatars, emphasizing a high degree of resemblance in

**Table 2: The survey we used in our study.**

| # | Item | Anchors of the Scales |
|---|---|---|
| | **Presence** | |
| 1 | To what extent were there times during the experience when the computer-generated world became the "reality" for you, and you almost forgot about the "real world" outside? | 1 = Not at all, 7 = All the time |
| 2 | When you think back to the experience, do you think of the virtual environment more as images that you saw or more as somewhere that you visited? | 1 = Images, 7 = Somewhere I visited |
| 3 | During the time of the experience, which was the strongest on the whole, your sense of being in the virtual environment or of being elsewhere? | 1 = Being elsewhere, 7 = Being in the virtual environment |
| 4 | During the time of your experience, did you often think to yourself that you were actually in the virtual environment? | 1 = Not at all, 7 = Very much |
| | **Body Ownership** | |
| 5 | I felt as if the virtual body was my body. | 1 = Strongly disagree, 7 = Strongly agree |
| 6 | It felt as if the virtual body I saw was someone else. | 1 = Strongly disagree, 7 = Strongly agree |
| 7 | It seemed as if I might have more than one body. | 1 = Strongly disagree, 7 = Strongly agree |
| 8 | The body I saw in the mirror felt like mine. | 1 = Strongly disagree, 7 = Strongly agree |
| 9 | The body I saw in the mirror felt like that of another person. | 1 = Strongly disagree, 7 = Strongly agree |
| | **Agency and Motor Control** | |
| 10 | I could control the virtual body as if it were my own. | 1 = Strongly disagree, 7 = Strongly agree |
| 11 | The virtual body's movements were caused by mine. | 1 = Strongly disagree, 7 = Strongly agree |
| 12 | The virtual body's movements influenced mine. | 1 = Strongly disagree, 7 = Strongly agree |
| 13 | The virtual body moved by itself. | 1 = Strongly disagree, 7 = Strongly agree |
| | **Location of the Body** | |
| 14 | My body felt located where I saw the virtual body. | 1 = Strongly disagree, 7 = Strongly agree |
| 15 | I felt out of my body. | 1 = Strongly disagree, 7 = Strongly agree |
| 16 | My body or the avatar felt like they were drifting toward each other. | 1 = Strongly disagree, 7 = Strongly agree |
| | **External Appearance** | |
| 17 | My real body felt like it was turning into an avatar. | 1 = Strongly disagree, 7 = Strongly agree |
| 18 | My real body felt like it was taking on the avatar's posture. | 1 = Strongly disagree, 7 = Strongly agree |
| 19 | The avatar resembled my real body. | 1 = Strongly disagree, 7 = Strongly agree |
| | **Emotional Reactivity** | |
| 20 | I felt uneasy with this virtual avatar. | 1 = Strongly disagree, 7 = Strongly agree |
| 21 | The avatar's motion made me uncomfortable. | 1 = Strongly disagree, 7 = Strongly agree |
| 22 | The avatar's appearance made me uncomfortable. | 1 = Strongly disagree, 7 = Strongly agree |
| | **Frustration** | |
| 23 | I felt frustrated in this virtual environment. | 1 = Strongly disagree, 7 = Strongly agree |
| | **Self-attribution** | |
| 24 | I felt like the virtual avatar was me. | 1 = Strongly disagree, 7 = Strongly agree |
| 25 | I could identify with the virtual avatar. | 1 = Strongly disagree, 7 = Strongly agree |
| 26 | The virtual avatar behaved how I would. | 1 = Strongly disagree, 7 = Strongly agree |
| 27 | The avatar had similar attributes to me. | 1 = Strongly disagree, 7 = Strongly agree |
| | **Overall Experience** | |
| 28 | Please provide additional comments on your overall experience. | Open-ended response |

appearance and voice. This phase also included a two-stage sanity check involving our research team and the participants to validate the avatars' self-similarity in appearance and voices. We also applied the sanity checks to the non-self-similar avatars, confirming that their appearance and voices were not inadvertently too close to participants, as detailed in Section 3.2.2.

*3.5.3 Appointment 2: Experimental Engagement.* During the second visit, participants engaged in a brief tutorial to familiarize themselves with the VR setup and controls, ensuring a smooth and comfortable user experience, as prior research indicated that tutorials substantially improve control learnability and user experience [51]. We then introduced participants to the four distinct experimental conditions, each exploring various combinations of avatar appearance and voice (i.e., self-similar vs. non-self-similar) and autonomy (i.e., autonomous vs. non-autonomous). We employed the Latin squares method [109] to balance these conditions, eliminating potential carry-over (residual) effects. Following each experimental condition, we asked participants to complete our survey (see Section 3.4.1 and Table 2). After completing all experimental conditions and the corresponding questionnaires, we thanked the participants for their contribution before concluding the session. We provide an overview of the study procedure in Figure 6.

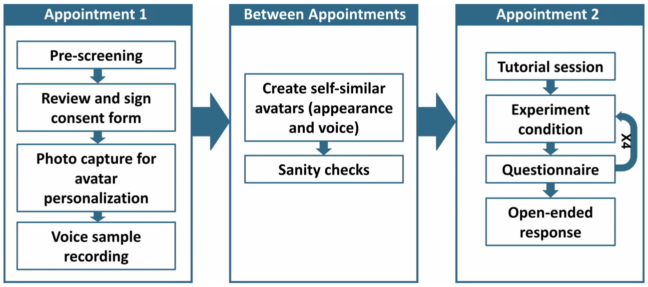

**Figure 6: Overview of the study procedure. Appointment 1: pre-screening, consent, photo capture, and voice recording. Between appointments: self-similar avatar creation and sanity checks. Appointment 2: tutorial, four counterbalanced experimental conditions, questionnaires, and open-ended responses.**

## 4 Results

In our statistical analyses, we used the autonomy and self-similarity factors as independent variables, and the self-reported ratings and behavioral responses as dependent variables. The Q-Q plots of the residuals and the Shapiro-Wilk test at the 5% level confirmed the normality of the collected data. Thus, we performed a two-way repeated measures analysis of variance (RM-ANOVA) for each variable using IBM's SPSS software v.25. We provide detailed results of our self-reported data in Table 3 and behavioral responses in Table 4. For any post-hoc pairwise tests, we applied a Bonferroni correction to control for multiple comparisons.

## 4.1 Self-reported Data

### 4.1.1 Embodiment.

*Body Ownership.* We did not find a statistically significant main effect of autonomy (Wilk's $\Lambda = .994$, $F[1,23] = .145$, $p = .707$, $\eta_p^2 = .006$), self-similarity (Wilk's $\Lambda = .997$, $F[1,23] = .067$, $p = .797$, $\eta_p^2 = .003$), or an autonomy × self-similarity interaction effect (Wilk's $\Lambda = .986$, $F[1,23] = .317$, $p = .579$, $\eta_p^2 = .014$).

*Agency and Motor Control.* We found a statistically significant main effect of autonomy (Wilk's $\Lambda = .291$, $F[1,23] = 56.100$, $p < .001$, $\eta_p^2 = .709$), indicating that participants rated their agency and motor control higher when exposed to non-autonomous ($M = 4.40$, $SE = .14$) than autonomous ($M = 3.34$, $SE = .11$) avatars. However, we did not find a significant main effect of self-similarity (Wilk's $\Lambda = .879$, $F[1,23] = 3.160$, $p = .089$, $\eta_p^2 = .121$), nor a significant autonomy × self-similarity interaction effect (Wilk's $\Lambda = .995$, $F[1,23] = .126$, $p = .726$, $\eta_p^2 = .005$).

*Location of the Body.* We did not find a statistically significant main effect of autonomy (Wilk's $\Lambda = .985$, $F[1,23] = .355$, $p = .557$, $\eta_p^2 = .015$). In contrast, self-similarity showed a statistically significant effect (Wilk's $\Lambda = .779$, $F[1,23] = 6.542$, $p = .018$, $\eta_p^2 = .221$), as participants reported a stronger sense that their body was located where they saw the virtual body in the self-similar ($M = 4.21$, $SE = .14$) than the non-self-similar ($M = 3.70$, $SE = .18$) conditions. Finally, we did not find a statistically significant autonomy × self-similarity interaction effect (Wilk's $\Lambda = .999$, $F[1,23] = .016$, $p = .901$, $\eta_p^2 = .001$).

*External Appearance.* We found a statistically significant main effect of autonomy (Wilk's $\Lambda = .828$, $F[1,23] = 4.779$, $p = .039$, $\eta_p^2 = .172$), indicating that participants rated their external appearance higher when exposed to non-autonomous ($M = 3.89$, $SE = .26$) than autonomous ($M = 3.38$, $SE = .17$) avatars. Also, self-similarity was statistically significant (Wilk's $\Lambda = .455$, $F[1,23] = 27.550$, $p < .001$, $\eta_p^2 = .545$), with participants rating self-similar avatars higher ($M = 4.15$, $SE = .23$) than non-self-similar avatars ($M = 3.12$, $SE = .19$). However, we did not find a statistically significant autonomy × self-similarity interaction effect (Wilk's $\Lambda = .970$, $F[1,23] = .706$, $p = .410$, $\eta_p^2 = .030$).

