International
Communication
Association

OXFORD

People, places, and time: a large-scale, longitudinal study of transformed avatars and environmental context in group interaction in the metaverse

Eugy Han 🔾 [1,*], Mark R. Miller[2], Cyan DeVeaux[1], Hanseul Jun[1], Kristine L. Nowak 🔾 [3], Jeffrey T. Hancock[1], Nilam Ram[1,4] and Jeremy N. Bailenson[1]

[1]Department of Communication, Stanford University, California, USA
[2]Department of Computer Science, Stanford University, California, USA
[3]Department of Communication, University of Connecticut, Connecticut, USA
[4]Department of Psychology, Stanford University, California, USA

*Corresponding author: Eugy Han, 450 Jane Stanford Way. Building 120. Stanford, CA 94305, USA. Email: eugyoung@stanford.edu

Abstract

As the metaverse expands, understanding how people use virtual reality to learn and connect is increasingly important. We used the Transformed Social Interaction paradigm (Bailenson et al., 2004) to examine different avatar identities and environments over time. In Study 1 ($n = 81$), entitativity, presence, enjoyment, and realism increased over 8 weeks. Avatars that resembled participants increased synchrony, similarities in moment-to-moment nonverbal behaviors between participants. Moreover, self-avatars increased self-presence and realism, but decreased enjoyment, compared to uniform avatars. In Study 2 ($n = 137$), participants cycled through 192 unique virtual environments. As visible space increased, so did nonverbal synchrony, perceived restorativeness, entitativity, pleasure, arousal, self- and spatial presence, enjoyment, and realism. Outdoor environments increased perceived restorativeness and enjoyment more than indoor environments. Self-presence and realism increased over time in both studies. We discuss implications of avatar appearance and environmental context on social behavior in classroom contexts over time.

Lay Summary

Understanding how people connect socially via avatars in immersive virtual reality has become increasingly important given the prolific rise of the metaverse. In two large-scale, longitudinal field experiments, we extended predictions of the Transformed Social Interaction paradigm to investigate how the appearance of avatars and the characteristics of the virtual environment influenced people's behaviors and attitudes over time. In Study 1, we demonstrated the effects of time: group cohesion, presence, enjoyment, and realism measures increased over time, and the effects of appearance: When represented by avatars that looked like themselves, people displayed more synchronous nonverbal behaviors, or were more "in sync" with others, and reported the image quality of the environment and people as more realistic. On the other hand, when people wore the same uniform avatar, they experienced more enjoyment. In Study 2, we demonstrated the effects of the environment: When in more spacious virtual environments, there was more synchronous movement and people reported feeling greater restoration, group cohesion, pleasure, arousal, presence, enjoyment, and realism, than constrained environments. When in outdoor environments with elements of nature, people reported feeling greater restoration and enjoyment than indoor environments.

Keywords: time, networked, virtual reality, avatars, environments, context, nonverbal behavior

The metaverse, persistent immersive virtual worlds often viewed through virtual reality (VR) headsets, is receiving increasing attention in industry, media, and academia. What makes these virtual worlds unique is that people can easily transform their avatar's appearance or environmental context. With the touch of a button, a person can be anyone and anywhere. As the medium moves from gaming arcades and laboratories to consumers' homes and universities, it is becoming increasingly important from both a theoretical and societal well-being standpoint to understand how these Transformed Social Interactions (TSI, Bailenson et al., 2004) influence people and their relationships with others, especially over prolonged periods of time.

One way of understanding the metaverse is through literature on collaborative virtual environments (CVEs). While researchers have been studying CVEs for decades (for a recent review, see Aseeri & Interrante, 2021), several important

challenges have limited these investigations (for exceptions, see Khojasteh & Won, 2021, Moustafa & Steed, 2018, and Bailenson & Yee, 2006). First, due to the high expense and technical challenges involving VR implementation, researchers are often forced to rely on small sample sizes and either one-shot or a limited number of sessions (Lanier et al., 2019). This raises the question: As the novelty wears off over time, will people have a "better" or "worse" experience as they adapt to the medium? Second, while current metaverse platforms feature groups of various sizes, the majority of research on CVEs feature dyads or occasionally triads (for a review, see Han et al., 2022), which take on very different turn-taking, gaze behavior, and other group dynamics from larger groups.

The current research used two longitudinal field experiments to systematically examine multiple sets of larger groups and how social dynamics evolve over time in CVEs. From a statistical standpoint, we take a multivariate approach to

observe how multiple constructs change over time and how they may be interrelated. From a theoretical standpoint, we extend the TSI paradigm to CVEs. Study 1 examines the visual appearance of avatars, a construct that has received much attention in the literature overall (for a review, see Ratan et al., 2020), by investigating the effects of assigning avatars customized to look like the self or uniform avatars to mask visual identity cues. Study 2 examines an environmental context by leveraging the ability to easily create VR scenes that differ in spaciousness and setting.

Background and previous work
Transformed social interaction

In CVEs, people are represented by avatars, or visual representations of the self. CVEs track various verbal and nonverbal signals and map them onto these digital beings. CVEs can also systematically filter avatars' behavioral actions and physical appearance, amplifying or suppressing features and nonverbal signals in real-time. As predicted by TSI, the behaviors and appearance of avatars in CVEs can have a drastic impact on people's perceptions, as well as their persuasive and instructional abilities (for a review, see Bailenson et al., 2008b). Of the three categories outlined by TSI: self-representations (i.e., avatars), contextual situations (i.e., virtual environments), and sensory capabilities, this current study focuses on the first two categories.

While research on TSI has occurred over two decades, there are still gaps in how these transformations affect people. For example, a recent article by Szolin et al. (2022) systematically reviewed the literature of avatar transformation on behavioral and attitudinal changes in the context of video games. The authors underscored how previous studies fail to separate different types of virtual environments, such as commercially available videogames and bespoke research-focused virtual environments. They concluded that this field of research is still in a relatively new stage of psychological investigation, and that further empirical investigation of avatar transformation and different contexts is needed to understand how they affect people both during and after the virtual experience. In the same vein, there has been growing literature on how other transformations lead to changed behavior, such as how body manipulations produce social, perceptual, and behavioral effects (for a review, see Gonzalez-Franco & Peck, 2018). However, most studies do not examine these transformations in the context of networked, social interactions or examine how the virtual environment itself may influence behaviors and attitudes.

Transforming self-representations

Previous research shows that even subtle transformations in avatars' behavioral or visual (e.g., photographic or anthropomorphic) resemblance can impact the way people engage with and perceive others (Blascovich, 2002; Nowak & Biocca, 2003; Yee et al., 2011). For example, Roth et al. (2018) manipulated the type of gaze—natural, hybrid, synthesized, and random—exhibited by avatars in dyadic social interactions. They reported that, based on trends found in perceived virtual rapport, interpersonal attraction, and trust, natural gaze was superior, and synthetic and hybrid gaze were better than random gaze. Other research found that, overriding avatars' actual head movements to mimic motions and increase synchrony led to greater liking between interaction partners

(Bailenson & Yee, 2005). Finally, a longitudinal study by Bailenson and Yee (2006) implemented TSI conditions to examine the impact of visual and behavioral similarity on group cohesion and task performance. They found that in certain tasks, groups performed better when they saw their own face on their partners' avatars (i.e., shared visual similarity). They also reported that, even with TSI manipulations, entitativity increased over time. However, the small sample size of this study precluded the generalizability of these data.

Transforming the situation

In VR, each person can have a unique viewpoint and sensory information, as it is possible to vary the number of visible avatars, spatial arrangement, or even colors in a scene. TSI has shown that these factors alter sensory perception, social interaction, and performance (Bailenson et al., 2008b). For example, Hasenbein et al. (2022) transformed the seating position of a student in a virtual classroom and manipulated how many rows of peer learners were between themselves, the teacher, and the screen. They used eye-tracking to show subsequent changes to attention and found that different seating arrangements led to different focus of gaze transitions on their virtual peer learners, teachers, and the screen, as well as different gaze distributions. Such spatial transformations of students' seat positions have also been shown to influence memory (Bailenson et al., 2008a) and persuasion (McCall et al., 2009). Similarly, Miller et al. (2021) found that teams of designers had more positive interactions when working in a VR conference room compared to working in a VR garage. Other factors, such as the size of a virtual room and what kind of objects are placed within it, have also been shown to influence outcomes such as attention and navigation (Kim et al., 2022).

Visible space: panoramic and constrained environments

One environmental factor of interest pertains to the amount of space visible from a given viewpoint. Previous research has shown differences in outcomes resulting from being in a constrained or panoramic (i.e., environments in which people can see wide and far) environment. For example, taller (i.e., more spacious) ceilings have been shown to prime feelings of freedom and encourage a more global, abstract way of processing information (Meyers-Levy & Zhu, 2007). Similarly, compared to confining environments, spacious environments were found to foster more self-disclosure (Okken et al., 2012). Finally, compared to smaller rooms, larger rooms were found to promote more engagement in informal learning among students (Wu et al., 2017).

Setting: outdoor and indoor environments

Natural settings have been shown to have beneficial effects (Bratman et al., 2015; Lederbogen et al., 2011). Living in environments where people have access to green spaces have shown to lead to lower mental distress and greater well-being in the long term (White et al., 2013). Even shorter, simulated exposure to nature, such as viewing slides or pictures containing landscapes, have been shown to reduce stress (van den Berg et al., 2014), improve self-esteem (Barton & Pretty, 2010), and physiological restoration (Ulrich et al., 1991), and increase the ability to focus (Berto, 2005).

VR is a medium that is uniquely suited for simulating natural settings. For instance, Anderson et al. (2017) showed that 360° videos of natural environments, compared to those of indoor environments, led to greater relaxation and reduced

negative affect. VR's immersion enhances the restorative potential of mediated natural environments on physiological well-being (for a review, see Browning et al., 2020). To our knowledge, the effect of virtual nature has not been studied in the presence of other avatars within CVEs.

Group behavior

When individuals are linked by social relationships that make up a group, they become interdependent and influence each other's behaviors, emotions, and perceptions (Janis, 1973; Milgram, 1963; Sherif, 1937). The moment a collection of individuals perceives itself as a group, a construct known as *entitativity* (Campbell, 1958), a series of psychological and interpersonal changes, occurs (Forsyth, 1990; Harasty, 1996). Given the social nature of CVEs, to meaningfully understand the effects of virtual experiences in such systems, it is important to investigate not only individuals, but also individuals as members of a group.

Nonverbal dynamics also play a critical role in groups. During the course of face-to-face communication, interactants tend to be in "synchrony," or be in similar states or have similar behaviors at similar times (Condon & Ogston, 1966). VR's ability to track motion allows for the examination of *motion synchrony*. Higher motion synchrony has been shown to be related to greater rapport between teachers and students (LaFrance, 1979) and more creativity (Won et al., 2014). Previously, researchers have manipulated synchrony with avatars displayed on screens (Oh Kruzic et al., 2020) and with agents (Tarr et al., 2018) and avatars (Sun et al., 2019; Miller et al., 2021) in VR.

Time

Human processes are complex and rarely, if ever, exist in isolation. It is critical to understand the processes by which human behavior and activity emerge as different components of a system, and influence and change one another over time (see *Dynamic systems theory*, Newman & Newman, 2020). Study of individuals repeatedly exposed to media stimuli and adapting to new technologies provide a unique opportunity to gain valuable insight about both the long-term and short-term processes invoked by those media (e.g., Bailenson & Yee, 2006; Brinberg et al., 2021). Inferences based on single-session exposures or obtained through analysis of just a few sessions when participants are adjusting to the novelty of a medium can be plagued with technical difficulties and be misleading. A number of researchers have theorized that, while first-time VR users may feel unfamiliar with the medium, with use, their experience in VR should improve (e.g., Loomis, 1992). Other perspectives suggest the opposite, arguing that the habituation effect can cause what was initially novel to diminish (e.g., Lombard & Ditton, 1997).

TSI is a paradigm that should be particularly sensitive to repeated exposures over time, as people learn to identify, adapt, and accommodate to the changes. For instance, while seeing a uniform avatar on everyone in a room may be jarring at first, perhaps with time people habituate. Alternatively, the effects of similarity may amplify. An early study by Bailenson and Yee (2006) followed three triads of participants for 15 sessions over a 10-week period as they collaborated for approximately 45 min per session. In addition to looking at time, the researchers manipulated two types of TSI—behavioral and visual similarity of group members. Results demonstrated changes in task performance, subjective ratings, nonverbal behavior, and simulator sickness over time as participants became familiar with the system. Furthermore, even in the presence of TSI, where there was a mismatch between types of behavior, entitativity was high over time, suggesting that people are able to retain symbolic meaning even with the starkest degree of social cues. However, the small sample reduced power, and further research is needed to examine how people evolve and respond to these transformations of people and place.

Current study

The current work aims to investigate how transformations of *who* you are and *where* you are evolve over time. Using TSI as its central framework, this work addresses two of the categories: self-representations and contextual situations, through two large-scale, longitudinal field experiments of how VR-based transformation of avatar appearance (Study 1) and environmental context (Study 2) influence interactants in group settings. The overarching research questions are as follows:

- (RQ1) How do people's behaviors and attitudes change over time?
- (RQ2) How do people's behaviors and attitudes change when they are embodying and surrounded by different avatars?
- (RQ3) How do people's behaviors and attitudes change when they are interacting in different environmental contexts?

Given the critical role that self-presentations have on how individuals perceive their experience and their communication partners in a virtual environment, we manipulated the visual appearance of the avatars such that members of the same group were represented by either avatars that resembled their physical self or avatars that were uniform across all members. Similarly, given the critical role that where you are and what kind of environment you are surrounded by can lead to differing outcomes, we manipulated the type of virtual environment in which group members interacted. Across both studies, we measured both behavioral and self-report variables central to understanding people's experiences in virtual environments, such as presence and realism. We additionally collected measures that aim to understand group outcomes, such as synchrony and entitativity. Using linear growth models with time-invariant and time-varying covariates, we built two models to understand how these outcomes change across time and vary based on individual differences.

Both field experiments were housed in a 10-week course about VR and its intersections with various disciplines. During each course, participants were provided with a VR headset, which they used to attend eight weekly instructor-led, medium-group size discussion sessions ($n1 = 9$–14, $n2 = 5$–8). The nature of each study being housed in a course allowed for naturalistic intervention of our variables of interest and unobtrusive measurement of behaviors. Nonverbal behavior was measured by recording 18 degrees of freedom of movement from each participant (e.g., pitch, yaw, and roll of head and both hands) to compute motion synchrony for each group. Attitudes of each participant's experience were measured using weekly surveys. We additionally explored how much these outcomes may be mediated by individual differences.

Our key contributions to the field are as follows:

- Time plays a critical role in people's experience in social VR. Across both studies, self-presence and realism increased over time and with VR use. In Study 1, entitativity, social and spatial presence, enjoyment, and realism increased over time and with VR use.
- Who you are and what you look like matter in social VR. When people are represented by avatars that resemble their physical selves, they are more nonverbally "in sync" with other people, and view the virtual environment and people as more realistic. On the other hand, having a uniform avatar leads to greater enjoyment.
- Where you are matters in social VR. When people are in more spacious virtual environments, they are more nonverbally "in sync" with other people, report feeling greater restoration, entitativity, pleasure, arousal, presence, enjoyment, and realism than when they are in constraining environments. This is an especially important finding as it is difficult to study very large indoor spaces in the real world, and the current findings are novel. Similarly, when there are elements of nature and people are in outdoor environments, they report feeling greater restoration and enjoyment than when they are in indoor environments.

Study 1

Study 1 focuses on the transformation of avatar appearance and investigates the following research questions: first, how will sharing visual similarity with group members influence nonverbal synchrony and entitativity over time? Second, how will perceived self, social, and spatial presence change over time and with different avatars? Third, how will perceived enjoyment of interacting in a virtual environment change over time and with different avatars? Finally, how will perceived realism change over time and with different avatars?

Method
Participants

Participants were 101 university students enrolled in a 10-week course about VR. At the beginning of the course, students were invited to participate in an Institutional Review Board-approved (IRB) study of how repeated exposure to VR influenced their individual and group behavior. While all students who were part of the course took part in all the VR activities, only those who consented to participate in the study had their data included in the study. Of the 101 students in the course, 93 consented to participate in the study. The 81 participants who participated in five or more of the eight weekly sessions (M = 47, F = 30, Other = 2, declined to or did not respond = 2) were between 18 and 58 years old (M = 22.26, SD = 5.19; $n_{18\sim23} = 68$, $n_{24\sim29} = 7$, $n_{30\sim35} = 3$, $n_{35\sim40} = 1$, $n_{55\sim60} = 1$, declined to or did not respond = 1) and identified as Asian or Asian-American (n = 30), White (n = 21), African, African-American, or Black (n = 11), Hispanic or Latinx (n = 9), multiracial (n = 5), Middle Eastern (n = 1), and declined to or did not respond (n = 4). Participants had varying levels of experience with VR, with 48 (59%) having never used VR before. Prior to the course, 38 participants were not familiar with anyone in their discussion group, and others reported knowing one ($n_1 = 13$) or more members ($n_2 = 12$, $n_3 = 1$, $n_4 = 2$, $n_5 = 2$). Safeguards implemented to ensure privacy and consent included review both by the IRB and a second university ethics organization, and third-party oversight of the consent process and data collection.

Hardware and VR equipment

Participants were provided with Oculus Quest 2 headsets (standalone head-mounted display with 1832 × 1920 resolution per eye, 104.00° horizontal FOV, 98.00° FOV, 90Hz refresh rate, and six-degree-of-freedom inside-out head and hand tracking, 503 g) and two hand controllers (126 g) for use in their personal environment. Of the 81 participants, 2 owned personal headsets (PC-based Valve Index) and participated using those devices.

Virtual environment: ENGAGE

Weekly sessions were hosted in ENGAGE, a collaborative social VR platform designed for education. Every week, the virtual environment consisted of a private (password restricted) "Engineering Workshop" room that was a large, open-space area that allowed participants to walk/teleport freely, create 3D drawings, write on personal whiteboards/stickies, add immersive effects/3D objects, and display media content. The large space accommodated the use of 3D audio, which allowed for splitting off into smaller groups without audio overlap.

Avatar: self vs. uniform

In ENGAGE, participants are represented by human avatars (Figure 1). Participants embodied one of two possible representations in the *avatar conditions*. In the *self-avatar* condition, participants were able to customize their avatars with various combinations of outfits, gender, age, skin complexion, weight, hairstyles, and facial features. In the *uniform avatar* condition, all participants used a pre-selected avatar within the customization options possible within ENGAGE. Through pre-testing and iteration, we chose an avatar that was gender and racially ambiguous. Prior to the study, we conducted a survey showing screenshots of five different avatars that varied in gender, skin tone, and facial features, and asked participants (n = 27) on their perceptions of each avatar's gender presentation, racial category, and whether they felt comfortable being visually represented by the avatar in a virtual environment. We created the uniform avatar based on features that resulted in the highest perceived neutrality in gender, racial, and comfort in representation. A detailed description of the items, results, and sample avatars from the pre-test can be found in Appendix A. The final uniform avatar had no hair (to avoid racial marking tendencies, MacLin & Malpass, 2001) and had a neutral (given the available option) skin color (to be racially ambiguous).[1]

Procedure

Participants selected a discussion group that fit their schedule and availability, resulting in eight consistent groups that met weekly for 8 weeks and varied in size from 9 to 14 members (M = 12.63, SD = 1.77). Two training sessions were held in the first 2 weeks of the course, during which participants were taught how to use the ENGAGE interface and navigate the virtual environment. During these training sessions, the teaching staff was available to assist in real-time both via video conferencing and within the virtual environment when participants faced technical mishaps. There was also a simultaneous Zoom call open during all discussion sessions, where participants could pull off their headsets and ask for technical support (Figure 2).

Figure 1. The uniform avatar, an example female customized self-avatar, and an example male customized self-avatar that participants embodied for some of their weekly discussion sessions.

Figure 2. A Zoom window with a subset of participants in their HMDs. A Zoom technical support call was open during all in-VR activities. Faces are blocked for the sake of privacy.

The structure of weekly activities varied with course content (Table 1). The sessions involved discussion with either the whole group or smaller groups of three to four, and typically followed a three-question format where participants were asked what they liked, what they were concerned about, and how what they learned might influence the future.[2] The

discussions allowed for preservation of nonverbal spatial constraints like interpersonal distance, head orientation, and spatialized sound (Figure 3, bottom right). Some sessions leveraged physical activity affordances of ENGAGE, including working together on a shared object (Figure 3, top left), creating new computer graphic content together (Figure 3,

Table 1. Study 1 weekly topics and activities

Week	Activity
1	• Acclimate participants to the headset and platform, leaving margin for technical or content issues • Discussion on being inside VR and how the experience compared to that in Zoom
2	• Full-group discussion reflecting on participants' experience visiting various sites in AltspaceVR (e.g., an art exhibition, solar system) • Sketch ideas of how one might teach and present content inside VR
3	• Full-group discussion reflecting on recording and performing as avatars inside of VR
4	• Small-group discussions reflecting on various VR empathy experiences
5	• Small-group discussion on how VR is used for medical applications and well-being
6	• Small-group activity in which participants chose a unique feature of VR and brainstormed how to communicate climate change based on this feature
7	• Activity, done either individually or in small groups, on creating and playtesting a VR-based game
8	• Small-group discussion reflecting on VR and its use cases, dangers, and potential direction

top right), and design-thinking with a shared whiteboard and stickies (Figure 3, bottom left).

Each week, groups were assigned to one of the two avatar conditions (self vs. uniform) via a Latin square randomization scheme that ensured that each group spent 4 weeks in each condition, counterbalanced across weeks and meeting days (Table 2). In each session, all members of a group wore either their self or uniform avatar to attend the discussion.

Measures

Multiple aspects of individuals' behaviors and attitudes were measured at the start of the study (pre-test), and during and after each of the eight weekly sessions (motion and weekly surveys; see Table 3).

Weekly repeated measures

Individual ratings were obtained after each weekly VR session through analysis of behavioral motion time series or surveys. To reduce fatigue, repetitiveness, and burden, item sets for each construct were purposely designed to be brief (Conner & Lehman, 2012).

Nonverbal behavior: motion synchrony

Following prior work on synchrony in VR (Miller et al., 2021; Sun et al., 2019), motion synchrony was first calculated

Figure 3. Participants represented either by their customized self-avatar or a uniform avatar (top left) interacting with immersive effects/3D objects, (top right) drawing in 3D space, (bottom left) utilizing a whiteboard, and (bottom right) having a discussion during the weekly sessions. The bars floating above the avatar are blocking the participants' names for the sake of privacy.

Table 2. Participants in each group ($n1 = 8$, $n2 = 24$) were randomly assigned to an avatar condition (self vs. uniform)

(A)

	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8
Group 1	Self	Self	Uniform	Uniform	Self	Self	Uniform	Uniform
Group 2	Self	Uniform	Self	Uniform	Self	Uniform	Self	Uniform
Group 3	Uniform	Self	Uniform	Self	Uniform	Self	Uniform	Self
Group 4	Uniform	Uniform	Self	Self	Uniform	Uniform	Self	Self
Group 5	Self	Self	Uniform	Uniform	Self	Self	Uniform	Uniform
Group 6	Self	Uniform	Self	Uniform	Self	Uniform	Self	Uniform
Group 7	Uniform	Self	Uniform	Self	Uniform	Self	Uniform	Self
Group 8	Uniform	Uniform	Self	Self	Uniform	Uniform	Self	Self

Note. This design ensured that each group experienced each condition once and that each condition appeared equally across the weekly schedule.

Table 3. Study 1 means and standard deviations (in parentheses) of repeated measures across 8 weeks

DV	Avatar	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Total
Synchrony	Self	0.024 (0.017)	0.049 (0.044)	0.028 (0.021)	0.044 (0.034)	0.025 (0.030)	0.035 (0.034)	0.013 (0.018)	0.028 (0.023)	0.032 (0.032)
	Uniform	0.031 (0.025)	0.023 (0.028)	0.015 (0.017)	0.030 (0.046)	0.022 (0.015)	0.0201 (0.025)	0.013 (0.023)	−0.0069 (0.034)	0.019 (0.029)
Entitativity	Self	5.18 (1.10)	5.07 (0.94)	5.18 (0.96)	5.05 (0.98)	5.63 (0.72)	5.33 (0.99)	5.33 (0.89)	5.28 (0.86)	5.25 (0.94)
	Uniform	4.82 (0.92)	5.13 (0.90)	5.17 (0.93)	5.44 (0.87)	5.12 (0.98)	5.07 (0.99)	5.62 (0.85)	5.63 (0.76)	5.25 (0.93)
Self-presence	Self	3.77 (1.29)	3.58 (1.70)	4.16 (1.38)	4.00 (1.57)	4.55 (1.53)	4.42 (1.59)	4.58 (1.41)	4.34 (1.65)	4.17 (1.55)
	Uniform	3.44 (1.37)	3.78 (1.46)	3.98 (1.57)	4.15 (1.59)	4.07 (1.84)	4.07 (1.43)	4.50 (1.85)	4.28 (1.66)	4.04 (1.62)
Social presence	Self	5.60 (0.96)	5.41 (1.24)	5.18 (0.99)	5.67 (1.04)	5.88 (0.75)	5.67 (1.04)	5.64 (0.97)	5.49 (1.13)	5.57 (1.03)
	Uniform	5.24 (1.34)	5.54 (0.94)	5.41 (1.32)	5.76 (0.86)	5.50 (0.95)	5.54 (1.02)	5.92 (1.06)	6.03 (0.67)	5.62 (1.05)
Spatial presence	Self	5.11 (1.07)	4.04 (1.65)	4.61 (1.25)	4.63 (1.34)	4.83 (1.27)	4.95 (1.40)	5.03 (1.43)	4.80 (1.41)	4.75 (1.39)
	Uniform	4.37 (1.56)	4.88 (1.26)	4.54 (1.47)	4.88 (1.19)	4.59 (1.44)	5.01 (1.23)	5.15 (1.37)	5.42 (1.21)	4.86 (1.36)
Enjoyment	Self	3.67 (0.85)	3.08 (1.08)	3.10 (0.88)	3.11 (1.02)	3.33 (0.83)	3.45 (0.98)	3.75 (0.97)	3.42 (1.03)	3.36 (0.98)
	Uniform	3.56 (1.01)	3.53 (0.75)	3.27 (1.01)	3.38 (0.89)	3.31 (0.90)	3.24 (1.06)	3.97 (0.92)	3.69 (0.96)	3.50 (0.96)
Realism	Self	37.00 (20.02)	35.81 (21.01)	32.90 (17.96)	37.40 (22.88)	45.82 (23.12)	43.68 (21.24)	42.31 (23.99)	41.76 (22.73)	39.65 (21.82)
	Uniform	29.62 (19.54)	34.28 (20.86)	36.82 (18.91)	38.21 (20.32)	36.31 (21.87)	39.15 (21.51)	45.70 (24.12)	40.78 (19.13)	37.66 (21.16)

for each pair of individuals in a session. Specifically, we calculated the Spearman correlation between all measurements of two individuals' avatar head speeds obtained every one-thirtieth of a second during the 30-min session (30 Hz) for all offsets of ± 2.5 s (not including offset $= 0$). These 150 correlations were then averaged to obtain a synchrony measure for each pair of participants for each week (pre-registration at https://osf.io/3c4aj/). Synchrony for a given individual on a given week was then calculated as the average of the synchrony scores for all pairs from a given session that included that individual. A detailed description of how motion synchrony was calculated can be found in Appendix B.

Entitativity

Entitativity was measured by seven items adapted from Rydell and McConnell (2005) using a 7-point Likert scale ($1 =$ Strongly disagree, $7 =$ Strongly agree). Sample items

include "My discussion group is important to its members" and "Members of my discussion group are affected by the behaviors of other members." Weekly entitativity scores were calculated as the mean of the seven items (Cronbach's $\alpha = 0.9$), with higher scores indicating greater entitativity.

Self, social, and spatial presence

Self, social, and spatial presence were measured by items adapted from prior work (Herrera et al., 2020; Oh et al., 2019) using a 7-point Likert scale ($1 =$ Strongly disagree to $7 =$ Strongly agree). *Self-presence* was measured as the level of agreement with two items: "I felt like my avatar's body was my own body," and "When something happened to my avatar, I felt like it was happening to me." *Social presence* was measured as the level of agreement with two items, "I felt like I was in the same room as my classmates," and "I felt like my classmates were aware of my presence." *Spatial presence* was

measured as the level of agreement with the two items: "I felt like I was really there inside the virtual environment" and "I felt as if I could reach out and touch the objects or people in the virtual environment." Weekly scores for each of the three types of presence were calculated as the mean of the two items, with higher scores indicating greater perceived presence. Internal consistencies (calculated using Spearman–Brown formula, as recommended for two-item measures, Eisinga et al., 2013), across all participants and weeks were 0.86 for self-presence, 0.80 for social presence, and 0.85 for spatial presence.

Enjoyment

Enjoyment was measured as the level of agreement with two items: "How much did you like interacting in the virtual environment?" and "How much fun did you have in the virtual environment?" using a 5-point Likert scale (1 = Not at all, 5 = Extremely). Weekly scores for enjoyment were calculated as the mean of the two items (Spearman–Brown coefficient = 0.91), with higher scores indicating greater enjoyment in the virtual environment.

Realism

Perceived photorealism of the virtual environment and people, which refers to the rendering quality of the image, was measured weekly by a single item adapted from Nowak et al. (2009) using a slider scale (0 = Cartoon-like, 100 = Photorealistic). We used a one-item scale that focuses on one of the dimensions of realism, photorealism. There are multiple dimensions of realism that have distinct effects on people's perceptions of the mediated environment and characters. In this study, the most critical dimension was the level of realism of the avatar that dealt not with whether it was a fantasy character or could occur offline, but instead the quality of the imagery. Because the original scale included other items that may relate to other dimensions, those items were excluded.

Individual differences measures

Individual differences measures were obtained during the pre-test and through analysis of motion data obtained throughout the entire study period.

Prior relationships

The number of discussion group members individuals were familiar with prior to the course was measured at the start of the study (e.g., 0, 1, 2, 3 people), to evaluate if there was an influence of having prior familiarity with any group members on how the dependent variables evolve over time ($M = 0.98$, $SD = 1.24$).

Prior VR use

Individuals' prior experience with VR was measured at the start of the study. Individuals were asked if they had ever used a VR headset before (1 = Yes, 0 = No), and if they had, how many times they had experienced VR ($n_0 = 41$, $n_1 = 6$, $n_2 = 6$, $n_3 = 7$, $n_{3+} = 20$, declined to or did not respond = 1).