### 4.1.2 User Experience.

*Presence.* We found a statistically significant main effect of autonomy (Wilk's $\Lambda = .573$, $F[1,23] = 17.166$, $p < .001$, $\eta_p^2 = .427$), with participants reporting a higher presence in the non-autonomous ($M = 4.85$, $SE = .20$) than the autonomous ($M = 4.16$, $SE = .24$) conditions. We also found a significant main effect of self-similarity (Wilk's $\Lambda = .735$, $F[1,23] = 8.282$, $p = .008$, $\eta_p^2 = .265$), as self-similar avatars elicited a higher presence ($M = 4.67$, $SE = .20$) than non-self-similar avatars ($M = 4.34$, $SE = .22$). However, we did not find a statistically significant autonomy × self-similarity interaction effect (Wilk's $\Lambda = .999$, $F[1,23] = .020$, $p = .890$, $\eta_p^2 = .001$).

*Emotional Reactivity.* We found a statistically significant main effect of autonomy (Wilk's $\Lambda = .500$, $F[1,23] = 23.043$, $p < .001$, $\eta_p^2 = .500$), with autonomous avatars eliciting stronger emotional responses ($M = 4.01$, $SE = .27$) compared to non-autonomous avatars ($M = 3.22$, $SE = .20$). However, there was no statistically significant main effect of self-similarity (Wilk's $\Lambda = .900$, $F[1,23] = 2.548$, $p = .124$, $\eta_p^2 = .100$), nor a significant autonomy × self-similarity interaction effect (Wilk's $\Lambda = .991$, $F[1,23] = .207$, $p = .654$, $\eta_p^2 = .009$).

*Frustration.* We found a statistically significant main effect of autonomy (Wilk's $\Lambda = .754$, $F[1,23] = 7.494$, $p = .012$, $\eta_p^2 = .246$), with autonomous avatars leading to higher frustration ($M = 3.38$, $SE = .31$) compared to non-autonomous avatars ($M = 2.52$, $SE = .22$). There was no statistically significant main effect of self-similarity (Wilk's $\Lambda = .990$, $F[1,23] = .222$, $p = .642$, $\eta_p^2 = .010$), nor a significant autonomy × self-similarity interaction effect (Wilk's $\Lambda = .958$, $F[1,23] = 1.000$, $p = .328$, $\eta_p^2 = .042$).

**Table 3: Detailed results from the self-report measures in our study (we present significant results with bold font). NSNA: Non-self-similar Appearance & Voice with Non-autonomous Behavior; NSA: Non-self-similar Appearance & Voice with Autonomous Behavior; SNA: Self-similar Appearance & Voice with Non-autonomous Behavior; and SA: Self-similar Appearance & Voice with Autonomous Behavior. Autonomy $df = 1$, Self-similarity $df = 1$, Interaction $df = 1$, and Error $df = 23$. Notes: (1) Body Ownership, (2) Agency and Motor Control, (3) Location of the Body, (4) External Appearance, (5) Presence, (6) Emotional Reactivity, (7) Frustration, and (8) Self-Attribution.**

| | (1) | | (2) | | (3) | | (4) | | (5) | | (6) | | (7) | | (8) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | M | SD | M | SD | M | SD | M | SD | M | SD | M | SD | M | SD | M | SD |
| NSNA | 4.00 | .77 | 4.28 | .83 | 3.64 | 1.01 | 3.44 | 1.31 | 4.70 | 1.05 | 2.99 | 1.37 | 2.67 | 1.31 | 3.08 | 1.28 |
| SNA | 3.98 | .51 | 4.51 | .88 | 4.17 | .86 | 4.33 | 1.46 | 5.00 | 1.08 | 3.46 | 1.38 | 2.38 | 1.01 | 4.58 | 1.40 |
| NSA | 3.99 | .97 | 3.18 | .73 | 3.76 | 1.20 | 2.80 | 1.02 | 3.99 | 1.38 | 3.69 | 1.65 | 3.33 | 1.83 | 2.25 | .93 |
| SA | 4.09 | .64 | 3.50 | .77 | 4.25 | .74 | 3.96 | 1.11 | 4.33 | 1.11 | 4.33 | 1.63 | 3.42 | 1.72 | 3.94 | 1.32 |
| Autonomy (Main Effect) | | | | | | | | | | | | | | | | |
| F | .145 | | **56.100** | | .355 | | **4.779** | | **17.166** | | **23.043** | | **7.494** | | **23.286** | |
| p | .707 | | < **.001** | | .557 | | **.039** | | < **.001** | | < **.001** | | **.012** | | < **.001** | |
| $\eta_p^2$ | .006 | | **.709** | | .015 | | **.172** | | **.427** | | **.500** | | **.246** | | **.503** | |
| Self-similarity (Main Effect) | | | | | | | | | | | | | | | | |
| F | .067 | | 3.160 | | **6.542** | | **27.550** | | **8.282** | | 2.548 | | .222 | | **37.951** | |
| p | .797 | | .089 | | **.018** | | < **.001** | | **.008** | | .124 | | .642 | | < **.001** | |
| $\eta_p^2$ | .003 | | .121 | | **.221** | | **.545** | | **.265** | | .100 | | .010 | | **.623** | |
| Autonomy × Self-similarity (Interaction Effect) | | | | | | | | | | | | | | | | |
| F | .317 | | .126 | | .016 | | .706 | | .020 | | .207 | | 1.000 | | .323 | |
| p | .579 | | .726 | | .901 | | .410 | | .890 | | .654 | | .328 | | .575 | |
| $\eta_p^2$ | .014 | | .005 | | .001 | | .030 | | .001 | | .009 | | .042 | | .014 | |

*Self-Attribution.* We found a statistically significant main effect of autonomy (Wilk's $\Lambda = .497$, $F[1, 23] = 23.286$, $p < .001$, $\eta_p^2 = .503$), with non-autonomous avatars beeing rated higher ($M = 3.83$, $SE = .23$) than autonomous avatars ($M = 3.09$, $SE = .17$). Similarly, we found a significant main effect of self-similarity (Wilk's $\Lambda = .377$, $F[1, 23] = 37.951$, $p < .001$, $\eta_p^2 = .623$), with self-similar avatars receiving higher ratings ($M = 4.26$, $SE = .26$) compared to non-self-similar avatars ($M = 2.67$, $SE = .20$). However, we did not find a statistically significant autonomy × self-similarity interaction effect (Wilk's $\Lambda = .986$, $F[1, 23] = .323$, $p = .575$, $\eta_p^2 = .014$).

## 4.2 Behavioral Responses

*Environment.* We found a statistically significant main effect of autonomy on the percentage of time participants spent orienting their head toward the environment (Wilk's $\Lambda = .481$, $F[1, 23] = 24.843$, $p < .001$, $\eta_p^2 = .519$), with participants devoting more dwell time to the environment in the non-autonomous condition ($M = .57$, $SE = .03$) compared to the autonomous condition ($M = .36$, $SE = .04$). However, there was no statistically significant main effect of self-similarity (Wilk's $\Lambda = .871$, $F[1, 23] = 3.400$, $p = .078$, $\eta_p^2 = .129$), nor a significant autonomy × self-similarity interaction effect (Wilk's $\Lambda = .877$, $F[1, 23] = 3.229$, $p = .086$, $\eta_p^2 = .123$).

*Mirror.* We found a statistically significant main effect of autonomy (Wilk's $\Lambda = .302$, $F[1, 23] = 53.096$, $p < .001$, $\eta_p^2 = .698$), with participants devoting significantly more dwell time to the mirror in the autonomous condition ($M = .63$, $SE = .04$) compared to the non-autonomous condition ($M = .34$, $SE = .02$). We also found a significant main effect of self-similarity (Wilk's $\Lambda = .777$, $F[1, 23] = 6.595$, $p = .017$, $\eta_p^2 = .223$), with self-similar avatars leading to greater mirror-directed head orientation ($M = .52$, $SE = .03$) than non-self-similar avatars ($M = .45$, $SE = .03$). However, the autonomy × self-similarity interaction effect was not statistically significant (Wilk's $\Lambda = .916$, $F[1, 23] = 2.116$, $p = .159$, $\eta_p^2 = .084$).