Group identification

Individual ratings for group identification, a person's identification to a group they belong to, such as an organization, club, or sports team, were measured at the start of the study using eight items adapted from an in-group identification

scale and an organizational identification scale (Leach et al., 2008; Mael & Ashforth, 1992). Sample items, each answered using a 7-point Likert scale (1 = Strongly disagree, 7 = Strongly agree), included: "The fact that I am part of my group is an important part of my identity" and "When I talk about my group, I usually say 'we' rather than 'they.'" Individual group identification scores were calculated as the mean of the eight items (Cronbach's $\alpha = 0.89$), with higher scores indicating greater identification with the group ($M = 5.35$, $SD = 0.97$).

Other individual differences

Additional individual differences predictors were examined in our preliminary models, including gender, computer and online learning self-efficacy, loneliness, Zoom fatigue, and video game usage, but were eventually trimmed from the reporting because none of these variables were related to baseline levels, rates of change, or avatar effects for any of the seven outcomes.

Data analysis

Individual differences in how individuals' behaviors and attitudes changed across the 8 weeks and in relation to the type of avatar (self vs. uniform), and how these effects were related to individual differences in prior relationships, prior VR experience, and group identification were examined using linear growth models with time-invariant and time-varying covariates (Grimm et al., 2016). Small between-group variance suggested use of a two-level structure with the repeated measures nested within individuals. Specifically, each of the seven weekly repeated measures outcomes were modeled as

$$outcome_{ti} = \beta_{0i} + \beta_{1i}(week_{ti}) + \beta_{2i}(avatar_{ti}) + e_{ti} \quad (1)$$

where the outcome of interest for person i at occasion t, $outcome_{ti}$ (e.g., social presence) is modeled as a function of a person-specific intercept, β_{0i}, a person-specific linear slope, β_{1i}, that indicates rate of change over time, a person-specific avatar effect, β_{2i}, that indicates the difference between avatar conditions, and residual error, e_{ti} that is assumed normally distributed with standard deviation σ_e. The person-specific intercepts, linear slopes, and avatar condition effects are simultaneously modeled as

$$\beta_{0i} = \gamma_{00} + \gamma_{01}(priorRelationships_i) + \gamma_{02}(priorVR_i) + \gamma_{03}(groupIdentification_i) + u_{0i} \quad (2)$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11}(priorRelationships_i) + \gamma_{12}(priorVR_i) + u_{1i} \quad (3)$$

$$\beta_{2i} = \gamma_{20} + u_{1i} \quad (4)$$

where γ_{00} and γ_{01} describe the linear trajectory of change for the prototypical individual, γ_{20} describes the prototypical effect of the uniform avatar manipulation; γ_{01}, γ_{02}, and γ_{03}, indicate how prior relationships, prior VR experience, and group identification, respectively, are related to individual differences in the initial level; γ_{11} and γ_{12} indicate how prior relationships and prior VR experience are related to individual differences in rate of change; and u_{0i}, u_{1i}, and u_{2i} are residual

unexplained differences that are assumed multivariate normal distributed with standard deviations σ_{u0}, σ_{u1}, σ_{u2}, and correlations r_{u0u1}, r_{u0u2}, and r_{u1u2}. All models were fit to the data in R using lme4 (Bates et al., 2015) with restricted maximum likelihood estimation, incomplete data treated as missing at random, and statistical significance evaluated at alpha = 0.05. Preliminary models allowed for moderation of the avatar effect, but the week × avatar interaction was not significant in any of the seven models and so was removed. In a few cases where the data did not support estimation of all random effects, the u_{2i} term was removed. After the main models were run, a variety of follow-up models were used to check sensitivity and robustness of results. These included an examination of the random effects structure through expansion of the residual error terms so that they could be time-specific (i.e., removing the homogeneity of error assumption) and sensitivity to potential outlier observations. In all cases, the pattern of results remained intact. Thus, results from the more parsimonious models are reported.

Results

Results from growth models with time-varying predictors [week and avatar (uniform = 1 vs. self = 0)] and time-invariant predictors (prior relationships, prior VR experience, and group identification) are presented separately for all seven outcomes (synchrony, entitativity, self, social, and spatial presence, enjoyment, and realism). Plots of the raw data overlaid with relevant prototypical trajectories are given in Figure 5.

Synchrony

The prototypical participant's synchrony decreased from an initial value of $\gamma_{00} = 0.0381$, $p = .006$ (on the −1 to 1 correlation scale) at a rate of $\gamma_{10} = -0.0034$, $p < .001$, per week. There was a significant effect of avatar manipulation on synchrony, such that participants synchronized less, $\gamma_{20} = -0.0122$, $p < .001$, in sessions with uniform avatars than in sessions with self-avatars. There was no evidence that individual differences in prior relationships, prior VR experience, or group identification were uniquely related to baseline levels of synchrony ($ps > 0.21$), or rates of change in synchrony ($ps > 0.14$). Figure 4 indicates the relationship of synchrony to time offset.

Entitativity

The prototypical participant's entitativity increased from an initial value of $\gamma_{00} = 5.010$, $p < .001$ (on a 7-point scale) at a rate of $\gamma_{10} = 0.059$, $p = .002$ points per week. There was no evidence that the avatar manipulation influenced entitativity, $\gamma_{20} = -0.022$, $p = .62$. A prototypical trajectory showing how entitativity changed over time is in Panel A of Figure 5. Individuals with more prior relationships had higher baseline levels of entitativity, $\gamma_{01} = 0.22$, $p = .04$, as evident in the contrast between the blue solid (+1SD on prior relationships) and dashed (−1 SD on prior relationships) lines in Panel A of Figure 5. There was no evidence that individual differences in group identification or prior VR experience were uniquely related to baseline levels of entitativity ($ps > 0.07$), or that individual differences in prior relationships or prior VR experience were uniquely related to the rate of increase in entitativity ($ps > 0.59$).
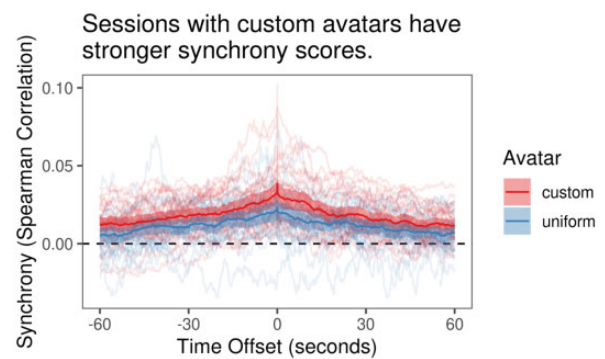

Figure 4. Effect of avatar on synchrony. This plot demonstrates that as the time offset of motion signals shifts away from zero (i.e., as one looks toward the right and left away from the center), synchrony (Y-axis) decreases. In this plot, synchrony for each group in each session is traced as a separate partially transparent line (60 total). The average of all sessions for a given avatar condition is the darker line, with the ribbon indicating 95% confidence intervals based on the underlying distribution. Each line is produced as the average of all unordered pairs in that session (from 6 to 78, $M = 36.8$, $SD = 17.9$), which is itself calculated from about 30 min of data per participant.

Presence

Self-presence

The prototypical participant's self-presence increased from an initial value of $\gamma_{00} = 3.75$, $p < .001$ (on a 7-point scale) at a rate of $\gamma_{10} = 0.101$, $p = .004$ points per week. There was a significant effect of the avatar manipulation, such that individuals reported lower self-presence when using uniform avatars than self-avatars, $\gamma_{20} = -0.21$, $p = .021$.

Social presence

The prototypical participant's social presence increased from an initial value of $\gamma_{00} = 5.23$, $p < .001$ (on a 7-point scale) at a rate of $\gamma_{10} = 0.068$, $p = .014$ points per week. There was no evidence that the avatar manipulation influenced social presence, $\gamma_{20} = 0.055$, $p = .40$. Individuals with higher group identification had higher baseline levels of social presence, $\gamma_{03} = 0.25$, $p = .021$, as evident in the contrast between the green solid (+1 SD on group identification) and dashed (−1 SD on group identification) lines in Panel C of Figure 5. There was no evidence that individual differences in prior relationships or prior VR experience were uniquely related to baseline levels of social presence ($ps > 0.30$).

Spatial presence

The prototypical participant's spatial presence increased from an initial value of $\gamma_{00} = 4.33$, $p < .001$ (on a 7-point scale) at a rate of $\gamma_{10} = 0.083$, $p = .014$ points per week. There was no evidence that the avatar manipulation influenced spatial presence $\gamma_{20} = 0.069$, $p = .38$.

There was no evidence that individual differences in group identification, prior relationships, or prior VR experience were uniquely related to baseline levels of self ($ps > 0.35$) or spatial ($ps > 0.106$) presence, or rates of increase in self ($ps > 0.63$), social ($ps > 0.49$), or spatial ($ps > 0.43$) presence.

Prototypical trajectories showing how self-presence changed over time for hypothetical individuals who alternated weekly between the two avatar conditions are shown as bold black lines in Panel B of Figure 5. Prototypical trajectories showing how social and spatial presence changed over time are in Panel C and D, respectively, of Figure 5.
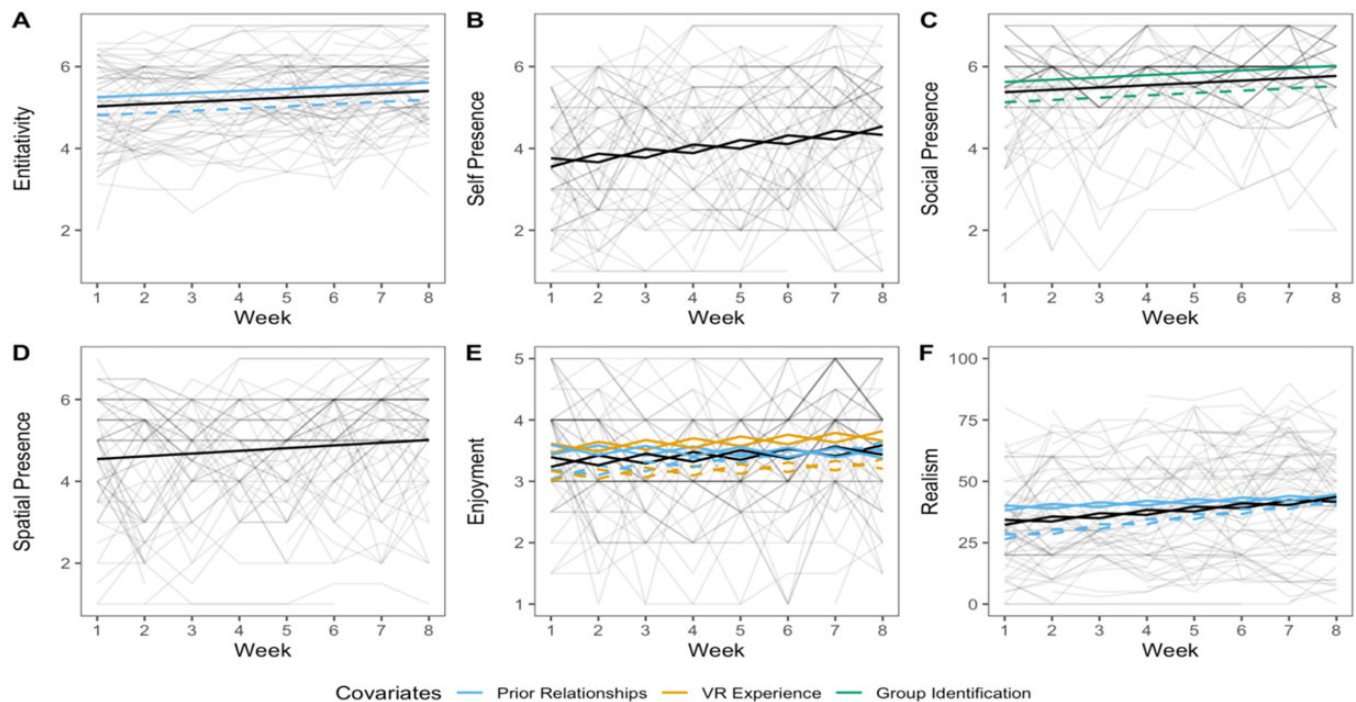
Figure 5. Dependent variables over time. (**A–F**) Change over time and in relation to the avatar manipulation for each of the six survey outcome variables. Individual trajectories (raw data) are indicated by the light gray lines. Model-implied prototypical trajectories are indicated by the thick black lines and are shown for two hypothetical cases where the avatar conditions alternated weekly (and thus produce oscillations). When individual differences were related to baseline or rate of change, additional model implied trajectories for individuals 1 SD above (solid color) and 1 SD below (dashed color) the average score are indicated by thick colored lines.

Enjoyment

The prototypical participant's enjoyment increased from an initial value of $\gamma_{00} = 3.057$, $p < .001$ (on a 5-point scale) at a rate of $\gamma_{10} = 0.061$, $p = .002$ points per week. There was a significant effect of the avatar manipulation, such that individuals reported greater enjoyment during weeks when using uniform avatars than self-avatars, $\gamma_{20} = 0.16$, $p = .011$. Prototypical trajectories showing how enjoyment changed over time for individuals who alternated weekly between the two avatar conditions are shown as bold black lines in Panel E of Figure 5. Individuals with more prior relationships had higher baseline levels of enjoyment, $\gamma_{01} = 0.22$, $p = .023$, and more prior VR experience had higher baseline levels of enjoyment, $\gamma_{02} = 0.24$, $p = .035$, as evident in the contrast between the colored lines (blue prior relationships, yellow prior VR experience; +1 SD solid, −1 SD dashed) in Panel E of Figure 5. There was no evidence that individual differences in group identification were uniquely related to baseline levels of enjoyment ($p = .37$). Although there was no evidence that individual differences in prior relationships were uniquely related to rate of increase in enjoyment ($p = .96$), the enjoyment of individuals with more prior VR experience did not increase as much as those with no prior VR experience, $\gamma_{12} = -0.044$, $p = .0079$, as seen in differential rates of increase of the yellow solid and dashed lines.

Realism

The prototypical participant's perception of realism increased from an initial value of $\gamma_{00} = 35.62$, $p < .001$ (on a 0–100, cartoon-like to photorealistic scale) at a rate of $\gamma_{10} = 0.88$, $p = .057$ points per week. There was a significant effect of the avatar manipulation, such that individuals reported lower realism (i.e., more "cartoon-like") when using uniform avatars than self-avatars, $\gamma_{20} = -2.028$, $p = .035$. Prototypical trajectories showing how realism changed over time for individuals who alternated weekly between the two avatar conditions are shown as bold black lines in Panel F of Figure 5. Individuals with more prior relationships had higher baseline levels of realism, $\gamma_{01} = 5.89$, $p = .0106$, as evident in the contrast between the blue solid (+1 SD on prior relationship) and dashed (−1 SD on prior relationship) lines in Panel F of Figure 5. There was no evidence that individual differences in group identification or prior VR experience were uniquely related to baseline levels of realism ($ps > 0.41$), or rate of change in realism ($p = .108$). The realism of individuals with more prior relationships increased less than that of individuals with fewer prior relationships, $\gamma_{11} = -0.704$, $p = .0405$, as evident in the contrast between the slopes of the blue solid and dashed lines in Panel F of Figure 5.

Discussion

Study 1 examined the role of time and transformed visual appearance on participants' experience and group dynamics. Every week for 8 weeks, 81 participants, separated into eight groups, met for approximately 30 min in a CVE to engage in a discussion on the course material. Overall, the results showed that almost all measures, including entitativity, presence (self, social, and spatial), enjoyment, and realism increased over time. The remaining measure, synchrony, decreased over time. These effects underscore the critical role that time plays in how people's experience in VR evolves. Given this, it is possible that once participants adapt to the medium and are no longer uncomfortable with the novelty of

the technology, they can reap the advantages that VR and CVEs provide and feel more presence and connectedness.

The investigation of synchrony demonstrated that motion synchrony occurs even when mediated in VR, consistent with previous research. It also indicated that synchrony both decreased over time and was lower in the uniform avatar condition (i.e., visually similar to one another) compared to the self-avatar condition. This may mean that synchrony serves a balancing function, where synchrony acts as a tool to increase entitativity when needed (Dale et al., 2020). Indeed, it is possible that transforming nonverbal behavior to induce synchrony can improve entitativity (Bailenson & Yee, 2005). However, future work should examine these possibilities.

Furthermore, when participants were in the uniform avatar, they had lower motion synchrony, reported lower self-presence, and perceived the virtual environment and others as more cartoon-like (less photorealistic), but reported greater enjoyment interacting in the virtual environment. Furthermore, while entitativity did increase over time, visual uniformity did not have an effect on entitativity. Similarly, while those who had prior relationships with group members did start with a higher level of entitativity, there was no evidence that this individual difference was uniquely related to the increase in entitativity.

Returning to the TSI paradigm, having limited cues about others' offline bodies in a virtual environment may make differences among group members more salient and interfere with the group identification process, though this does not hold true over time. It is possible that sharing identical visual features with everyone in the group creates a more recreative environment (i.e., leading to lowered photorealism), which may place less stress on how an individual is presented in front of others and less emphasis on individual behavior, ultimately leading to a lowered sense of self-presence and greater enjoyment. If all members of the group look the same, the stress of individuality and being present in the environment may be distributed across group members. The visual cue that every member of the group shares identical features may lower an individual's sense of ownership of their self and embodiment, affecting their sense of self as an individual more than how it affects their identification with the group.

Visual uniformity or similarity is often taken into consideration when wanting to create a stronger sense of group identity (Kim, 2009). However, how this transformation influences social interactions and behavior in virtual environments with time and use has remained open to question. It could be argued that visual appearance used in certain contexts can serve specific purposes in shaping social interactions. Given avatar appearance did not have an effect on entitativity and lowered motion synchrony, it may not make sense to use visual cues as a unifier for group identification. Conversely, avatar appearance did have an effect on variables such as self-presence, which, given its role in immersion, and in turn, attention in and connection to the environment, it may be unfavorable to have a uniform avatar in a group setting and suppress individuals' visual cues. At the same time, if the goal of social interactions is for enjoyment purposes, having uniform avatars may allow people to focus less on their individual role in a group setting and more on enjoying the task at hand. However, we note that further research is needed to better understand how such transformations impact enjoyment. While we designed our own measurement of enjoyment, it may require a more nuanced understanding to draw definite conclusions about how it interacts with shared visual cues.

Lastly, we found that it is important to consider individual differences in how people's experience in VR changes over time, as evidenced by our findings that individual differences accounted for different initial baselines and differences in how people's experiences evolved to varying degrees.

Study 2

Complementary to Study 1's focus on transformation of avatar appearance, Study 2 focuses on the transformation of environmental context. Based on the preliminary findings of Study 1, we generated and pre-registered hypotheses related to time and the virtual environment for Study 2 (pre-registration at https://osf.io/s37xc). As Table 4 lays out, given the beneficial effects that being in spacious, panoramic environments, and outdoor, natural environments provide, we hypothesized that participants will be able to interact with one another more freely in panoramic environments than in constrained environments. We anticipate that this increase in interaction and engagement will foster a greater sense of entitativity and enjoyment. Similarly, outdoor, natural environments have been shown to have restorative properties, which should improve perceived restorativeness for these environments.

In this study, participants at each weekly session were exposed to one of four possible types of virtual environments (2 *spaciousness* × 2 *setting conditions*): a panoramic outdoor environment, a panoramic indoor environment, a constrained outdoor environment, and a constrained indoor environment. Along with the dependent variables examined in Study 1, Study 2 examines the influence of time and virtual environment on additional variables such as perceived restorativeness and affect (pleasure and arousal).

Table 4. Pre-registered self-report measure and nonverbal behavior hypotheses

Independent variable	Hypotheses	
Time	Entitativity (H1)	Increase over time
	Presence: self (H2a), social (H2b), and spatial (H2c)	Increase over time
	Photographic realism (H3)	Increase over time
Panoramic vs. constrained	Nonverbal behavior or synchrony (H4)	Increase in nonverbal synchrony in panoramic
	Perceived restorativeness (H5)	Greater in panoramic
	Entitativity (H6)	Greater in panoramic
	Affect: pleasure (H7a), arousal (H7b)	Greater in panoramic
	Enjoyment (H8)	Greater in panoramic
Outdoors vs. indoors	Nonverbal behavior or synchrony (H9)	Greater outdoors
	Perceived restorativeness (H10)	Greater outdoors
	Enjoyment (H11)	Increase in nonverbal synchrony outdoors
Other	Nonverbal synchrony exists in VR (H12)	

Method
Participants
Participants were 171 university students enrolled in a 10-week course about VR. At the beginning of the course, students were invited to participate in an IRB-approved study of how repeated exposure to VR influenced their individual and group behavior. While all students who were part of the course took part in all the VR activities, only those who consented to participate in the study had their data included in the study. Of the 171 students in the course, 158 consented to participate in the study. The 137 participants who participated in five or more of the eight weekly sessions ($M = 78$, $F = 59$) were between 18 and 49 years old ($M = 20.9$, $SD = 2.78$; $n_{18\sim20} = 62$, $n_{21\sim23} = 71$, $n_{24\sim49} = 5$) and identified as Asian or Asian-American ($n = 47$), White ($n = 41$), multiracial ($n = 19$), African, African-American, or Black ($n = 12$), Hispanic or LatinX ($n = 8$), Native Hawaiian or other Pacific Island ($n = 5$), Indigenous/Native American, Alaska Native, First Nations ($n = 2$), declined to or did not respond ($n = 2$), Middle Eastern ($n = 1$), and a racial group not listed ($n = 1$). Participants had varying levels of experience with VR ($n_0 = 50$, $n_1 = 29$, $n_2 = 23$, $n_{3\sim10} = 26$, $n_{20\sim50} = 4$, $n_{90} = 2$, $n_{100} = 4$). Prior to the course, 86 participants were not familiar with anyone in their discussion group and others reported knowing one ($n_1 = 40$) or more members ($n_1 = 13$, $n_3 = 5$, $n_4 = 6$, $n_5 = 1$, $n_7 = 1$).

Virtual environments
As in Study 1, weekly discussion sessions were hosted in ENGAGE. There were four types of virtual environments (2 spaciousness × 2 setting): (a) panoramic outdoors, (b) panoramic indoors, (c) constrained outdoors, or (d) constrained indoors (Figure 6). Each environment was built by research personnel using 3D objects. In total, there were 192 uniquely-built environments that differed in size of moving area and height. As suggested by Reeves et al. (2015), as variance in media is growing, so should variance in media research. As the authors argue, any media chosen as a stimulus can have a list of features that may be psychologically relevant and interact with the primary factors in an experiment. Selecting one idealized representative stimuli from each end of the distribution can increase Type I, II, and III errors. Through stimulus sampling and statistical methods (e.g., using a mixed statistical model that factors in fixed and random effects), we are able to better understand media that may be found in real-world experiences (Judd et al., 2012; Westfall et al., 2014). To evaluate whether these manipulations work across a range of environments and examine generalization of results across stimuli, we created 192 unique environments which were rigorously controlled in terms of our theoretical variables related to context, but also contained diverse thematic features, as opposed to relying on a single stimuli manipulation.

The moving area of the environments was measured by adding markers to the corners and ceilings of the environments inside ENGAGE and then calculating the areas using the positional data of the markers. By design, the panoramic environments ($n = 48$, $M = 39494.52$, $SD = 51231.27$) were 1778.67% larger than constrained environments ($n = 48$, $M = 2102.26$, $SD = 9367.35$) [$t(50.139) = 4.97$, $p < .0001$] and were 84.17% greater in max heights ($n = 48$, $M = 29.062$, $SD = 26.05$) than those of constrained environments ($n = 48$, $M = 15.78$, $SD = 15.78$), [$t(77.41) = 2.781$,

$p = .0068$]. As a design artifact (i.e., indoor environments could not be infinitely large), the moving areas of the outdoor environments ($n = 48$, $M = 36745.89$, $SD = 53455.15$) were 657.51% larger than the indoor environments ($n = 48$, $M = 4850.89$, $SD = 7030.96$), ($t(48.63) = 4.099$, $p = .00016$), but did not differ in max height [outdoor environments ($n = 48$, $M = 22.33$, $SD = 23.45$), indoor environments ($n = 48$, $M = 23.56$, $SD = 21.29$), [$t(93.13) = 0.27$, $p = .79$]].

Avatar
All participants were asked to use the customization tool to make an avatar that looked and felt like their offline selves.

Procedure[3]
Participants selected a discussion group that fit their schedule and availability, resulting in 24 groups that met weekly for 8 weeks and varied in size from five to eight members ($M = 6.71$, $SD = 0.81$). The sizes of actual attended groups ranged from 2 to 11 members (Week 1 $M = 6.38$, $SD = 1.47$; Week 2 $M = 6.25$, $SD = 1.48$; Week 3 $M = 6.08$, $SD = 1.18$; Week 4 $M = 6.29$, $SD = 1.23$; Week 5 $M = 6.38$, $SD = 1.35$; Week 6 $M = 6.25$, $SD = 1.33$, Week 7 $M = 6.00$, $SD = 1.50$; Week 8 $M = 5.75$, $SD = 2.01$). Each week, each group was assigned to a set of four between-subject conditions (2 × 2 design) via a Latin square randomization scheme that ensured each group experienced each condition once and that each condition appeared equally across the weekly schedule (Table 5). The sessions were led by one of three instructors. Each instructor led the same eight groups every week.

A training session was held in the first week of the course, during which participants were guided through how to use the ENGAGE interface and the controllers to navigate the virtual environment. As in Study 1, during these training sessions, the teaching staff was available to assist via Zoom when participants faced technical mishaps in hardware and software.

The first discussion session began in the second week, during which participants completed a series of small-group activities to further familiarize them with the ENGAGE environment and its tools. All discussions, except in the fifth session, had a creative activity, which involved creating, brainstorming, or prototyping an idea using the tools available on ENGAGE (e.g., drawing with the 3D pen, bringing in 3D models, writing on whiteboards). The 30-min sessions were divided into a 10-min full-group discussion and recap of the course material, a 15-min individual creative activity based on a prompt, and a 5-min sharing of the final product of the activity portion (Table 6).

Measures
As in Study 1, multiple aspects of individuals' behaviors and attitudes were measured at the start of the study (pre-test), and during and after each of the eight weekly sessions (see Table 7[4]).

Weekly repeated measures
Nonverbal behavior: motion synchrony
As in Study 1, synchrony was computed for each participant for each week as the rank correlation of head speed over the entire (approximately 30 min) session.

Perceived restorativeness
Perceived restorativeness, the restorative quality and potential of environments, was measured using four items adapted from the Perceived Restorativeness Scale (Hartig et al., 1996)

Figure 6. Environment types used every session. There were four possible types of virtual environments (2 spaciousness × 2 setting): (1) panoramic outdoors, (2) panoramic indoors, (3) constrained outdoors, or (4) constrained indoors.

Table 5. Participants in each group ($n1 = 8$, $n2 = 24$) were randomly assigned to spaciousness and setting condition (panoramic vs. constrained, outdoors vs. indoors) via a Latin square randomization scheme

	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8
Group 1	Outdoors, panoramic	Outdoors, panoramic	Outdoors, constrained	Indoors, panoramic	Indoors, constrained	Indoors, constrained	Indoors, panoramic	Outdoors, constrained
Group 2	Outdoors, panoramic	Indoors, panoramic	Outdoors, panoramic	Indoors, constrained	Outdoors, constrained	Outdoors, constrained	Indoors, constrained	Indoors, panoramic
Group 3	Indoors, panoramic	Indoors, constrained	Outdoors, panoramic	Outdoors, constrained	Outdoors, panoramic	Indoors, panoramic	Outdoors, constrained	Indoors, constrained
Group 4	Indoors, constrained	Outdoors, constrained	Indoors, panoramic	Indoors, panoramic	Outdoors, panoramic	Indoors, constrained	Outdoors, panoramic	Outdoors, constrained
Group 5	Outdoors, constrained	Indoors, panoramic	Indoors, constrained	Indoors, constrained	Indoors, panoramic	Outdoors, constrained	Outdoors, panoramic	Outdoors, panoramic
Group 6	Indoors, panoramic	Indoors, constrained	Outdoors, constrained	Outdoors, constrained	Indoors, constrained	Outdoors, panoramic	Indoors, panoramic	Outdoors, panoramic
Group 7	Indoors, constrained	Outdoors, constrained	Indoors, panoramic	Outdoors, panoramic	Outdoors, constrained	Outdoors, panoramic	Indoors, constrained	Indoors, panoramic
Group 8	Outdoors, constrained	Outdoors, panoramic	Indoors, constrained	Outdoors, panoramic	Indoors, panoramic	Indoors, panoramic	Outdoors, constrained	Indoors, constrained
Group 9	Outdoors, panoramic	Outdoors, panoramic	Outdoors, constrained	Indoors, panoramic	Indoors, constrained	Indoors, constrained	Indoors, panoramic	Outdoors, constrained
Group 10	Outdoors, panoramic	Indoors, panoramic	Outdoors, panoramic	Indoors, constrained	Outdoors, constrained	Outdoors, constrained	Indoors, constrained	Indoors, panoramic
Group 11	Indoors, panoramic	Indoors, constrained	Outdoors, panoramic	Outdoors, constrained	Outdoors, panoramic	Indoors, panoramic	Outdoors, constrained	Indoors, constrained
Group 12	Indoors, constrained	Outdoors, constrained	Indoors, panoramic	Indoors, panoramic	Outdoors, panoramic	Indoors, constrained	Outdoors, panoramic	Outdoors, constrained
Group 13	Outdoors, constrained	Indoors, panoramic	Indoors, constrained	Indoors, constrained	Indoors, panoramic	Outdoors, constrained	Outdoors, panoramic	Outdoors, panoramic
Group 14	Indoors, panoramic	Indoors, constrained	Outdoors, constrained	Outdoors, constrained	Indoors, constrained	Outdoors, panoramic	Indoors, panoramic	Outdoors, panoramic
Group 15	Indoors, constrained	Outdoors, constrained	Indoors, panoramic	Outdoors, panoramic	Outdoors, constrained	Outdoors, panoramic	Indoors, constrained	Indoors, panoramic
Group 16	Outdoors, constrained	Outdoors, panoramic	Indoors, constrained	Outdoors, panoramic	Indoors, panoramic	Indoors, panoramic	Outdoors, constrained	Indoors, constrained
Group 17	Outdoors, panoramic	Outdoors, panoramic	Outdoors, constrained	Indoors, panoramic	Indoors, constrained	Indoors, constrained	Indoors, panoramic	Outdoors, constrained
Group 18	Outdoors, panoramic	Indoors, panoramic	Outdoors, panoramic	Indoors, constrained	Outdoors, constrained	Outdoors, constrained	Indoors, constrained	Indoors, panoramic
Group 19	Indoors, panoramic	Indoors, constrained	Outdoors, panoramic	Outdoors, constrained	Outdoors, panoramic	Indoors, panoramic	Outdoors, constrained	Indoors, constrained
Group 20	Indoors, constrained	Outdoors, constrained	Indoors, panoramic	Indoors, panoramic	Outdoors, panoramic	Indoors, constrained	Outdoors, panoramic	Outdoors, constrained
Group 21	Outdoors, constrained	Indoors, panoramic	Indoors, constrained	Indoors, constrained	Indoors, panoramic	Outdoors, constrained	Outdoors, panoramic	Outdoors, panoramic
Group 22	Indoors, panoramic	Indoors, constrained	Outdoors, constrained	Outdoors, constrained	Indoors, constrained	Outdoors, panoramic	Indoors, panoramic	Outdoors, panoramic
Group 23	Indoors, constrained	Outdoors, constrained	Indoors, panoramic	Outdoors, panoramic	Outdoors, constrained	Outdoors, panoramic	Indoors, constrained	Indoors, panoramic
Group 24	Outdoors, constrained	Outdoors, panoramic	Indoors, constrained	Outdoors, panoramic	Indoors, panoramic	Indoors, panoramic	Outdoors, constrained	Indoors, constrained

Note. This design ensured that each group experienced each condition once and that each condition appeared equally across the weekly schedule.

using a 5-point Likert scale (1 = Not at all to 5 = Extremely). Sample items include "Spending time here gave me a good break from my day-to-day routine" and "There is too much going on in this environment." Weekly perceived restorativeness scores were calculated as the mean of four item responses (Cronbach's $\alpha = 0.71$), with higher scores indicating greater perceived restorativeness of the environment.