*Body.* We found a statistically significant main effect of autonomy (Wilk's $\Lambda = .723$, $F[1, 23] = 8.811$, $p = .007$, $\eta_p^2 = .277$), with participants devoting more dwell time to their virtual body in the non-autonomous condition ($M = .05$, $SE = .02$) compared to the autonomous condition ($M = .01$, $SE = .00$). We did not find a significant main effect of self-similarity (Wilk's $\Lambda = .995$, $F[1, 23] = .108$, $p = .745$, $\eta_p^2 = .005$), nor a significant autonomy × self-similarity interaction effect (Wilk's $\Lambda = .974$, $F[1, 23] = .622$, $p = .438$, $\eta_p^2 = .026$).

**Table 4: Detailed results from the application log in our study (we present significant results with bold font). NSNA: Non-self-similar Appearance & Voice with Non-autonomous Behavior; NSA: Non-self-similar Appearance & Voice with Autonomous Behavior; SNA: Self-similar Appearance & Voice with Non-autonomous Behavior; and SA: Self-similar Appearance & Voice with Autonomous Behavior. Self-similarity $df = 1$, Autonomy $df = 1$, Interaction $df = 1$, and Error $df = 23$. Notes: (1) Body, (2) Environment, (3) Mirror, (4) Task, and (5) Time.**

| | (1) | | (2) | | (3) | | (4) | | (5) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | M | SD | M | SD | M | SD | M | SD | M | SD |
| NSNA | .05 | .08 | .57 | .18 | .32 | .13 | .02 | .02 | 84.40 | 27.53 |
| NSA | .06 | .09 | .57 | .14 | .36 | .13 | .02 | .02 | 82.30 | 28.56 |
| SNA | .01 | .01 | .41 | .25 | .58 | .26 | .01 | .01 | 87.18 | 1.08 |
| SA | .01 | .01 | .31 | .19 | .68 | .19 | .00 | .00 | 87.20 | 1.00 |
| Autonomy (Main Effect) | | | | | | | | | | |
| F | | **8.811** | | **24.843** | | **53.096** | | **14.444** | | .498 |
| p | | **.007** | | < **.001** | | < **.001** | | **.001** | | .487 |
| $\eta_p^2$ | | **.277** | | **.519** | | **.698** | | **.386** | | .021 |
| Self-similarity (Main Effect) | | | | | | | | | | |
| F | | .108 | | 3.400 | | **6.595** | | .710 | | .062 |
| p | | .745 | | .078 | | **.017** | | .408 | | .806 |
| $\eta_p^2$ | | .005 | | .086 | | **.223** | | .030 | | .003 |
| Autonomy × Self-similarity (Interaction Effect) | | | | | | | | | | |
| F | | .622 | | 3.229 | | 2.116 | | **4.516** | | .068 |
| p | | .438 | | .086 | | .159 | | **.045** | | .796 |
| $\eta_p^2$ | | .026 | | .123 | | .084 | | **.164** | | .003 |

*Task.* We found a statistically significant main effect of autonomy (Wilk's $\Lambda = .614$, $F[1, 23] = 14.444$, $p = .001$, $\eta_p^2 = .386$), with participants devoting more dwell time to task-related objects in the non-autonomous condition ($M = .02$, $SE = .00$) compared to the autonomous condition ($M = .01$, $SE = .00$). There was no statistically significant main effect of self-similarity (Wilk's $\Lambda = .970$, $F[1, 23] = .710$, $p = .408$, $\eta_p^2 = .030$). However, we found a significant autonomy × self-similarity interaction effect (Wilk's $\Lambda = .836$, $F[1, 23] = 4.516$, $p = .045$, $\eta_p^2 = .164$). To examine this interaction further, we conducted post-hoc pairwise $t$-tests. When the avatar was non-self-similar, participants spent more dwell time with their head oriented toward task-related objects in the non-autonomous ($M = .02$, $SD = .02$) than in the autonomous ($M = .01$, $SD = .01$) condition ($t[23] = 2.36$, $p = .027$, $d = .48$). When the avatar was self-similar, participants spent more dwell time with their head oriented toward task-related objects in the non-autonomous ($M = .02$, $SD = .02$) than in the autonomous ($M = .00$, $SD = .00$) condition ($t[23] = 4.34$, $p < .001$, $d = .89$). We also examined the simple effects of self-similarity under non-autonomous and autonomous conditions; neither comparison was significant.

*Time.* We did not find a statistically significant main effect of autonomy (Wilk's $\Lambda = .979$, $F[1, 23] = .498$, $p = .487$, $\eta_p^2 = .021$), self-similarity (Wilk's $\Lambda = .997$, $F[1, 23] = .062$, $p = .806$, $\eta_p^2 = .003$), or an autonomy × self-similarity interaction effect (Wilk's $\Lambda = .997$, $F[1, 23] = .068$, $p = .796$, $\eta_p^2 = .003$) on the time participants spent finishing the task.

## 4.3 Qualitative Data

To complement the quantitative findings, we thematically analyzed participants' open-ended reflections on their self-avatar experiences, grouped into themes of *appearance*, *voice*, *movement*, and *overall impressions*.

Three participants commented on the self-avatar's visual resemblance (*appearance*): P6 noted it "*resembled me to some extent*" and would feel "*very real*" with additional customization, while P12 found mismatched body proportions and mouth movement that "*made me uneasy,*" and P16 felt "*strange to see myself on a body with different body features.*" Mirror reflections shaped identification for others; P2 wanted more mirror coverage to "*feel more confidence,*" and P8 described strongly identifying with the self-similar avatar in one session, stating it felt like "*a copy or clone of me... I was clashing with it.*" In contrast, the dissimilar avatar allowed them to "*lose myself in the virtual environment more.*"

*Voice* feedback varied: P11 found it "*really cool*" to hear a familiar voice in the environment, while P13 described it as strange to hear their voice saying "*artificial things,*" and P15 noted that the lip-sync animation "*looked very weird and scary.*"

Participants emphasized the importance of interactivity and *movement* realism. P9 preferred when they could "*freely move my hands,*" while P4 suggested refining "*the grab and hold, the length of* *the arms, and the volume of the body,*" and found the mouth movement "*very weird.*" P21 proposed adding elbow tracking to improve realism, suggesting "*add some sensor on the user's elbow to make the hand movement more vivid.*"

*Overall impressions* were largely positive: P3 described the experience as "*nice,*" and P10 wrote, "*Great job! I'm curious to see the results.*" P14 emphasized how "*the experience was more immersive due to the environment,*" while P21 was excited by face similarity, stating "*I always imagine that we could reconstruct our face in VR.*" Finally, three noted areas for improvement: P24 described moments of feeling "*disconnected*" and called the self-similar avatar "*creepy,*" while P5 mentioned "*I wasn't really focusing on the mirror*" during the task, and P26 noted "*it took me sometime to understand that there was a mirror.*"

## 5 Discussion

In this study, we examined the effects of multimodal self-similarity and avatar autonomy on embodiment, user experience, and behavioral responses in VR, and we discuss these findings in the following sections.

### 5.1 RQ1: Embodiment

The self-reported data on embodiment provide insights into how autonomy and self-similarity shape different facets of embodiment in complementary ways, revealing both overlaps and dissociations across body ownership, agency and motor control, location of the body, and external appearance.

- **RQ1.1:** Body ownership appeared relatively robust, as it did not vary across the conditions.

This suggested that the body ownership aspect of embodiment may be less sensitive to these manipulations. This finding contrasts with prior research emphasizing the role of visuo-proprioceptive synchronization in enhancing body ownership [101]. A possible explanation lies in the distinction between agency and body ownership: while agency arises from the perception of control over an avatar's actions, body ownership relies on multisensory integration, such as visuo-tactile and visuo-proprioceptive congruence, to create a unified sense of self [103]. Specifically, Steed et al. [99] found that a self-avatar alone was insufficient to enhance embodiment unless accompanied by meaningful interaction. This suggests that ownership is not solely a function of visual self-resemblance but instead requires sensorimotor coherence. In our study, the controller-based input, while sufficient to elicit agency, may not have provided the sensory congruence necessary to enhance body ownership.

- **RQ1.2:** Agency and motor control was closely tied to autonomy, with participants reporting stronger agency when directly operating the avatar.

This aligns with prior work that shows agency arises when users perceive a match between their motor intentions and the avatar's movements [66]. In these conditions, the avatar acted as an extension of the self, strengthening the perception of agency. In contrast, autonomous avatars executing predefined actions disrupted this mapping, leading to diminished agency ratings. The lack of a significant effect for self-similarity supports prior findings that visual resemblance alone is insufficient to enhance agency without motor congruence [80]. According to Kilteni et al. [53], agency relies critically on the match between predicted and actual sensory outcomes of movement, rather than on appearance. Moreover, recent findings by Boban et al. [10] suggest that while users may spontaneously follow an autonomous avatar under uncertainty, this follower effect does not preserve the same sense of agency and may instead lead to a dissociation between the self and the avatar.

- **RQ1.3:** Self-similarity reinforced body location, as users felt more strongly anchored in the body when the avatar resembled them. In contrast, autonomy did not significantly influence participants' perception of body location.