Pleasure and arousal

Individual ratings for perceived pleasure and arousal were obtained after each weekly VR session using the Self-Assessment Manikin (Bradley & Lang, 1994) non-verbal pictorial scale accompanied by a pair of adjectives associated with the pleasure and arousal dimensions (Pleasure: 1 = Bored to 9 = Relaxed; Arousal: 1 = Calm to 9 = Excited). Higher scores indicate greater pleasure (i.e., relaxation) or arousal (i.e., excited).

Self, social, and spatial presence

Items were adapted from Study 1 to include an additional item and utilize a 5-point Likert scale (1 = Not at all to 5 = Extremely). Self, social, and spatial presence were measured as the level of agreement with three items (Cronbach's α = 0.84 for self-presence, 0.79 for social presence, and 0.82 for spatial presence). Weekly scores for each of the three types of presence were calculated as the mean of the three items, with higher scores indicating greater perceived presence.

Table 6. Study 2 weekly topics and activities

Session	Activity
1	• Acclimate participants to the headset and platform, leaving margin for technical or content issues • Activity: Consider the affordances of VR and create a prototype of something that leverages the uniqueness of VR
2	• Full-group discussion on what activities heightened sense of presence in VR • Activity: Create something frightening that induces a feeling of high presence
3	• Full-group discussion reflecting on participants' experience visiting various sites in AltspaceVR (e.g., an art exhibition, solar system) • Activity: Consider the affordances of VR to make a difficult concept easier to understand
4	• Full-group discussion on how to improve ENGAGE's avatar if the participant were in charge of ENGAGE • Activity: Create something that reimagines avatars and representations of the self
5	• Small-group discussions reflecting on various VR empathy experiences
6	• Full-group discussion how VR is used for medical applications and well-being • Activity: Create a meditation room or "safe-space"
7	• Full-group discussion on VR's role in people's attitudes and actions toward climate change • Activity: Brainstorm an idea of how to communicate a message about climate change
8	• Full-group discussion on VR's role in the future of sports and fitness • Activity: Create and playtest a VR-based game

Individual differences measures

Additional individual differences predictors included in the model building process, including environmental identification and prior VR use, were trimmed from the reporting because none of these variables were related to baseline levels, rates of change, or environment conditions for any of the 10 outcomes.

Data analysis

Individual differences in how individuals' behaviors and attitudes changed across the 8 weeks and in relation to spaciousness and setting conditions, and how these effects were related to gender were examined using linear growth models with time-invariant and time-varying covariates (Grimm et al., 2016). Specifically, each of the 10 repeated measures outcomes was modeled as

$$outcome_{ti} = \beta_{0i} + \beta_{1i}(week_{ti}) + \beta_{2i}(spaciousness_{ti}) \\ + \beta_{3i}(setting_{ti}) + \beta_{4i}(spaciousness_{ti} \cdot setting_{ti}) \\ + e_{ti} \quad (1)$$

where the outcome of interest for person i at occasion t, $outcome_{ti}$ is modeled as a function of person-specific intercepts, β_{0i}, person-specific linear slopes, β_{1i}, that indicate rate of change across weeks, person-specific spaciousness effects, β_{2i}, that indicate the difference between panoramic and constrained conditions, person-specific setting effects, β_{3i}, that

indicate the difference between outdoors and indoors conditions, an interaction term β_{4i}, that indicates extent of moderation between the spaciousness and setting manipulations, and residual error, e_{ti} that is assumed normally distributed with standard deviation σ_e. The person-specific intercepts, linear slopes, and spaciousness and setting effects are simultaneously modeled as

$$\beta_{0i} = \gamma_{00} + \gamma_{01}(gender_i) + u_{0i} \quad (2)$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11}(gender_i) \quad (3)$$

$$\beta_{2i} = \gamma_{20} + \gamma_{21}(gender_i) \quad (4)$$

$$\beta_{3i} = \gamma_{30} + \gamma_{31}(gender_i) \quad (5)$$

$$\beta_{4i} = \gamma_{40} + \gamma_{41}(gender_i) \quad (6)$$

where γ_{00} and γ_{01} describe the linear trajectory of change for the prototypical individual, γ_{20} describes the prototypical effect of the spaciousness manipulation, γ_{30} describes the prototypical effect of the setting manipulation; γ_{40} describes the prototypical spaciousness and setting interaction effect; γ_{01}, γ_{11}, γ_{21}, and γ_{31} indicate how individual differences in level, change, and the manipulations are related to gender, and u_{0i} is residual unexplained differences that are assumed normally distributed with standard deviation σ_{u0}. As in Study 1, all models were fit to the data in R using the lme4 and lmerTest libraries with restricted maximum likelihood estimation, incomplete data treated as missing at random, and statistical significance evaluated at alpha $= 0.05$.

Results

Results from growth models with time-varying predictors [week, spaciousness (panoramic $= 1$ vs. constrained $= 0$), and setting (outdoors $= 1$ vs. indoors $= 0$)] and a time-invariant predictor (gender), are presented separately for all 10 outcomes (synchrony, perceived restorativeness, entitativity, pleasure, arousal, and self, social, and spatial presence).

Synchrony

The prototypical participant's motion synchrony was positive, $\gamma_{00} = 0.015$, $p < .001$, confirming H12. Motion synchrony increased slightly, but not significantly at a rate of $\gamma_{10} = 0.00026$, $p = .559$ points per week over the 8 weeks of study. There was a significant effect of the spaciousness manipulation, such that individuals had higher synchrony when in panoramic environments than constrained environments, $\gamma_{20} = 0.010005$, $p = .0004$ (H4). There was no evidence that the setting manipulation influenced synchrony, $\gamma_{30} = 0.0019$, $p = .507$ (H9), interaction effects, or gender differences. Figure 7 shows the strength of synchrony over time offset, indicating the time dependence of synchrony.

Perceived restorativeness

The prototypical participant's perceived restorativeness decreased from an initial value of $\gamma_{00} = 3.169$, $p < .001$ (on a 5-point scale) at a rate of $\gamma_{10} = -0.027$, $p < .001$ points per week. There was a significant effect of both the setting and

Table 7. Study 2 means and standard deviations (in parentheses) of repeated measures across 8 weeks

DV	Environmental Condition	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Total
Synchrony	Constrained	0.012 (0.025)	0.015 (0.025)	0.025 (0.032)	0.012 (0.023)	0.020 (0.032)	0.025 (0.035)	0.012 (0.027)	0.019 (0.033)	0.017 (0.029)
	Indoors	0.017 (0.028)	0.025 (0.026)	0.026 (0.033)	0.007 (0.022)	0.031 (0.036)	0.030 (0.032)	0.018 (0.025)	0.016 (0.039)	0.022 (0.031)
	Constrained indoors	0.011 (0.029)	0.022 (0.024)	0.021 (0.037)	0.002 (0.025)	0.015 (0.032)	0.020 (0.024)	0.020 (0.023)	0.022 (0.032)	0.016 (0.029)
	Panoramic indoors	0.022 (0.028)	0.029 (0.027)	0.031 (0.030)	0.011 (0.019)	0.047 (0.032)	0.038 (0.036)	0.015 (0.028)	0.012 (0.044)	0.026 (0.033)
	Outdoors	0.016 (0.024)	0.019 (0.034)	0.024 (0.025)	0.018 (0.024)	0.038 (0.035)	0.036 (0.046)	0.016 (0.031)	0.018 (0.029)	0.023 (0.033)
	Constrained outdoors	0.014 (0.020)	0.009 (0.024)	0.029 (0.026)	0.019 (0.020)	0.025 (0.032)	0.030 (0.043)	0.005 (0.030)	0.016 (0.033)	0.018 (0.029)
	Panoramic outdoors	0.018 (0.028)	0.030 (0.040)	0.016 (0.022)	0.016 (0.029)	0.050 (0.035)	0.040 (0.048)	0.026 (0.028)	0.019 (0.027)	0.028 (0.036)
	Panoramic	0.020 (0.028)	0.029 (0.034)	0.026 (0.028)	0.014 (0.025)	0.049 (0.033)	0.039 (0.042)	0.021 (0.028)	0.016 (0.036)	0.027 (0.034)
Perceived restorativeness	Constrained	3.32 (0.72)	3.20 (0.73)	3.17 (0.66)	3.21 (0.67)	2.92 (0.79)	3.11 (0.70)	3.03 (0.67)	2.95 (0.80)	3.12 (0.72)
	Indoors	3.44 (0.65)	3.06 (0.67)	3.10 (0.72)	3.14 (0.71)	3.03 (0.61)	3.19 (0.74)	3.11 (0.65)	2.99 (0.78)	3.14 (0.70)
	Constrained indoors	3.30 (0.77)	3.13 (0.60)	3.03 (0.67)	3.15 (0.59)	3.09 (0.61)	3.01 (0.68)	2.88 (0.69)	2.80 (0.70)	3.05 (0.67)
	Panoramic indoors	3.57 (0.49)	2.98 (0.73)	3.16 (0.75)	3.13 (0.83)	2.97 (0.61)	3.35 (0.76)	3.33 (0.53)	3.30 (0.81)	3.23 (0.70)
	Outdoors	3.25 (0.64)	3.23 (0.77)	3.38 (0.67)	3.21 (0.71)	3.08 (0.94)	3.23 (0.74)	3.22 (0.65)	3.23 (0.71)	3.23 (0.74)
	Constrained outdoors	3.34 (0.68)	3.27 (0.84)	3.29 (0.64)	3.26 (0.72)	2.73 (0.91)	3.22 (0.72)	3.19 (0.62)	3.20 (0.89)	3.19 (0.76)
	Panoramic outdoors	3.14 (0.59)	3.19 (0.69)	3.48 (0.70)	3.15 (0.71)	3.42 (0.85)	3.24 (0.77)	3.25 (0.69)	3.25 (0.59)	3.26 (0.71)
	Panoramic	3.41 (0.56)	3.08 (0.71)	3.29 (0.74)	3.14 (0.76)	3.19 (0.77)	3.29 (0.76)	3.30 (0.60)	3.27 (0.67)	3.25 (0.70)
Entitativity	Constrained	2.95 (0.58)	3.11 (0.60)	3.03 (0.62)	3.28 (0.67)	3.05 (0.63)	2.94 (0.77)	3.11 (0.75)	2.97 (0.77)	3.06 (0.68)
	Indoors	3.05 (0.57)	3.28 (0.60)	3.02 (0.72)	3.11 (0.82)	3.01 (0.70)	2.89 (0.74)	3.18 (0.69)	3.12 (0.76)	3.08 (0.70)
	Constrained indoors	2.86 (0.57)	3.17 (0.55)	2.95 (0.68)	3.36 (0.74)	2.96 (0.57)	2.68 (0.80)	3.15 (0.75)	3.04 (0.72)	3.02 (0.69)
	Panoramic indoors	3.22 (0.52)	3.39 (0.63)	3.08 (0.76)	2.84 (0.83)	3.05 (0.82)	3.06 (0.64)	3.21 (0.62)	3.25 (0.82)	3.14 (0.71)
	Outdoors	3.10 (0.60)	3.01 (0.65)	3.10 (0.58)	3.14 (0.66)	3.10 (0.74)	3.23 (0.68)	2.98 (0.79)	3.13 (0.77)	3.10 (0.68)
	Constrained outdoors	3.04 (0.58)	3.05 (0.66)	3.11 (0.54)	3.22 (0.62)	3.14 (0.68)	3.22 (0.65)	3.06 (0.76)	2.88 (0.86)	3.10 (0.66)
	Panoramic outdoors	3.20 (0.63)	2.96 (0.65)	3.10 (0.63)	3.04 (0.70)	3.06 (0.80)	3.23 (0.71)	2.90 (0.82)	3.29 (0.68)	3.10 (0.71)
	Panoramic	3.21 (0.56)	3.18 (0.67)	3.09 (0.70)	2.95 (0.76)	3.06 (0.80)	3.14 (0.68)	3.07 (0.73)	3.28 (0.73)	3.12 (0.72)
Pleasure	Constrained	6.33 (1.75)	5.89 (1.99)	5.68 (2.05)	5.77 (1.93)	5.45 (2.06)	5.73 (2.10)	5.31 (2.19)	5.90 (2.37)	5.76 (2.06)
	Indoors	6.66 (1.59)	6.37 (1.76)	5.53 (2.08)	5.44 (2.15)	5.48 (2.02)	6.05 (1.86)	5.69 (2.08)	5.41 (2.35)	5.86 (2.02)
	Constrained indoors	6.24 (1.94)	6.00 (1.80)	5.48 (1.90)	5.68 (1.74)	5.59 (1.88)	5.93 (2.12)	5.38 (2.28)	5.27 (2.27)	5.70 (2.00)
	Panoramic indoors	7.03 (1.13)	6.74 (1.67)	5.57 (2.24)	5.19 (2.51)	5.35 (2.18)	6.15 (1.62)	6.00 (1.85)	5.63 (2.52)	6.01 (2.03)
	Outdoors	6.47 (1.44)	5.84 (1.95)	6.23 (2.04)	5.71 (2.10)	5.84 (2.28)	5.46 (2.23)	5.70 (2.23)	6.26 (1.95)	5.92 (2.06)
	Constrained outdoors	6.42 (1.54)	5.78 (2.20)	5.87 (2.19)	5.84 (2.08)	5.30 (2.26)	5.52 (2.10)	5.22 (2.12)	6.89 (2.23)	5.82 (2.12)
	Panoramic outdoors	6.55 (1.30)	5.90 (1.68)	6.68 (1.77)	5.58 (2.15)	6.35 (2.21)	5.41 (2.36)	6.19 (2.27)	5.87 (1.67)	6.02 (2.00)
	Panoramic	6.85 (1.20)	6.33 (1.71)	6.02 (2.12)	5.40 (2.31)	5.85 (2.24)	5.78 (2.05)	6.08 (2.04)	5.78 (2.01)	6.01 (2.02)
Arousal	Constrained	4.62 (1.95)	4.44 (1.96)	4.33 (2.14)	3.98 (1.92)	4.06 (2.19)	3.91 (1.94)	3.81 (2.01)	3.96 (1.85)	4.15 (2.00)
	Indoors	5.25 (2.08)	4.95 (1.98)	3.91 (1.97)	3.85 (1.91)	3.87 (1.88)	3.95 (2.04)	4.08 (2.05)	3.67 (1.92)	4.23 (2.05)