This suggests that the avatar's control mode did not alter their sense of spatial alignment with the virtual body. In the meantime, self-similarity had a significant effect, aligning with prior research showing that visual cues dominate spatial localization [83] and that visual realism, particularly self-resemblance, enhances self-location [34, 55]. Notably, this contrasts with Jung et al. [49], who found that motion, not appearance, influenced self-location. This divergence may result from our inclusion of both visual and vocal self-similarity, which likely reinforced self-recognition through multimodal cues. Moreover, our interactive task may have increased attentional engagement, further anchoring the body to the self-similar avatar. These differences showed how task demands and multimodal personalization can shape spatial embodiment in distinct ways.

- **RQ1.4:** Perceptions of external appearance were jointly influenced by autonomy and self-similarity, indicating that how the avatar looks cannot be separated from how it behaves.

This result reinforces the idea that visual resemblance strengthens self-attribution and identification, as proposed by Biocca et al. [9]. Importantly, our findings suggest that participants evaluated their avatars not only on aesthetic preference but also on the perceived alignment between the avatar and their self-image. Banakou et al. [6] provide critical evidence that altering avatar appearance, such as embodying a child-like body, can shift perceptual judgments and even implicit self-associations. This supports the idea that appearance influences how users internalize and project identity within a virtual body. Moreover, the higher ratings for self-similar avatars align with findings by Krogmeier and Mousas [62], who observed that participants tend to look more often at avatars they perceive as similar to themselves, suggesting a link between perceived similarity and visual attention patterns. The significant autonomy effect on external appearance ratings could further be explained by Kilteni et al.'s [53] research on embodiment, where motor control enhances agency and strengthens the sense of body ownership, leading users to incorporate the avatar into their self-representation and evaluate its appearance more favorably. This effect may also reflect self-affirmation processes, as Kang and Kim [50] demonstrated that avatar control serves as a self-affirming activity, creating psychological investment in the avatar's positive evaluation.

### 5.2 RQ2: User Experience

The self-reported data on user experience provide insights into how autonomy and self-similarity exert distinct influences, with control primarily modulating presence and frustration, and resemblance driving self-attribution.

- **RQ2.1:** Presence benefited from both self-similarity and control, with participants feeling more immersed when avatars resembled them and when they had agency over the avatar's body.

The increased presence in non-autonomous conditions likely resulted from the active engagement required to control the avatar, which enhances the sense of being an active participant in the virtual environment [97]. Self-similarity further enhanced presence by triggering self-referential processing [77] and promoting identification with the avatar [9]. Interestingly, this contrasts with Jung et al. [49], who found that presence was enhanced by motion synchrony but not significantly affected by appearance. This discrepancy may reflect differences in modality and task. While their study focused on visual personalization, our use of visual and vocal self-similarity may have enhanced the sense of being there.

- **RQ2.2:** Emotional responses intensified under autonomy, showing that observing one's avatar act independently can heighten affective engagement.

This may reflect the discomfort associated with observing one's avatar act independently, disrupting the coherence between self-perception and perceived control. This idea is supported by Gall et al. [37], who demonstrated that virtual embodiment amplifies emotional responses to visual stimuli. In their study, participants reported significantly higher arousal, dominance, and valence for positive stimuli when the illusion of body ownership and agency was successfully induced through synchronous visuotactile and visuomotor feedback. Interestingly, self-similarity did not significantly affect emotional reactivity, suggesting that the emotional impact of losing control outweighed the effect of visual or vocal resemblance. While Jung et al. [49] found no significant emotional effects from appearance and only modest effects from motion synchrony, these differences may be attributed to variations in study design, task demands, and emotion measurement approaches. In contrast, our results indicate the emotional impact of autonomous avatar behavior during embodied experiences, particularly in a first-person perspective where users lack direct control over the avatar.

- **RQ2.3:** Frustration was triggered by loss of control, even when the avatar was self-similar, highlighting the delicate balance between resemblance and autonomy.

This suggests that predefined avatar actions introduced a mismatch between participants' expectations and the avatar's behavior. We argue that this disruption likely weakened the sense of seamless control, increasing cognitive demands during the task. This aligns with recent findings by Kim et al. [56], who showed that reduced avatar control significantly lowers agency and disrupts embodiment. Frisco et al. [36] further emphasize that spatial prediction of one's body plays a central role in body ownership, and misalignment between expected and actual actions may lead to discomfort. In our study, the increased frustration in self-similar autonomous conditions may result from increased self-referential processing, as participants watched avatars resembling themselves act independently and unexpectedly. In contrast, participants in non-autonomous conditions could directly control their avatars, leading to alignment between intention and motion and reduced cognitive strain.

- **RQ2.4:** Self-attribution was consistently boosted by self-similarity, confirming that resemblance plays a central role in identification with the avatar.

This result aligns with self-awareness theory, which posits that self-focused attention intensifies when individuals observe actions that reflect their own traits or intentions [16]. In our study, the virtual mirror functioned as a tool for self-focusing, which potentially enhanced participants' awareness of their avatars. Recent work supports this interpretation. Frisco et al. [36] showed that alignment between predicted and observed body position plays a central role in embodiment. Kim et al. [56] found that, more than appearance alone, motor congruence drives agency and self-identification. Jo et al. [48] also demonstrated that visual resemblance enhances body ownership, particularly when users are actively engaged. Mirror-based self-observation further enhances these effects, which promote self-referential processing through alignment between internal body models and visual feedback [28, 69].

## 5.3 RQ3: Behavioral Responses

The logged head-direction data provides insights into how autonomy and self-similarity shape head orientation patterns, which serve as a proxy for visual attention. Despite using pseudo-gaze tracking rather than precise eye-tracking technologies, our findings demonstrated interesting trends across different gaze categories.

- **RQ3.1:** Autonomy was correlated with increased head orientation toward mirrors, whereas direct control was correlated with focus on the task and the avatar's body, illustrating how control dynamics shape exploratory behavior.

In the Environment category, we found a significant main effect of autonomy, with participants orienting their head more toward the environment in the non-autonomous conditions than in the autonomous conditions. This difference is consistent with the task demands across conditions: when participants directly controlled the avatar, they needed to actively complete the assembly task and could reasonably scan the surrounding environment for task-relevant objects and spatial cues. This result can be interpreted using the cognitive load theory [94], as participants in the non-autonomous conditions may have directed their head orientation outward to situate themselves within the task context. Conversely, the reduced environment-directed head orientation in the autonomous conditions likely reflects a cognitive shift toward observing the avatar's autonomous actions, as participants were relieved of direct task control. We also observed a significant effect of self-similarity, with participants spending more time observing the environment in the non-self-similar conditions. This may be because participants in the self-similar conditions were more captivated by their avatars, which closely resembled themselves, drawing their orientational focus away from the environment.

In the Mirror category, we found a significant effect of autonomy, with participants devoting more head orientation time to the mirror in the autonomous condition. This could reflect increased engagement with the avatar's actions, as users monitor their avatar while performing tasks independently. While increased head-direction toward the mirror may also suggest confusion or anxiety about the avatar's autonomous behavior, we must be cautious in our interpretation, as head orientation data alone cannot confirm the user's affective state. Prior research suggests virtual mirrors promote self-focused attention and facilitate self-recognition by visually linking agency and appearance [54, 110]. Moreover, prior studies demonstrate that mirror feedback enhances the sense of embodiment by supporting self-evaluation and full-body awareness, especially when the avatar's motion is visible but not directly controllable [74]. Our finding that self-similar avatars induce more mirror-directed head orientation complements prior works, which show that visual resemblance contributes to self-avatar recognition and intensifies self-referential processing, particularly when congruent visual and motion cues are present [54]. Thus, in our study, mirrors appear to serve not only as visual aids but also as focal points for head orientation during autonomous avatar behavior, especially when participants observe autonomous behaviors performed by a self-similar avatar. While this increased orientational focus may support observational learning, the specific cognitive and emotional processes underlying this gaze behavior require further investigation.

We observed a significant effect of autonomy on the Body category, with participants orienting their head more frequently toward their virtual bodies in non-autonomous conditions. This pattern likely reflects a greater need for self-monitoring when users directly control the avatar, consistent with prior findings that visuomotor synchrony is a key contributor to the sense of agency and embodiment [28]. In contrast, the reduced body-directed head orientation in the autonomous condition may suggest that participants shifted their attention to the mirror for indirect self-observation, as mirrors can function as external feedback sources when direct control is absent. Nakano and Narumi [74] demonstrated that mirrors and extended visual fields support embodiment by facilitating visual self-monitoring, especially when full-body feedback is inaccessible. Interestingly, self-similarity did not significantly affect body-focused head orientation behavior, suggesting that visual resemblance alone is insufficient to influence attention allocation when agency is constrained. While a prior study has shown that self-similar avatars enhance self-recognition and embodiment, they also emphasize the dominant role of motion congruence over appearance in driving self-identification [54]. This supports the idea that task demands, particularly those related to action monitoring, may override visual resemblance in directing attention toward the body.