Table 7. (continued)

DV	Environmental Condition	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Total
	Constrained indoors	4.85	4.50	3.72	3.82	4.16	3.89	3.78	3.80	4.08
		(1.97)	(1.83)	(1.83)	(1.63)	(2.03)	(2.04)	(2.04)	(1.71)	(1.91)
	Panoramic indoors	5.61	5.42	4.05	3.89	3.58	4.00	4.36	3.47	4.37
		(2.14)	(2.05)	(2.09)	(2.19)	(1.69)	(2.06)	(2.04)	(2.25)	(2.17)
	Outdoors	4.70	4.72	4.52	4.00	3.89	3.72	3.96	4.30	4.21
		(1.95)	(2.00)	(2.19)	(1.97)	(2.27)	(1.90)	(2.01)	(1.81)	(2.04)
	Constrained outdoors	4.39	4.38	4.90	4.11	3.97	3.93	3.85	4.21	4.23
		(1.93)	(2.11)	(2.27)	(2.12)	(2.39)	(1.86)	(2.01)	(2.07)	(2.10)
	Panoramic outdoors	5.14	5.10	4.04	3.88	3.81	3.56	4.07	4.35	4.20
		(1.93)	(1.84)	(2.03)	(1.82)	(2.18)	(1.94)	(2.04)	(1.66)	(1.98)
	Panoramic	5.43	5.27	4.05	3.88	3.69	3.78	4.23	4.02	4.29
		(2.06)	(1.94)	(2.05)	(1.98)	(1.94)	(2.00)	(2.03)	(1.93)	(2.08)
Self-presence	Constrained	2.52	2.68	2.65	2.77	2.66	2.84	2.78	2.68	2.69
		(0.78)	(0.85)	(0.79)	(0.77)	(0.85)	(0.88)	(0.78)	(0.94)	(0.83)
	Indoors	2.54	2.81	2.73	2.67	2.78	2.71	2.78	2.83	2.73
		(0.73)	(0.88)	(0.80)	(0.77)	(0.72)	(0.81)	(0.83)	(0.93)	(0.81)
	Constrained indoors	2.35	2.72	2.74	2.77	2.62	2.60	2.78	2.67	2.65
		(0.76)	(0.91)	(0.82)	(0.75)	(0.78)	(0.87)	(0.84)	(0.89)	(0.83)
	Panoramic indoors	2.69	2.90	2.72	2.57	2.95	2.81	2.77	3.09	2.80
		(0.67)	(0.86)	(0.80)	(0.80)	(0.64)	(0.76)	(0.85)	(0.96)	(0.78)
	Outdoors	2.63	2.76	2.62	2.72	2.67	2.98	2.73	2.81	2.74
		(0.76)	(0.81)	(0.76)	(0.79)	(0.94)	(0.89)	(0.81)	(0.78)	(0.82)
	Constrained outdoors	2.69	2.64	2.57	2.77	2.69	3.10	2.78	2.70	2.74
		(0.77)	(0.81)	(0.77)	(0.79)	(0.93)	(0.83)	(0.72)	(1.04)	(0.83)
	Panoramic outdoors	2.54	2.90	2.69	2.67	2.66	2.89	2.68	2.87	2.75
		(0.76)	(0.80)	(0.75)	(0.80)	(0.96)	(0.94)	(0.91)	(0.58)	(0.82)
	Panoramic	2.64	2.90	2.71	2.62	2.80	2.85	2.73	2.95	2.77
		(0.70)	(0.82)	(0.77)	(0.79)	(0.82)	(0.85)	(0.86)	(0.74)	(0.80)
Social presence	Constrained	3.09	3.40	3.34	3.56	3.34	3.22	3.24	3.16	3.30
		(0.88)	(0.76)	(0.72)	(0.77)	(0.80)	(0.84)	(0.89)	(0.87)	(0.82)
	Indoors	3.28	3.62	3.30	3.18	3.33	3.22	3.14	3.31	3.30
		(0.90)	(0.71)	(0.81)	(0.98)	(0.73)	(0.87)	(0.82)	(0.76)	(0.83)
	Constrained indoors	3.02	3.58	3.34	3.61	3.18	2.95	3.27	3.27	3.28
		(0.93)	(0.74)	(0.70)	(0.78)	(0.77)	(0.92)	(0.92)	(0.69)	(0.83)
	Panoramic indoors	3.51	3.66	3.27	2.74	3.48	3.45	3.01	3.37	3.32
		(0.82)	(0.68)	(0.90)	(0.98)	(0.66)	(0.77)	(0.70)	(0.87)	(0.83)
	Outdoors	3.19	3.17	3.40	3.40	3.46	3.38	3.21	3.17	3.31
		(0.83)	(0.76)	(0.75)	(0.87)	(0.92)	(0.69)	(0.89)	(0.86)	(0.83)
	Constrained outdoors	3.17	3.21	3.34	3.53	3.51	3.51	3.21	3.00	3.33
		(0.83)	(0.74)	(0.75)	(0.77)	(0.82)	(0.64)	(0.87)	(1.09)	(0.81)
	Panoramic outdoors	3.23	3.14	3.47	3.24	3.42	3.27	3.21	3.28	3.28
		(0.85)	(0.78)	(0.75)	(0.96)	(1.02)	(0.71)	(0.92)	(0.68)	(0.84)
	Panoramic	3.41	3.41	3.35	3.02	3.45	3.36	3.10	3.31	3.30
		(0.84)	(0.77)	(0.84)	(0.99)	(0.85)	(0.74)	(0.80)	(0.75)	(0.83)
Spatial presence	Constrained	3.39	3.29	3.17	3.29	3.01	3.11	3.09	3.06	3.18
		(0.79)	(0.82)	(0.73)	(0.82)	(0.82)	(0.80)	(0.79)	(0.92)	(0.81)
	Indoors	3.52	3.31	3.27	3.14	3.07	3.15	3.10	3.10	3.22
		(0.80)	(0.86)	(0.79)	(0.95)	(0.77)	(0.79)	(0.79)	(0.80)	(0.83)
	Constrained indoors	3.32	3.29	3.21	3.33	2.89	2.99	3.05	3.01	3.14
		(0.86)	(0.91)	(0.80)	(0.90)	(0.87)	(0.84)	(0.73)	(0.85)	(0.85)
	Panoramic indoors	3.68	3.33	3.32	2.95	3.26	3.28	3.14	3.25	3.29
		(0.72)	(0.81)	(0.79)	(0.98)	(0.59)	(0.74)	(0.85)	(0.72)	(0.80)
	Outdoors	3.44	3.30	3.26	3.17	3.19	3.25	3.13	3.09	3.23
		(0.76)	(0.72)	(0.70)	(0.79)	(0.82)	(0.75)	(0.89)	(0.82)	(0.78)
	Constrained outdoors	3.45	3.28	3.13	3.26	3.13	3.24	3.14	3.14	3.23
		(0.72)	(0.73)	(0.67)	(0.77)	(0.74)	(0.75)	(0.88)	(1.04)	(0.77)
	Panoramic outdoors	3.42	3.31	3.41	3.07	3.25	3.25	3.12	3.06	3.23
		(0.82)	(0.72)	(0.72)	(0.80)	(0.89)	(0.77)	(0.92)	(0.68)	(0.79)
	Panoramic	3.59	3.32	3.35	3.02	3.25	3.27	3.13	3.13	3.26
		(0.77)	(0.76)	(0.76)	(0.88)	(0.75)	(0.75)	(0.88)	(0.69)	(0.79)
Enjoyment	Constrained	3.27	3.18	3.07	3.16	2.74	2.87	2.84	2.87	3.01
		(0.79)	(0.83)	(0.74)	(0.94)	(0.94)	(0.95)	(0.85)	(0.98)	(0.89)
	Indoors	3.56	3.29	2.86	2.95	2.79	2.93	2.92	2.76	3.02
		(0.83)	(0.86)	(0.86)	(0.94)	(0.80)	(0.94)	(0.80)	(1.00)	(0.91)
	Constrained indoors	3.26	3.33	2.88	3.16	2.78	2.61	2.76	2.65	2.93
		(0.86)	(0.82)	(0.74)	(0.82)	(0.85)	(0.93)	(0.78)	(0.89)	(0.87)
	Panoramic indoors	3.83	3.26	2.84	2.74	2.79	3.21	3.08	2.92	3.11
		(0.71)	(0.91)	(0.99)	(1.02)	(0.75)	(0.88)	(0.79)	(1.15)	(0.94)

(continued)

Table 7. (continued)

DV	Environmental Condition	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Total
	Outdoors	3.21 (0.72)	3.09 (0.74)	3.24 (0.75)	2.99 (0.99)	2.92 (0.96)	3.07 (0.97)	2.93 (0.88)	3.18 (0.86)	3.07 (0.87)
	Constrained outdoors	3.27 (0.72)	3.03 (0.82)	3.24 (0.71)	3.16 (1.03)	2.70 (1.03)	3.15 (0.91)	2.93 (0.92)	3.21 (1.03)	3.09 (0.90)
	Panoramic outdoors	3.11 (0.74)	3.16 (0.66)	3.24 (0.82)	2.80 (0.93)	3.13 (0.85)	3.00 (1.02)	2.93 (0.85)	3.16 (0.75)	3.06 (0.84)
	Panoramic	3.57 (0.794)	3.21 (0.793)	3.00 (0.919)	2.78 (0.96)	2.96 (0.81)	3.10 (0.95)	3.01 (0.82)	3.07 (0.92)	3.09 (0.89)
Realism	Constrained	32.56 (21.90)	38.86 (22.51)	52.55 (21.03)	45.83 (21.51)	44.66 (22.82)	46.73 (24.14)	45.05 (22.86)	46.25 (21.98)	43.86 (22.89)
	Indoors	32.69 (19.74)	41.22 (23.45)	49.24 (25.53)	48.11 (21.93)	48.63 (19.79)	47.54 (23.14)	47.76 (20.60)	46.25 (23.00)	44.91 (22.73)
	Constrained indoors	30.03 (21.49)	38.97 (21.63)	51.00 (21.35)	46.54 (21.97)	50.48 (21.06)	41.89 (24.34)	44.03 (22.48)	43.30 (22.72)	43.06 (22.74)
	Panoramic indoors	35.00 (18.04)	43.55 (25.34)	47.87 (28.60)	49.74 (22.18)	46.77 (18.60)	52.33 (21.28)	51.15 (18.41)	50.90 (23.29)	46.69 (22.62)
	Outdoors	34.68 (22.91)	40.49 (22.44)	51.63 (21.90)	43.71 (20.87)	41.41 (23.79)	50.34 (21.08)	50.35 (24.66)	52.36 (18.53)	45.48 (22.71)
	Constrained outdoors	35.26 (22.37)	38.75 (23.70)	54.00 (20.96)	45.30 (21.44)	38.63 (23.34)	51.74 (23.31)	46.19 (23.66)	50.90 (20.47)	44.67 (23.07)
	Panoramic outdoors	33.86 (24.16)	42.41 (21.22)	48.68 (23.10)	41.94 (20.40)	44.10 (24.29)	49.24 (19.42)	54.52 (25.37)	53.26 (17.53)	46.29 (22.36)
	Panoramic	34.58 (20.30)	43.00 (23.25)	48.19 (26.32)	45.45 (21.40)	45.44 (21.50)	50.76 (20.26)	52.67 (21.69)	52.36 (19.72)	46.50 (22.47)
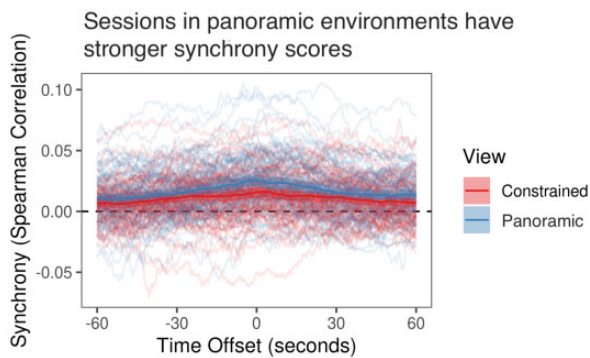
Figure 7. Effect of view on synchrony. This plot demonstrates that as the time offset of motion signals shifts away from zero (i.e., as one looks toward the right and left away from the center), synchrony (Y-axis) decreases. In this plot, synchrony for each group in each session is traced as a separate partially transparent line (185 total). The average of all sessions for a given avatar condition is the darker line, with the ribbon indicating 95% confidence intervals based on the underlying distribution. Each line is produced as the average of all unordered pairs in that session (from 3 to 36, $M = 18.7$, $SD = 8.59$), which is itself calculated from about 30 min of data per participant.

spaciousness manipulations, such that individuals reported greater perceived restorativeness when in panoramic environments than constrained environments, $\gamma_{20} = 0.168$, $p = .0005$ (H5), or in outdoor environments than indoor environments, $\gamma_{30} = 0.14$, $p = .004$ (H10). There was no evidence of interaction effects or gender differences.

Entitativity

The prototypical participant's entitativity decreased from an initial value of $\gamma_{00} = 3.03$, $p < .001$ (on a 7-point scale), though not significantly, at a rate of $\gamma_{10} = -0.005$, $p = .34$ points per week (H1). There was a significant effect of the spaciousness manipulation, such that individuals reported greater entitativity when in panoramic environments than constrained environments, $\gamma_{20} = 0.093$, $p = .0092$ (H6).

There was no evidence that the setting manipulation influenced entitativity, $\gamma_{30} = 0.048$, $p = .187$, interaction effects, or gender differences.

Pleasure

The prototypical participant's pleasure decreased from an initial value of $\gamma_{00} = 6.17$, $p < .001$ (on a 9-point scale) at a rate of $\gamma_{10} = -0.11$, $p < .001$ points per week. There was a significant effect of the spaciousness manipulation, such that individuals reported greater pleasure when in panoramic environments than constrained environments, $\gamma_{20} = 0.28$, $p = .037$ (H7a). There was no evidence of setting effects, $\gamma_{30} = 0.094$, $p = .504$, interaction effects, or gender differences.

Arousal

The prototypical participant's arousal decreased from an initial value of $\gamma_{00} = 4.54$, $p < .001$ (on a 9-point scale) at a rate of $\gamma_{10} = -0.14$, $p < .001$ points per week. There was a significant effect of the spaciousness manipulation, such that individuals reported greater arousal when in panoramic environments than constrained environments, $\gamma_{20} = 0.307$, $p = .0339$ (H7b). There was no evidence of setting effects, $\gamma_{30} = 0.118$, $p = .42$, interaction effects, or gender differences.