Autonomy significantly influenced the percentage of dwell time participants spent with their head oriented toward task-related objects, with participants allocating more orientational focus to these elements in the non-autonomous condition ($M$ = .02) compared to the autonomous condition ($M$ = .005). These mean values represent the proportion of total gaze samples directed at task-relevant objects during each condition. Although the absolute numbers are small, the relative difference suggests that participants were more visually engaged with the task when they had direct control over their avatar's actions. In contrast, during autonomous conditions, participants may have shifted their focus toward observing the avatar's independent movements. Self-similarity did not significantly impact task-related head-direction, suggesting that avatar resemblance played a minimal role in directing visual attention. However, a significant interaction effect revealed that participants spent more dwell time with their head oriented toward task-related objects when the autonomous avatar was not self-similar, possibly indicating reduced social engagement with the dissimilar avatar and a redirection of head orientation toward the task.

- **RQ3.2:** Task completion time remained stable across conditions, indicating that experiential differences did not translate into measurable performance changes.

We thus argue that this consistency in exposure duration strengthens the internal validity of our study, as the observed differences in user experience and behavior are less likely to be attributed to time spent in the virtual environment and more likely to reflect the intended experimental manipulations.

## 5.4 Implications

Our findings suggest that self-similar appearance and voice can preserve aspects of embodiment even when users do not control the avatar directly. Although autonomous behavior reduced agency and self-attribution, self-similar avatars maintained presence and enhanced the sense of body location, indicating that visual and vocal resemblance may support embodiment during passive observation. This has practical relevance for systems where demonstration is necessary, such as in rehabilitation or training scenarios that require users to observe correct techniques.

Because we found significant main effects of self-similarity and autonomy (with no significant interaction), these factors appear to influence embodiment independently, suggesting that designers can adjust them individually. Since we did not detect a statistically significant autonomy × self-similarity interaction in our self-report analyses, we argue that our data are consistent with largely additive contributions of these factors to embodiment in this study, suggesting that designers can adjust them individually. Multimodal personalization can be used to improve user identification without interfering with motion autonomy, providing flexibility in applications that require both observational learning and a sense of personal relevance. By separating control from self-recognition, systems can support a wider range of user needs and capabilities without compromising engagement.

Last, behavioral responses further support this distinction. The increase in mirror-focused head-direction during autonomous and self-similar conditions, combined with reduced head orientation toward task elements, indicates that users shift their orientational focus from action execution to self-reflection. These patterns point to the potential of head-direction tracking as a non-invasive proxy for monitoring user engagement and embodiment in real-time. VR systems that respond to such behavioral cues may be able to adaptively balance demonstration and interaction phases more effectively.

## 5.5 Limitations

While this study provides meaningful insights into how autonomy and self-similarity influence user perception when embodying a virtual avatar, it is essential to acknowledge the limitations in our work, as they help to better interpret our findings. First, the categorization of head-direction tracking data relied on grouping object names into predefined categories (i.e., Body, Mirror, Task, and Environment). While this approach provided structured insights, it inherently simplified the complexity of attentional behaviors. Notably, the data analyzed in this study were based on head orientation derived from avatar head and body orientations rather than actual eye-tracking data. As a result, some subtleties of gaze direction and intent might have been overlooked. Thus, we argue that integrating eye-tracking technologies would enable researchers to achieve more precise and context-aware gaze tracking, offering richer insights into gaze behaviors.

Second, although the avatar's appearance and voice were carefully configured to reflect self-similarity, the fidelity of these representations was constrained by the tools used for voice cloning and avatar creation. For example, participants with distinct accents may have experienced reduced identification with their avatars, which could potentially affect their responses. We obtained consent for voice cloning from all participants, but we did not assess how they felt about this process or how it affected their sense of self. Recent work conceptualizes voice cloning as a tension between voice as personal identity and as commodified data and examines cases where cloned voices are misused in ways that compromise identity and security [8, 105], suggesting that comfort with voice cloning and perceived control over one's vocal identity may shape the acceptance of self-similar avatars. While self-similar representations went through a sanity check to ensure resemblance, perceived similarity in the non-self-similar condition was not explicitly verified, and there is a possibility that some participants perceived the "non-similar" avatar or voice as unintentionally familiar. Thus, we think that enhancing customization technologies and incorporating user feedback on perceived non-similarity would improve the robustness of similar studies.

Third, our autonomy manipulation focused specifically on motor control during the block assembly task, while voice behavior was intentionally pre-scripted across all conditions to isolate motor autonomy effects and ensure consistent instructional content. While this design choice enabled clear interpretation of motor autonomy effects, it limits the generalizability to fully multimodal autonomous interactions. As such, the findings should be interpreted within the context of this controlled, unimodal autonomy framework.

Fourth, we can argue that interaction realism was limited by the technologies used for motion capture and avatar control. Instead of employing a high-quality motion capture system, we relied on an inverse kinematics solver and VR controllers as inputs to control the avatar's movements. While this approach provided functional control, it may have lacked the precision required to achieve fully naturalistic avatar behaviors; therefore, the resulting interactions might not have fully reflected the subtleties of real-world movement dynamics.

Fifth, our avatar design approach prioritized morphological consistency, but imposed a binary sex constraint for non-self-similar avatars based on participants' sex assigned at birth. While effective for standardization, this approach may not reflect the full spectrum of gender identity, as it could risk gender dysphoria in non-cisgender participants [84] or impact perceived humanness [81]. We also did not screen for body dysmorphia, which may interact negatively with high-fidelity self-avatars. These are important considerations for future studies involving hyper-realistic personalization. Additionally, although generic avatars reflected the dominant demographics at our institution, we did not collect participants' ethnicity information or control avatar ethnicity or skin tone, which limits ecological validity and prevents us from assessing how ethnicity shapes embodiment. Although no concerns emerged in participant feedback, we argue that researchers should support diverse skin-tone calibration and non-binary morphologies. Finally, our sample consisted primarily of young adult university students; thus, generalization to clinical or older populations should be made cautiously.

Sixth, in the non-autonomous conditions, although participants retained full motor control over their avatars, the avatars' verbal output remained pre-scripted and autonomous. This mismatch between motor agency and vocal autonomy may have introduced an unintended hybrid autonomy condition, potentially reducing the perceived sense of full control. In our study, hearing an avatar speak independently, even while controlling its body, may have disrupted participants' perception of their avatar as an extension of themselves, particularly if the speech content felt incongruent with their intentions or self-image.

Finally, the scope of interactions in this study was constrained to a block-assembling task. While participants engaged with avatars in this specific context, the restricted task scope may not fully represent how autonomy and self-similarity influence interactions in broader, less-structured VR environments. Introducing more varied and complicated tasks, such as free-form conversations [115] or collaborative problem-solving [22], could provide a more comprehensive understanding of user-avatar interactions. Furthermore, we focused on immediate subjective and behavioral outcomes rather than long-term performance gains; thus, we argue that researchers should examine whether the embodiment benefits observed here translate into improved skill acquisition or task performance over time.

## 6 Conclusions and Future Work

We investigated the effects of multimodal self-similarity (i.e., visual appearance and voice) and avatar autonomy on embodiment, user experience, and behavioral responses in virtual reality. Our findings reveal that autonomy and self-similarity each had significant main effects on user experience, with no significant interactions observed. Critically, we demonstrated that self-similar avatars can maintain presence and self-attribution even during autonomous demonstrations, while autonomous conditions increased emotional reactivity and frustration compared to user-controlled conditions. Non-autonomous avatars enhanced agency, motor control, and self-attribution ratings. Our behavioral analysis revealed distinct head-direction patterns. In autonomous conditions, participants were more often oriented toward mirrors, while in non-autonomous conditions, they were more often oriented toward task elements and their own bodies. A single significant interaction effect was observed, indicating increased task-focused head orientation when non-self-similar avatars exhibited autonomous behavior.

Looking forward, several promising research directions emerge from our findings. For example, advanced pose estimation through systems like MobilePoser [114] opens possibilities for real-time personalized avatar control using consumer devices, making our multimodal self-similarity approach more accessible. Integration of neural acoustic fields [100] could enhance our voice synthesis approach with spatial audio that adapts to virtual environments, creating more convincing demonstrations. Moreover, as machine learning approaches have demonstrated promising accuracy in gaze-based intention prediction [19], integrating such methods could allow VR systems to anticipate when users wish to transition from observation to control, paving the way for more fluid and intuitive interaction paradigms. Streamlining the preparation pipeline for voice and avatar personalization, which is currently reliant on manual photo and audio capture, will also be key to more efficient deployment beyond a research lab. Future studies should also explore full multimodal autonomy by incorporating real-time voice control alongside movement control, and investigate these effects across more diverse task contexts beyond structured assembly tasks.