Presence
Self-presence

The prototypical participant's self-presence increased from an initial value of $\gamma_{00} = 2.46$, $p < .001$ (on a 5-point scale) at a rate of $\gamma_{10} = 0.022$, $p = .0021$ points per week (H2a). There was a significant effect of the spaciousness manipulation, such that individuals reported higher self-presence when in panoramic environments than constrained environments, $\gamma_{20} = 0.129$, $p = .0048$.

Social presence

The prototypical participant's social presence decreased from an initial value of $\gamma_{00} = 3.22$, $p < .001$ (on a 5-point scale) at a rate of $\gamma_{10} = -0.0159$, $p = .03$ points per week (H2b).

There was no evidence that the spaciousness manipulation influenced social presence, $\gamma_{20} = 0.015$, $p = .74$.

Spatial presence

The prototypical participant's spatial presence decreased from an initial value of $\gamma_{00} = 3.22$, $p < .001$ (on a 5-point scale) at a rate of $\gamma_{10} = -0.049$, $p < .001$ points per week (H2c). There was a significant effect of the spaciousness manipulation, such that individuals reported higher spatial presence when in panoramic environments than constrained environments, $\gamma_{20} = 0.128$, $p = .008$.

There was no evidence that the setting manipulation influenced self ($\gamma_{30} = 0.074$, $p = .109$), social ($\gamma_{30} = 0.038$, $p = .41$), or spatial ($\gamma_{30} = 0.071$, $p = .14$) presence. There was no interaction between the spaciousness and setting manipulations on self, social, or spatial presence (ps > 0.055). Individuals who identified as female had higher baseline levels of self ($\gamma_{01} = 0.24$, $p = .039$), social ($\gamma_{01} = 0.28$, $p = .016$), and spatial ($\gamma_{01} = 0.236$, $p = .029$) presence.

Enjoyment

The prototypical participant's enjoyment decreased from an initial value of $\gamma_{00} = 3.19$, $p < .001$ (on a 5-point scale) at a rate of $\gamma_{10} = -0.064$, $p < .001$ points per week. There was evidence that both the setting manipulation and spaciousness manipulation influenced enjoyment, such that individuals reported higher enjoyment when in panoramic environments than constrained environments, $\gamma_{20} = 0.166$, $p = .0043$ (H8) or in outdoor environments than in indoor environments, $\gamma_{30} = 0.13$, $p = .0267$ (H11). However, when the environment was one that was both outdoors and panoramic, there was a lower baseline level of enjoyment, $\gamma_{40} = -0.188$, $p = .024$. There was no evidence of gender differences.

Realism

The prototypical participant's realism increased from an initial value of $\gamma_{00} = 38.34$, $p < .001$ (on a 0–100, cartoon-like to photorealistic scale) at a rate of $\gamma_{10} = 1.76$, $p < .001$ points per week (H3). There was a significant effect of the spaciousness manipulation, such that individuals had higher realism when in panoramic environments than constrained environments, $\gamma_{20} = 3.57$, $p = .0083$. There was no evidence that the setting manipulation influenced realism, $\gamma_{30} = 1.9009$, $p = .166$. However, when the environment was one that was both outdoors and panoramic, there was a lower baseline level of realism, $\gamma_{40} = -4.0803$, $p = .035$. There was no evidence of gender differences.

Discussion

Study 2 examined the role of time and environmental context (spaciousness and setting) on participants' experience and group dynamics. Overall, the results showed that self-presence and realism increased over time, and social presence, spatial presence, and enjoyment decreased over time. While the effects of time were less robust in Study 2, the results hold true that people's behaviors and attitudes in VR changes with time and use.

In line with our hypotheses of the beneficial effects of being in a spacious, panoramic environment, during the weeks where participants were in a panoramic environment (i.e., environments in which people can see wide and far), their synchrony increased, and they reported greater perceived restorativeness, entitativity, pleasure, arousal, self and spatial presence, enjoyment, and realism. As panoramic environments naturally come with more visual components (i.e., there is more visible space, and more content that fills that space), this may have caused the surrounding environment to be more stimulating, leading to greater arousal. In panoramic environments, participants had the freedom to look around and focus their attention on different features, be it the other members of the group or what was in the immediate or far surrounding space. In contrast, a constrained environment may have led to feelings of confinement and forced people to pay their full attention to a limited amount of options. Whereas a constrained environment may have acted as a stressor to an individual's experience, potentially influencing social interactions that took place in the space, as well as resulting in a more critical evaluation of their sense of self, group members, and perception of experience, a panoramic environment may have provided a more restorative, open space that allowed them the freedom to let their mind wander.

Similarly, in line with our hypotheses of the restorative effects of being in an outdoor environment, during the weeks where participants were in an outdoor environment with elements of nature, their perceived restorativeness and enjoyment were greater. In addition to considering the beneficial, restorative properties that outdoor, natural environments provide, it is also important to note the context in which these environments were used. Oftentimes group discussions and social interactions occur in indoor environments in classrooms, conference rooms, or common spaces. The context of meeting with group members and engaging in a discussion in an outdoor environment—in between boulders, near ponds, or surrounded by a forest—may have provided an experience that is not common or easily accessible, leading to novelty, and in turn, greater enjoyment. The novelty of the environmental context in which the group interaction took place may have enriched not only one's perception of the experience, but also the social experience.

However, if the environment was one that was both panoramic and outdoors, reported enjoyment and realism were lower. Theories from evolutionary perspectives, namely Appleton's (1975) prospect refuge theory may lend a hand in understanding these outcomes. The prospect refuge theory argues that there is an innate human preference for environments that allow for both prospect and refuge. Ideal environments allow for a clear view of the scene and evaluation of opportunities (e.g., resources, places for hiding) and threats (e.g., predators, hazards). Environments that pose threats to survival may trigger negative reactions such as fear and avoidance (Ulrich, 1983). It is possible that interacting in large, open spaces with elements of nature that do not provide a sense of protection led to participants not enjoying their experience as much and being more alert and critical of their virtual surroundings. If an individual's experience is instilled with a sense of fear and endangerment, this may negatively influence any social interactions that take place, and as TSI would predict, this would continue to alter their behavior during and after the engagement in the virtual world.

Given that elements of the surrounding environment, such as how much space is visible and whether they are outdoors surrounded by nature, influence people's behaviors and attitudes within CVEs, the virtual environments in which such interactions are designed can be transformed in different ways. In particular, depending on what the desired goals of

these interactions are (e.g., social, team building, educational), the ways in which the virtual environment are structured can meet different needs and foster specific dynamics within groups.

General discussion
Summary of results
In Study 1, we examined the transformation of the self and others in a CVE by manipulating the avatar appearance of the participants. Participants wore either a self-avatar or a uniform avatar. We found that over time, presence (self, social, and spatial), enjoyment, entitativity, and realism all increased. Wearing a self-avatar increased nonverbal synchrony, self-presence, and realism, but decreased enjoyment. We also explored how much these outcomes may be mediated by individual differences and reported that those with more prior relationships had higher baseline levels of entitativity, enjoyment, and realism, but over time these individuals' perception of realism increased less than that of individuals with fewer prior relationships; those with prior VR experience had higher baseline levels of enjoyment; and those with higher group identification had a higher baseline level of social presence.

In Study 2, we examined the transformation of an environmental context by manipulating the virtual environment. Results showed that, as visible space increased, so did nonverbal synchrony, perceived restorativeness, entitativity, pleasure, arousal, self and spatial presence, enjoyment, and realism. Moreover, being in an outdoor environment led to greater reported perceived restorativeness and enjoyment. However, when the virtual environment was both panoramic and outdoors, reported enjoyment and realism were lower.

Based on the preliminary findings of Study 1, we hypothesized that there would be a robust effect of time. In line with Study 1, Study 2 results show that self-presence and realism increased over time. Oppositely, social and spatial presence, and enjoyment slightly decreased over time. We measured additional variables in Study 2, including perceived restorativeness, pleasure, and arousal, which also decreased over time.

Limitations
This study is the first large-scale, longitudinal, quantitative study of large groups in HMD-based CVEs. However, there are a variety of limitations. First, both studies were field experiments, which come with strengths and limitations. While field experiments allow for researchers to implement interventions and measure outcomes in naturalistic settings, there are constraints on how much control the researchers have on external conditions and potential intervening variables. Typical research studies rely on participant pools in social science departments, or online participants recruited through various panels, such as Mechanical Turk. These samples also have their own strengths and limitations, and the same holds true for a field study embedded within a class on VR. While our sample was heterogeneous in terms of race and previous VR use, it still reflects a convenience sample of college students and college students learning about the medium of VR, which makes them a very particular sample. It is possible that students learning about the medium could have served as a third variable explanation for our temporal effects. At the same time, we point out that it is critical to

allow students to grow accustomed and learn about the medium before we can investigate how response to VR changes over time and understand its full potential. The current study implemented novel strategies aimed at observing the robustness of how these effects hold over time. Future work should investigate how these effects hold over different contexts and populations.

In a similar vein, the current study utilized a stimulus sampling method to see how our effects can generalize across different types of environments. While stimulus sampling serves as one of the strengths for the robustness of the observed effects, in order to isolate and strengthen the causal argument of our manipulation, we suggest that future work explore the moderators of our variables of interest with a more narrowed lens.

Third, while the Oculus Quest 2 headsets and the ENGAGE platform were surprisingly robust compared to our previous experience with immersive VR technology, many sessions were lost due to technological error, and our final sample for both studies was slightly smaller than we had hoped. Our choice to focus on groups over time made this study unique for many reasons, but had its own costs, such as handling software updates to the platform that changed features of the avatars, or network issues that led to participants being unable to join. Furthermore, due to the nature of the study simultaneously being a course, there was a need for flexibility to accommodate participants' schedules. This included allowing participants to attend different discussion sessions when needed, which affected the members and size of the group across weeks. Moreover, our choice of using the ENGAGE platform was driven by its features to easily create content and record data. However, there are specific aspects of this platform which will likely not generalize to all platforms, which have unique affordances and overall qualities (Barreda-Ángeles & Hartmann, 2022).

Another limitation draws from the avatar design process. Although the selections were informed by previous research, the design was heavily limited by what options were available. Factors such as gender, which only came with binary options of female and male, or skin color, which we tried to keep as close to gray and racially ambiguous as possible, may have contributed to creating an avatar that, while uniform, gave off cues of a recognizable gender and race. Continuing this discussion of limitations brought on by avatars, in VR, a person's experience is presented from a first-person point of view. Consequently, people are unable to see their own selves. This raises the possibility of other cues overriding avatar perception. Although participants inevitably had to see their avatar in the customization page every time they were randomly assigned to the uniform avatar condition, this was only for half of the sessions. In future work, we hope to incorporate a mechanism in which participants are able to see their own selves to be reminded of what their avatars look like throughout their experience.

Lastly, the time variable was confounded with topic, in that the topics changed each week. While there was no pattern that dictated which topics were discussed early versus late (i.e., it was not the case that topics got more difficult or more technical over time), it is important to acknowledge that a better temporal manipulation would have had similar content over time or used a design that randomized topic over time across groups.

Future directions

There is a growing importance to understanding the social dynamics of how people use CVEs. Many questions remain unanswered on how the components of the TSI paradigm—self-representation, sensory capabilities, and contextual situations—shape people as they navigate the virtual world and form groups. We examined the transformation of avatar appearance by utilizing a uniform avatar in which all avatars looked the same. While we suggest that having a uniform avatar in a group setting may suppress individuals' visual cues and be unfavorable as it lowers self-presence and realism, it is possible that some degree of visual similarity, rather than complete similarity, may be advantageous for group-building. One avenue of research that demands further research is varying the degree of similarity or the number of similar cues shared amongst group members.

Similarly, there is research showing that there are other factors that may influence how people present their avatars, such as individual motivations (i.e., the individual is on the platform to be immersed in a virtual environment, or have social interactions, or achieve goals specific to the platform) or the functionality of the platform (i.e., the customization options available) (Harari et al., 2015). These differences can result in different ways of creating and expressing the self via avatars, which ultimately shape the type of avatar an individual uses to represent the self. In other words, the avatar people select to represent themselves may not be representative of their true self, but other versions of the self, such as an "ideal self" (Bessière et al., 2007; Ducheneaut et al., 2009).

Implications

The current study is one of the first large-scale, longitudinal field experiments to investigate how multiple sets of larger groups and social dynamics evolve over time in CVEs. From an experimental design standpoint, the study implemented a unique design that allowed for observations of behaviors in a naturalistic setting, rather than a controlled laboratory setting. From a statistical standpoint, we contribute to the field by using linear growth models to understand constructs and their changes across time, not in isolation, but in an interrelated way. We showed that choices of how avatars are created and scene size change nonverbal synchrony—a hallmark measure of the success of group interaction. Minor decisions made by metaverse designers will have psychological impacts on users.

In recent years, VR headsets and content have become more accessible to the general public. As there is interest in making a digital migration to the metaverse, there is a growing need to understand how transformations resulting from CVEs affect people's behaviors and attitudes, and how they should be taken into account when designing said platforms. In particular, as the metaverse is being used for purposes such as training, learning, and team building—which are often social activities that involve multiple individuals—what representation looks like is critical.

Currently, there is a wide breadth of research that has been conducted regarding how the two dimensions of interest in the TSI framework influence outcomes (e.g., for self-representations see work related to the Proteus Effect, Praetorius and Görlich, 2020; Ratan et al., 2020; Roth et al., 2018; for contextual situations see Bolouki, 2022; Lee et al., 2022; Nukarinen et al., 2022). The current work contributes to this breadth of literature from a theoretical standpoint by examining the effects of time and group interaction, as well as self-representation and contextual situations.

Transforming avatar appearance

Previous research has pointed out the gains to customizing one's avatar. Results show that avatar customizing and similarity to the self do indeed increase presence (Waltemate et al., 2018). However, what is unique in this study is the finding that uniform avatars provide greater enjoyment than self-avatars. Hence, depending on the goal of the platform, one should take these findings into account. For applications in which self-presence and realism are the goal—such as training—customizing is best. On the other hand, for recreation and social interaction, fostering visual similarity is recommended.

The results have implications for designers of such platforms on how avatars are presented and what options are made available. Previous research suggests that the way in which avatars are presented gives rise to differences in cues that are more useful or appropriate for different contexts. For instance, Dobre et al. (2022) report that in a work setting, realistic avatars and their nonverbal behavior are more appropriate compared to cartoon-like avatars. Moreover, Tanis and Postmes (2003) argue that a lack of cues in a communication partner may lead to ambiguity and uncertainty. However, in a different context, such as gaming, oftentimes simple cartoon-like avatars are used and have been shown to have a more positive impact and engagement (e.g., Monteiro et al., 2018). Depending on how many customizable options are made available, avatars can be altered to be made as individualized and as close to a tailored avatar as possible, or oppositely, be reduced to a limited number of options that result in avatars that are highly similar to one another. Beyond aesthetic goals of a platform and its avatars, designers should consider the goal of their platform, and adjust for how much control and customization people can have when creating their avatars.

Transforming context

Context is a term often used in theories and models in social science and human–computer interaction, but is difficult to explicate. Some studies actively manipulate context through means such as randomly assigning students to various classrooms. Such studies are often limited by cost, as it is expensive to physically build dozens of rooms that only vary on a single parameter. For example, Meyers-Levy and Zhu (2007) examined ceiling height by constructing two false ceilings to create four rooms that differed only on the Y-axis, and needed to employ professional engineers to build the ceilings for the study. Consequently, due to the cost involved, there are very few studies that look at more than a handful of different rooms. Moreover, in most studies, there are confounds in the variables of interest (e.g., bigger rooms also have different furnishings or light patterns than smaller rooms).