## References

[1] Isayas Berhe Adhanom, Paul MacNeilage, and Eelke Folmer. 2023. Eye tracking in virtual reality: a broad review of applications and challenges. *Virtual Reality* 27, 2 (2023), 1481–1505.

[2] Claudio Alves da Silva, Bernhard Hilpert, Chirag Bhuvaneshwara, Patrick Gebhard, Fabrizio Nunnari, and Dimitra Tsovaltzi. 2023. Visual similarity for socially interactive agents that support self-awareness. In *ACM Iternational Conference on Intelligent Virtual Agents*. 1–3.

[3] K Carrie Armel and Vilayanur S Ramachandran. 2003. Projecting sensations to external objects: evidence from skin conductance response. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270, 1523 (2003), 1499–1506.

[4] Laura Aymerich-Franch, René F Kizilcec, and Jeremy N Bailenson. 2014. The relationship between virtual self similarity and social anxiety. *Frontiers in human neuroscience* 8 (2014), 944.

[5] Jeremy N Bailenson, Andrew C Beall, Jim Blascovich, Mike Raimundo, and Max Weisbuch. 2001. Intelligent agents who wear your face: Users' reactions to the virtual self. In *International workshop on intelligent virtual agents*. 86–99.

[6] Domna Banakou, Raphaela Groten, and Mel Slater. 2013. Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences* 110, 31 (2013), 12846–12851.

[7] Domna Banakou and Mel Slater. 2017. Embodiment in a virtual body that speaks produces agency over the speaking but does not necessarily influence subsequent real speaking. *Scientific reports* 7, 1 (2017), 14227.

[8] Adam Eric Berkowitz and Miriam E Sweeney. 2025. Look Who's Talking: Voice Cloning as Tension Point Between Identity and Data. *Philosophy & Technology* 38, 4 (2025), 131.

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

[10] Loën Boban, Ronan Boulic, and Bruno Herbelin. 2024. In case of doubt, one follows one's self: the implicit guidance of the embodied self-avatar. *IEEE Transactions on Visualization and Computer Graphics* (2024).

[11] Matthew Botvinick and Jonathan Cohen. 1998. Rubber hands 'feel'touch that eyes see. *Nature* 391, 6669 (1998), 756–756.

[12] Niclas Braun, Stefan Debener, Nadine Spychala, Edith Bongartz, Peter Sörös, Helge HO Müller, and Alexandra Philipsen. 2018. The senses of agency and ownership: a review. *Frontiers in psychology* 9 (2018), 535.

[13] Giovanni Buccino. 2014. Action observation treatment: a novel tool in neurorehabilitation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369, 1644 (2014), 20130185.

[14] Karin A Buetler, Joaquin Penalver-Andres, Özhan Özen, Luca Ferriroli, René M Müri, Dario Cazzoli, and Laura Marchal-Crespo. 2022. "Tricking the brain" using immersive virtual reality: modifying the self-perception over embodied avatar influences motor cortical excitability and action initiation. *Frontiers in human neuroscience* 15 (2022), 787487.

[15] Luis Vicente Calderita, Luis J Manso, Pablo Bustos, Cristina Suárez-Mejías, Fernando Fernández, and Antonio Bandera. 2014. THERAPIST: towards an autonomous socially interactive robot for motor and neurorehabilitation therapies for children. *JMIR rehabilitation and assistive technologies* 1, 1 (2014), e3151.

[16] Charles S Carver and Michael F Scheier. 1978. Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence. *Journal of personality and social psychology* 36, 3 (1978), 324.

[17] Valentina Catania, Francesco Rundo, Simonetta Panerai, and Raffaele Ferri. 2023. Virtual reality for the rehabilitation of acquired cognitive disorders: a narrative review. *Bioengineering* 11, 1 (2023), 35.

[18] Matteo Ceradini, Elena Losanno, Silvestro Micera, Andrea Bandini, and Silvia Orlandi. 2024. Immersive VR for upper-extremity rehabilitation in patients with neurological disorders: A scoping review. *Journal of NeuroEngineering and Rehabilitation* 21, 1 (2024), 75.

[19] Xiao-Lin Chen and Wen-Jun Hou. 2022. Gaze-based interaction intention recognition in virtual reality. *Electronics* 11, 10 (2022), 1647.

[20] Minsoo Choi, Dixuan Cui, Alexandros Koilias, and Christos Mousas. 2024. The Effects of Virtual Character's Intelligence and Task's Complexity during an Immersive Jigsaw Puzzle Co-solving Task. In *Proceedings of the 17th ACM SIGGRAPH Conference on Motion, Interaction, and Games*. 1–12.

[21] Minsoo Choi, Dixuan Cui, Matias Volonte, Alexandros Koilias, Dominic Kao, and Christos Mousas. 2025. Toward Understanding the Effects of Intelligence of a Virtual Character during an Immersive Jigsaw Puzzle Co-Solving Task. *ACM Transactions on Applied Perception* 22, 2 (2025), 1–28.

[22] Minsoo Choi, Siqi Guo, Alexandros Koilias, Matias Volonte, Dominic Kao, and Christos Mousas. 2024. Exploring the Effects of Self-correction Behavior of an Intelligent Virtual Character during a Jigsaw Puzzle Co-solving Task. *ACM Transactions on Interactive Intelligent Systems* 14, 3 (2024), 1–33.

[23] Viviane Clay, Peter König, and Sabine Koenig. 2019. Eye tracking in virtual reality. *Journal of eye movement research* 12, 1 (2019), 10–16910.

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

[25] Dixuan Cui and Christos Mousas. 2022. Evaluating the sense of embodiment through out-of-body experience and tactile feedback. In *ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*. 1–7.

[26] Ilana Dubovi. 2022. Cognitive and emotional engagement while learning with VR: The perspective of multimodal methodology. *Computers & Education* 183 (2022), 104495.

[27] Louise Dupraz, Marine Beaudoin, Michel Guerraz, and Julien Barra. 2024. Does the avatar embodiment moderate the Proteus effect? *International Journal of Human-Computer Studies* 187 (2024), 103272.

[28] James Coleman Eubanks, Alec G Moore, Paul A Fishwick, and Ryan P McMahan. 2020. The effects of body tracking fidelity on embodiment of an inverse-kinematic avatar for male participants. In *IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 54–63.

[29] Alessandro Farne, Francesco Pavani, Francesca Meneghello, and Elisabetta Ladavas. 2000. Left tactile extinction following visual stimulation of a rubber hand. *Brain* 123, 11 (2000), 2350–2360.

[30] Pedro Guillermo Feijóo-García, Chase Wrenn, Alexandre Gomes de Siqueira, Rashi Ghosh, Jacob Stuart, Heng Yao, and Benjamin Lok. 2024. Exploring the Effects of User-Agent and User-Designer Similarity in Virtual Human Design to Promote Mental Health Intentions for College Students. *ACM Transactions on Applied Perception* 22, 1 (2024), 1–41.

[31] Luara Ferreira dos Santos, Oliver Christ, Kedar Mate, Henning Schmidt, Jörg Krüger, and Christian Dohle. 2016. Movement visualisation in virtual reality rehabilitation of the lower limb: a systematic review. *Biomedical engineering online* 15 (2016), 75–88.

[32] Marie Luisa Fielder, Erik Wolf, Nina Dollinger, David Mal, Mario Botsch, Marc Erich Latoschik, and Carolin Wienrich. 2024. From avatars to agents: Self-related cues through embodiment and personalization affect body perception in virtual reality. *IEEE Transactions on Visualization and Computer Graphics* (2024).

[33] Isabel Sophie Fitton, Elizabeth Dark, Manoela Milena Oliveira da Silva, Jeremy Dalton, Michael J Proulx, Christopher Clarke, and Christof Lutteroth. 2024. Watch this! observational learning in vr promotes better far transfer than active learning for a fine psychomotor task. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19.

[34] Rebecca Fribourg, Ferran Argelaguet, Anatole Lécuyer, and Ludovic Hoyet. 2020. 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 (2020), 2062–2072.

[35] Rebecca Fribourg, Nami Ogawa, Ludovic Hoyet, Ferran Argelaguet, Takuji Narumi, Michitaka Hirose, and Anatole Lécuyer. 2020. Virtual co-embodiment: evaluation of the sense of agency while sharing the control of a virtual body among two individuals. *IEEE Transactions on Visualization and Computer graphics* 27, 10 (2020), 4023–4038.

[36] Francesca Frisco, Vito Bruno, Daniele Romano, and Giorgia Tosi. 2024. I am where I believe my body is: The interplay between body spatial prediction and body ownership. *PloS one* 19, 12 (2024), e0314271.

[37] Dominik Gall, Daniel Roth, Jan-Philipp Stauffert, Julian Zarges, and Marc Erich Latoschik. 2021. Embodiment in virtual reality intensifies emotional responses to virtual stimuli. *Frontiers in Psychology* 12 (2021), 674179.

[38] Mar Gonzalez-Franco and Tabitha C Peck. 2018. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI* 5 (2018), 74.

[39] Mar Gonzalez-Franco, Anthony Steed, Christopher C Berger, and Ana Tajadura-Jiménez. 2024. The impact of first-person avatar customization on embodiment in immersive virtual reality. *Frontiers in Virtual Reality* 5 (2024), 1436752.

[40] Michael SA Graziano. 1999. Where is my arm? The relative role of vision and proprioception in the neuronal representation of limb position. *Proceedings of the National Academy of Sciences* 96, 18 (1999), 10418–10421.

[41] Michael SA Graziano, Dylan F Cooke, and Charlotte SR Taylor. 2000. Coding the location of the arm by sight. *Science* 290, 5497 (2000), 1782–1786.

[42] Joris Groen and Peter J Werkhoven. 1998. Visuomotor adaptation to virtual hand position in interactive virtual environments. *Presence* 7, 5 (1998), 429–446.

[43] 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.

[44] Martin Guy, Camille Jeunet-Kelway, Guillaume Moreau, and Jean-Marie Normand. 2022. Manipulating the Sense of Embodiment in Virtual Reality: a study of the interactions between the senses of agency, self-location and ownership. In *International Conference on Artificial Reality and Telexistence & Eurographics Symposium on Virtual Environments.* 1–11.

[45] Martin Guy, Jean-Marie Normand, Camille Jeunet-Kelway, and Guillaume Moreau. 2023. The sense of embodiment in virtual reality and its assessment methods. *Frontiers in Virtual Reality* 4 (2023), 1141683.

[46] Maureen K Holden. 2005. Virtual environments for motor rehabilitation. *Cyberpsychology & behavior* 8, 3 (2005), 187–211.

[47] Wijnand A IJsselsteijn, Yvonne A W de Kort, and Antal Haans. 2006. Is this my hand I see before me? The rubber hand illusion in reality, virtual reality, and mixed reality. *Presence: Teleoperators and Virtual Environments* 15, 4 (2006), 455–464.

[48] Dongsik Jo, Kangsoo Kim, Gregory F Welch, Woojin Jeon, Yongwan Kim, Ki-Hong Kim, and Gerard Jounghyun Kim. 2017. The impact of avatar-owner visual similarity on body ownership in immersive virtual reality. In *ACM Symposium on Virtual Reality Software and Technology.* 1–2.

[49] Myeongul Jung, Sangyong Sim, Jejoong Kim, Kwanguk Kim, et al. 2022. Impact of personalized avatars and motion synchrony on embodiment and users' subjective experience: empirical study. *JMIR Serious Games* 10, 4 (2022), e40119.

[50] Hyunjin Kang and Hye Kyung Kim. 2020. My avatar and the affirmed self: Psychological and persuasive implications of avatar customization. *Computers in Human Behavior* 112 (2020), 106446.

[51] 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.

[52] Dominic Kao, Rabindra Ratan, Christos Mousas, and Alejandra J Magana. 2021. The effects of a self-similar avatar voice in educational games. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (2021), 1–28.

[53] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments* 21, 4 (2012), 373–387.

[54] Bowon Kim, Eugene Hwang, and Jeongmi Lee. 2024. Self-Avatar Recognition in Virtual Reality: the Self Advantage Phenomenon and the Relative Importance of Motion and Visual Congruence. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR).* 643–651.

[55] Chang-Seop Kim, Myeongul Jung, So-Yeon Kim, Kwanguk Kim, et al. 2020. Controlling the sense of embodiment for virtual avatar applications: methods and empirical study. *JMIR Serious Games* 8, 3 (2020), e21879.

[56] DoHyung Kim, Halim Yeo, and Kyoungju Park. 2025. Effects of an Avatar Control on VR Embodiment. *Bioengineering* 12, 1 (2025), 32.

[57] Hayeon Kim, Jinhyung Park, and In-Kwon Lee. 2023. "To be or Not to be Me?": Exploration of Self-Similar Effects of Avatars on Social Virtual Reality Experiences. *IEEE Transactions on Visualization and Computer Graphics* 29, 11 (2023), 4794–4804.

[58] Daiki Kodama, Takato Mizuho, Yuji Hatada, Takuji Narumi, and Michitaka Hirose. 2023. Effects of collaborative training using virtual co-embodiment on motor skill learning. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2304–2314.

[59] Alexandros Koilias, Christos Mousas, and Christos-Nikolaos Anagnostopoulos. 2019. The effects of motion artifacts on self-avatar agency. In *Informatics*, Vol. 6. 18.

[60] Gaiqing Kong, Kang He, and Kunlin Wei. 2017. Sensorimotor experience in virtual reality enhances sense of agency associated with an avatar. *Consciousness and cognition* 52 (2017), 115–124.

[61] 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.

[62] 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 *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops.* IEEE, 127–134.

[63] Kate E Laver, Belinda Lange, Stacey George, Judith E Deutsch, Gustavo Saposnik, and Maria Crotty. 2017. Virtual reality for stroke rehabilitation. *Cochrane database of systematic reviews* 11 (2017).

[64] Timothy D Lee and Richard A Schmidt. 2025. *Motor learning and performance: From principles to application.* Human Kinetics.

[65] Muriel Deutsch Lezak. 2004. *Neuropsychological assessment.* Oxford University Press, USA.

[66] Hannah Limerick, David Coyle, and James W Moore. 2014. The experience of agency in human-computer interactions: a review. *Frontiers in human neuroscience* 8 (2014), 643.

[67] Lorraine Lin and Sophie Jörg. 2016. Need a hand? How appearance affects the virtual hand illusion. In *ACM Symposium on Applied Perception.* 69–76.

[68] Joan Llobera, Valentin Jacquat, Carmela Calabrese, and Caecilia Charbonnier. 2022. Playing the mirror game in virtual reality with an autonomous character. *Scientific Reports* 12, 1 (2022), 21329.

[69] Jean-Luc Lugrin, Maximilian Ertl, Philipp Krop, Richard Klüpfel, Sebastian Stierstorfer, Bianka Weisz, Maximilian Rück, Johann Schmitt, Nina Schmidt, and Marc Erich Latoschik. 2018. Any "body" there? avatar visibility effects in a virtual reality game. In *IEEE Conference on Virtual Reality and 3D User Interfaces.* IEEE, 17–24.

[70] Jessie Mann. 2020. The medical avatar and its role in neurorehabilitation and neuroplasticity: a review. *NeuroRehabilitation* 46, 4 (2020), 467–482.

[71] Zahira Merchant, Ernest T Goetz, Lauren Cifuentes, Wendy Keeney-Kennicutt, and Trina J Davis. 2014. Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & education* 70 (2014), 29–40.

[72] Akira Miyake, Naomi P Friedman, Michael J Emerson, Alexander H Witzki, Amy Howerter, and Tor D Wager. 2000. The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive psychology* 41, 1 (2000), 49–100.

[73] 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.

[74] Kizashi Nakano and Takuji Narumi. 2025. Avatars, Should We Look at Them Directly or through a Mirror?: Effects of Avatar Display Method on Sense of Embodiment and Gaze. *IEEE Transactions on Visualization and Computer Graphics* (2025).

[75] Sharlene D Newman, Mitchell T Hansen, and Arianna Gutierrez. 2016. An fMRI study of the impact of block building and board games on spatial ability. *Frontiers in psychology* 7 (2016), 1278.

[76] Masayasu Niwa, Katsutoshi Masai, Shigeo Yoshida, and Maki Sugimoto. 2023. Investigating effects of facial self-similarity levels on the impression of virtual agents in serious/non-serious contexts. In *Augmented Humans International Conference.* 221–230.

[77] Georg Northoff, Alexander Heinzel, Moritz De Greck, Felix Bermpohl, Henrik Dobrowolny, and Jaak Panksepp. 2006. Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage* 31, 1 (2006), 440–457.

[78] Hawkar Oagaz, Breawn Schoun, and Min-Hyung Choi. 2021. Performance improvement and skill transfer in table tennis through training in virtual reality. *IEEE Transactions on visualization and computer graphics* 28, 12 (2021), 4332–4343.

[79] Diriba Dereje Olana, Teklu Gemechu Abessa, Dheeraj Lamba, Lisa Tedesco Triccas, and Bruno Bonnechere. 2025. Effect of virtual reality-based upper limb training on activity of daily living and quality of life among stroke survivors: a systematic review and meta-analysis. *Journal of NeuroEngineering and Rehabilitation* 22, 1 (2025), 92.

[80] Sofia Adelaide Osimo, Rodrigo Pizarro, Bernhard Spanlang, and Mel Slater. 2015. Conversations between self and self as Sigmund Freud—A virtual body ownership paradigm for self counselling. *Scientific reports* 5, 1 (2015), 13899.

[81] Olaoluwa Oyedokun, Syed Tanzim Mubarrat, Amogh Joshi, Christos Mousas, and Dominic Kao. 2024. Exploring the influence of avatar skin tone in VR educational games. In *Companion Proceedings of the 2024 Annual Symposium on Computer-Human Interaction in Play.* 227–234.

[82] Stefan Pastel, Katharina Petri, Chien-Hsi Chen, Ana Milena Wiegand Cáceres, Meike Stirnatis, Carlo Nübel, Lasse Schlotter, and Kerstin Witte. 2023. Training in virtual reality enables learning of a complex sports movement. *Virtual Reality* 27, 2 (2023), 523–540.

[83] Francesco Pavani, Charles Spence, and Jon Driver. 2000. Visual capture of touch: Out-of-the-body experiences with rubber gloves. *Psychological science* 11, 5 (2000), 353–359.

[84] Tabitha C Peck, Jessica J Good, Austin Erickson, Isaac Bynum, and Gerd Bruder. 2022. Effects of transparency on perceived humanness: Implications for rendering skin tones using optical see-through displays. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (2022), 2179–2189.

[85] Quan-cheng Peng, Ling Yin, and Yi Cao. 2021. Effectiveness of virtual reality in the rehabilitation of motor function of patients with subacute stroke: a meta-analysis. *Frontiers in neurology* 12 (2021), 639535.

[86] Daniel Perez-Marcos, Mélanie Bieler-Aeschlimann, and Andrea Serino. 2018. Virtual reality as a vehicle to empower motor-cognitive neurorehabilitation. *Frontiers in psychology* 9 (2018), 2120.

[87] Beatrix Potter. 1990. *The tale of Peter rabbit.* Baker's Plays.

[88] Unnikrishnan Radhakrishnan, Lisheng Kuang, Konstantinos Koumaditis, Francesco Chinello, and Claudio Pacchierotti. 2023. Haptic feedback, performance and arousal: A comparison study in an immersive vr motor skill training task. *IEEE transactions on haptics* 17, 2 (2023), 249–262.

[89] Rabindra Ratan, David Beyea, Benjamin J Li, and Luis Graciano. 2020. Avatar characteristics induce users' behavioral conformity with small-to-medium effect sizes: a meta-analysis of the proteus effect. *Media Psychology* 23, 5 (2020), 651–675.

[90] Grégoire Richard, Thomas Pietrzak, Ferran Argelaguet, Anatole Lécuyer, and Géry Casiez. 2021. Studying the role of haptic feedback on virtual embodiment in a drawing task. *Frontiers in Virtual Reality* 1 (2021), 573167.

[91] Anca Salagean, Eleanor Crellin, Martin Parsons, Darren Cosker, and Danaë Stanton Fraser. 2023. Meeting your virtual twin: Effects of photorealism and personalization on embodiment, self-identification and perception of self-avatars in virtual reality. In *ACM CHI conference on human factors in computing systems*. 1–16.

[92] Roy Salomon, Melanie Lim, Oliver Kannape, Joan Llobera, and Olaf Blanke. 2013. "Self pop-out": agency enhances self-recognition in visual search. *Experimental brain research* 228 (2013), 173–181.

[93] Anri Sasaki, Eizaburo Suzuki, Kotaro Homma, Nariyuki Mura, and Katsuhiko Suzuki. 2025. Impact of Observation Duration in Action Observation Therapy: Manual Dexterity, Mirror Neuron System Activity, and Subjective Psychomotor Effort in Healthy Adults. *Brain Sciences* 15, 5 (2025), 457.

[94] Stoo Sepp, Steven J Howard, Sharon Tindall-Ford, Shirley Agostinho, and Fred Paas. 2019. Cognitive load theory and human movement: Towards an integrated model of working memory. *Educational Psychology Review* 31 (2019), 293–317.

[95] Meng Ting Shih, Yi-Chieh Lee, Chih-Mao Huang, and Liwei Chan. 2023. " A feeling of déjà vu": The Effects of Avatar Appearance-Similarity on Persuasiveness in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–31.

[96] Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. 2013. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review* 20, 1 (2013), 21–53.

[97] Mel Slater, Martin Usoh, and Anthony Steed. 1994. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments* 3, 2 (1994), 130–144.

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

[99] Anthony Steed, Sebastian Frlston, Maria Murcia Lopez, Jason Drummond, Ye Pan, and David Swapp. 2016. An 'in the wild'experiment on presence and embodiment using consumer virtual reality equipment. *IEEE transactions on visualization and computer graphics* 22, 4 (2016), 1406–1414.

[100] Guansen Tong, Johnathan Chi-Ho Leung, Xi Peng, Haosheng Shi, Liujie Zheng, Shengze Wang, Arryn Carlos O'Brien, Ashley Paula-Ann Neall, Grace Fei, Martim Gaspar, et al. 2025. Multimodal Neural Acoustic Fields for Immersive Mixed Reality. *IEEE Transactions on Visualization and Computer Graphics* (2025).

[101] Manos Tsakiris and Patrick Haggard. 2005. The rubber hand illusion revisited: visuotactile integration and self-attribution. *Journal of experimental psychology: Human perception and performance* 31, 1 (2005), 80.

[102] Manos Tsakiris, Matthew R Longo, and Patrick Haggard. 2010. Having a body versus moving your body: neural signatures of agency and body-ownership. *Neuropsychologia* 48, 9 (2010), 2740–2749.

[103] Manos Tsakiris, Gita Prabhu, and Patrick Haggard. 2006. Having a body versus moving your body: How agency structures body-ownership. *Consciousness and cognition* 15, 2 (2006), 423–432.

[104] Frank MF Verberne, Jaap Ham, and Cees JH Midden. 2015. Trusting a virtual driver that looks, acts, and thinks like you. *Human factors* 57, 5 (2015), 895–909.

[105] Kun Wang, Meng Chen, Li Lu, Jingwen Feng, Qianniu Chen, Zhongjie Ba, Kui Ren, and Chun Chen. 2025. From one stolen utterance: Assessing the risks of voice cloning in the aigc era. In *2025 IEEE Symposium on Security and Privacy (SP)*. IEEE, 4663–4681.

[106] Liu Wang, Mengjie Huang, Rui Yang, Hai-Ning Liang, Ji Han, and Ying Sun. 2022. Survey of movement reproduction in immersive virtual rehabilitation. *IEEE Transactions on Visualization and Computer Graphics* 29, 4 (2022), 2184–2202.

[107] William D Weisman and Jorge F Peña. 2021. Face the uncanny: the effects of doppelganger talking head avatars on affect-based trust toward artificial intelligence technology are mediated by uncanny valley perceptions. *Cyberpsychology, behavior, and social networking* 24, 3 (2021), 182–187.

[108] Robert B Welch. 1972. The effect of experienced limb identity upon adaptation to simulated displacement of the visual field. *Perception & Psychophysics* 12, 6 (1972), 453–456.

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

[110] Katja Windheim, David Veale, and Martin Anson. 2011. Mirror gazing in body dysmorphic disorder and healthy controls: Effects of duration of gazing. *Behaviour research and therapy* 49, 9 (2011), 555–564.

[111] Erik Wolf, Nina Döllinger, David Mal, Stephan Wenninger, Andrea Bartl, Mario Botsch, Marc Erich Latoschik, and Carolin Wienrich. 2022. Does distance matter? Embodiment and perception of personalized avatars in relation to the self-observation distance in virtual reality. *Frontiers in Virtual Reality* 3 (2022), 1031093.

[112] Andrea Stevenson Won and Shuo Zhou. 2024. Effects of first vs. third-person perspective and self-versus other-avatars on user movements in virtual reality. *International Journal of Human-Computer Studies* 187 (2024), 103259.

[113] Yuanjie Wu, Yu Wang, Sungchul Jung, Simon Hoermann, and Robert W Lindeman. 2021. Using a fully expressive avatar to collaborate in virtual reality: Evaluation of task performance, presence, and attraction. *Frontiers in Virtual Reality* 2 (2021), 641296.

[114] Vasco Xu, Chenfeng Gao, Henry Hoffmann, and Karan Ahuja. 2024. MobilePoser: Real-Time Full-Body Pose Estimation and 3D Human Translation from IMUs in Mobile Consumer Devices. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–11.

[115] 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).

[116] Haoran Yun, Jose Luis Ponton, Carlos Andujar, and Nuria Pelechano. 2023. Animation fidelity in self-avatars: impact on user performance and sense of agency. In *IEEE Conference Virtual Reality and 3D User Interfaces*. IEEE, 286–296.