Another strategy is to observe people as they move about the world and measure how various behaviors differ based on their location. Recent work examining smartphones can look to see how locations, as tracked by smartphone GPS signals, influence social interaction and other behaviors (Matz & Harari, 2021). This approach allows for larger variance in locations but is limited to places where people happen to go to, as opposed to locations which are designed specifically to

meet some type of theoretical question. The current study is focused on VR, but also presents an enhanced understanding of how the structure of outdoor and indoor spaces—specifically how far a person can see on the X–Z plane—influences nonverbal behaviors and attitudes. Our stimulus sampling strategy of presenting 192 distinct locations makes this one of the most rigorous studies to ever examine the effects of location on psychological outcomes. In particular, researchers have rarely examined panoramic indoor spaces, as they are incredibly expensive to access in the real world.

Second, we found that the benefits of being in a spacious, panoramic environment found in previous research translate even inside virtual environments. In VR, space is free, and by holding an event within a spacious environment, a host would be able to foster a sense of perceived restorativeness, pleasure, arousal, presence (self and spatial), and enjoyment for participants. Such environments will also be beneficial for creating a sense of community, as indicated by greater synchrony and entitativity, which may be of interest for training, teaching, and team-building purposes.

However, it should be noted that environments that are both panoramic and outdoors may result in lower enjoyment and realism. One potential explanation for this draws from Appleton's prospect refuge theory (1975), which draws from theories on evolutionary survival instincts, and posits that people innately prefer environments that provide both opportunity and safety. The constrained nature of the outdoor environments fits with that framing, as a lack of access for shelter may induce threat and fear (Ulrich, 1983). Another potential reason for lowered enjoyment and realism pulls from qualitative observations made by discussion session leaders, in which several participants pointed out increased pixelation and lag in environments where there were more rendered trees (i.e., panoramic outdoors were filled with more elements of nature).

Groups

One question of interest within group interactions in VR relates to the size of the group. Creating a sense of community and fostering group cohesion is often a desired goal. While we did find that entitativity increased over time in Study 1, we did not find the same results in Study 2. In building our models, we examined how much variance was accounted for by the time-varying group size (i.e., repeated measures nested within individuals nested within the discussion session they attended that week) and found little variance at the group level. This raises the question: can a group be too big in CVEs? While there is research on the role of group size in efficacy and collaboration (e.g., Guimerà et al., 2005; Kerr, 1989), more research on the ideal or maximum size of group interactions in CVEs is required to draw any conclusions. We provide some suggestions that draw from qualitative observations.

First, one theme that emerged was the value of a backup communication channel. We expected a small but serious likelihood of technical challenges that would prevent participants from reaching out for assistance. For example, a headset may have low battery or lose Internet connection, the participant may fail to log in, or the multi-user service may fail altogether. It was necessary and very helpful to have a fallback medium. In our case, we had a Zoom video conferencing window open that was operated by a different instructor that was not leading the discussion session. While technical challenges are inevitable in such settings, they can be addressed with ease and swiftness in a smaller group, or with a few number of students to assist. The ideal size of a group may be limited by how much technical support and resources can be provided.

Second, as CVEs are currently structured, audio issues may arise when there are many people occupying the same virtual space. Unless spatialized audio is used, it is difficult to have multiple people speaking at once due to how audio is outputted. This leads to social cues that are unique to CVEs to indicate turn-taking. For instance, in ENGAGE, a participant will raise and twist their wrist to indicate that they are planning to unmute and talk. Similarly, every participant had their usernames and a microphone icon floating above their avatars that showed whether they were muted. In a typical interaction, participants muted their microphones to prevent background noise from the real world bleeding into the virtual conversation. In order to speak, participants would have to turn their heads and look around the room to see if anyone had the microphone unmuted to speak. As the group size grows larger, such social cues may become less salient and challenging to pick up.

Time

Studies that examine individual or group behavior over time in VR, in particular CVEs, are extremely rare. In the current study, we were particularly interested in the evolution of groups over time, and how this evolution interacted with self-representations and contextual situations. As VR users grow more comfortable using the medium and the novelty wears off, how do transformations of the self and context manifest in changes in attitudes? We report that, across both studies, there was an increase in self-presence and realism. One possibility is that, as participants grow accustomed to the virtual body and environment, they grow more comfortable and present in their avatar. With time and use, participants may have been able to focus more on being present and pay attention to their surroundings, rather than focus on learning how to use the medium. However, with comfort comes familiarity, and the novelty of the medium may have worn off. This potentially explains why there was a decrease in pleasure, arousal, and enjoyment.

In addition to learning about the evolution of virtual behavior over time, another finding emerged here: people change substantially over time, continually up to Week 8 in our studies. Even outcomes that were not obvious in hindsight—for example, our finding that scenes are perceived as more realistic over time—consistently change with more experience. If one simply looks at the first session, an inaccurate picture emerges. In some instances, the noise from looking at the first session masks important findings which emerge later. In this sense, studies are "temporally underpowered." More problematic are the instances in which the pattern that one sees during the first session is actually opposite to the patterns that consistently emerge over the majority of subsequent weeks, such as our finding on the effect of panoramic viewing on synchrony. Given that most published research in VR only looks at a single dose at one time point, it is critical for future work to spend the extra resources to ensure experimental effects are robust temporally.

Notes

1. However, technical limitations led to variance, such as each avatar was customized on the participant's end, participants had to switch between the self and uniform avatars between sessions, and how

the skin colors were rendered in individuals' headsets differed. Additionally, as lower torso, age, and weight were not rendered in the HMDs, no specific instructions were provided for these features.

2. The study also varied the type of onboarding exercise participants did when first entering VR. However, due to students arriving at different times, and the lack of adherence to the movement instructions, the variable failed manipulation checks and we do not report it given space constraints. The nature of the variable is further described in Appendix B.

3. In addition to the two variables related to context, we also attempted to manipulate the amount of translation—movement within the VR scene. However, given the nature of the collaboration tasks, there was not enough physical translation for this variable to show differences, it failed manipulation checks, and we do not report it given space constraints. The nature of the variable is further described in Appendix B.

4. Measures for entitativity, enjoyment, and realism were the same across Study 1 and Study 2. One less item was included in entitativity for Study 2. Cronbach's α was 0.89 for enjoyment and 0.86 for entitativity.

Data availability

The data underlying this article cannot be shared publicly due to the privacy of individuals that participated in the study. The data will be shared on reasonable request to the corresponding author.

Funding

Acknowledgements

References

Anderson, A. P., Mayer, M. D., Fellows, A. M., Cowan, D. R., Hegel, M. T., & Buckey, J. C. (2017). Relaxation with immersive natural scenes presented using virtual reality. *Aerospace Medicine and Human Performance*, 88(6), 520–526. https://doi.org/10.3357/amhp.4747.2017

Appleton, J. (1975). *The experience of landscape*. Wiley.

Aseeri, S., & Interrante, V. (2021). The influence of avatar representation on interpersonal communication in virtual social environments. *IEEE Transactions on Visualization and Computer Graphics*, 27(5), 2608–2617. https://doi.org/10.1109/tvcg.2021.3067783

Bailenson, J. N., Beall, A. C., Loomis, J., Blascovich, J., & Turk, M. (2004). Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence: Teleoperators and Virtual Environments*, 13(4), 428–441. https://doi.org/10.1162/1054746041944803

Bailenson, J. N., & Yee, N. (2005). Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science*, 16(10), 814–819. https://doi.org/10.1111/j.1467-9280.2005.01619.x

Bailenson, J. N., & Yee, N. (2006). A longitudinal study of task performance, head movements, subjective report, simulator sickness, and transformed social interaction in collaborative virtual environments. *Presence: Teleoperators and Virtual Environments*, 15(6), 699–716. https://doi.org/10.1162/pres.15.6.699

Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008a). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *Journal of the Learning Sciences*, 17(1), 102–141. https://doi.org/10.1080/10508400701793141

Bailenson, J. N., Yee, N., Blascovich, J., & Guadagno, R. E. (2008b). Transformed social interaction in mediated interpersonal communication. In Konijn E. A., Utz S., Tanis M., Barnes S. B. (Eds.) *Mediated Interpersonal Communication* (pp. 91–113). Routledge. http://dx.doi.org/10.4324/9780203926864-14

Barreda-Ángeles, M., & Hartmann, T. (2022). Psychological benefits of using social virtual reality platforms during the covid-19 pandemic: The role of social and spatial presence. *Computers in Human Behavior*, 127, 107047. https://doi.org/10.1016/j.chb.2021.107047

Barton, J., & Pretty, J. (2010). What is the best dose of nature and green exercise for improving mental health? A multi-study analysis. *Environmental Science & Technology*, 44(10), 3947–3955. https://doi.org/10.1021/es903183r

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Berto, R. (2005). Exposure to restorative environments helps restore attentional capacity. *Journal of Environmental Psychology*, 25(3), 249–259. https://doi.org/10.1016/j.jenvp.2005.07.001

Bessière, K., Seay, A. F., & Kiesler, S. (2007). The Ideal Elf: Identity exploration in world of warcraft. *CyberPsychology & Behavior*, 10(4), 530–535. https://doi.org/10.1089/cpb.2007.9994

Blascovich, J. (2002). Social influence within immersive virtual environments. In R. Schroeder (Ed.), *The social life of avatars. Computer supported cooperative work*. Springer. pp. 127–145. https://doi.org/10.1007/978-1-4471-0277-9_8

Bolouki, A. (2022). The impact of virtual reality natural and built environments on affective responses: A systematic review and meta-analysis. *International Journal of Environmental Health Research*, 1–17. https://doi.org/10.1080/09603123.2022.2130881

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Bratman, G. N., Hamilton, J. P., Hahn, K. S., Daily, G. C., & Gross, J. J. (2015). Nature experience reduces rumination and subgenual prefrontal cortex activation. *Proceedings of the National Academy of Sciences*, 112(28), 8567–8572. https://doi.org/10.1073/pnas.1510459112

Brinberg, M., Vanderbilt, R. R., Solomon, D. H., Brinberg, D., & Ram, N. (2021). Using technology to unobtrusively observe relationship development. *Journal of Social and Personal Relationships*, 38(12), 3429–3450. https://doi.org/10.1177/02654075211028654

Browning, M. H. E. M., Shipley, N., McAnirlin, O., Becker, D., Yu, C.-P., Hartig, T., & Dzhambov, A. M. (2020). An actual natural setting improves mood better than its virtual counterpart: A meta-analysis of experimental data. *Frontiers in Psychology*, 11. https://doi.org/10.3389/fpsyg.2020.02200

Campbell, D. T. (1958). Common fate, similarity and other indices of the status of aggregates of persons as social entities. In Willner D. (Ed.) *Decisions, values and groups* (pp. 185–201). Elsevier. http://dx.doi.org/10.1016/b978-0-08-009237-9.50017-2

Condon, W. S., & Ogston, W. D. (1966). Sound film analysis of normal and pathological behavior patterns. *The Journal of Nervous and Mental Disease*, 143(4), 338–347. https://doi.org/10.1097/00005053-196610000-00005

Conner, T. S., & Lehman, B. (2012). Getting started: Launching a study in daily life. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 89–107). Guilford Press. http://dx.doi.org/10.5860/choice.50-1159

Dale R., Bryant G A., Manson J. H. & Gervais M. M. Body synchrony in triadic interaction. *Royal Society Open Science*. 2020. 7 (9) 200095. https://doi.org/10.1098/rsos.200095

Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., & Cohen, D. (2012). Interpersonal synchrony: A survey of evaluation methods across disciplines. *IEEE Transactions on Affective Computing*, 3(3), 349–365. https://doi.org/10.1109/t-affc.2012.12

Dobre, G. C., Wilczkowiak, M., Gillies, M., Pan, X., & Rintel, S. (2022, April 27). *Nice is different than good: Longitudinal communicative effects of realistic and cartoon avatars in real mixed reality work meetings*. CHI Conference on Human Factors in Computing Systems Extended Abstracts. http://dx.doi.org/10.1145/3491101.3519628

Ducheneaut, N., Wen, M.-H., Yee, N., & Wadley, G. (2009, April 4). *Body and mind*. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. http://dx.doi.org/10.1145/1518701.1518877

Eisinga, R., te Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, 58(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3

Forsyth, D. R. (1990). *Group dynamics*. Wadsworth Publishing Company.

Gonzalez-Franco, M., & Peck, T. C. (2018). Avatar embodiment. Towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5, 74. https://doi.org/10.3389/frobt.2018.00074

Grimm, K. J., Ram, N., & Estabrook, R. (2016). *Growth modeling: Structural equation and multilevel modeling approaches*. Guilford Publications.

Guimerà R., Uzzi B., Spiro J. & Amaral L. A. N. Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance. *Science*. 2005. 308 (5722) 697–702. https://doi.org/10.1126/science.1106340

Han, E., Miller, M. R., Ram, N., Nowak, K. L., & Bailenson, J. N. (2022, May). *Understanding group behavior in virtual reality: A large-scale, longitudinal study in the metaverse*. 72nd Annual International Communication Association Conference. https://ssrn.com/abstract=4110154

Harari, G. M., Graham, L. T., & Gosling, S. D. (2015). Personality impressions of World of Warcraft players based on their avatars and usernames. *International Journal of Gaming and Computer-Mediated Simulations*, 7(1), 58–73. https://doi.org/10.4018/ijgcms.2015010104

Harasty, A. S. (1996). *Perceiving groups as entities: The role of "entitativity" for impression formation processes and stereotype use*. The Ohio State University, ProQuest Dissertations Publishing.

Hartig, T., Korpela, K., & Evans, G. W. (1996). *Validation of the measure of perceived environmental restorativeness*. University of Göteborg, Department of Psychology.

Hasenbein, L., Stark, P., Trautwein, U., Queiroz, A. C. M., Bailenson, J., Hahn, J.-U., & Göllner, R. (2022). Learning with simulated virtual classmates: Effects of social-related configurations on students' visual attention and learning experiences in an immersive virtual reality classroom. *Computers in Human Behavior*, 133, 107282. https://doi.org/10.1016/j.chb.2022.107282

Herrera, F., Oh, S. Y., & Bailenson, J. N. (2020). Effect of behavioral realism on social interactions inside collaborative virtual environments. *Presence: Teleoperators and Virtual Environments*, 27(2), 163–182. https://doi.org/10.1162/pres_a_00324

Janis, I. L. (1973). Groupthink and group dynamics: A social psychological analysis of defective policy decisions. *Policy Studies Journal*, 2(1), 19–25. https://doi.org/10.1111/j.1541-0072.1973.tb00117.x

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, 103(1), 54–69. https://doi.org/10.1037/a0028347

Khojasteh, N., & Won, A. S. (2021). Working together on diverse tasks: A longitudinal study on individual workload, presence and emotional recognition in collaborative virtual environments. *Frontiers in Virtual Reality*, 2. https://doi.org/10.3389/frvir.2021.643331

Kim, D., Kim, J., Shin, J., Yoon, B., Lee, J., & Woontack W. (2022, January 12). *Effects of virtual room size and objects on relative translation gain thresholds in redirected walking*. ArXiv.Org. https://arxiv.org/abs/2201.04273

Kim, J. (2009). "I want to be different from others in cyberspace" The role of visual similarity in virtual group identity. *Computers in Human Behavior*, 25(1), 88–95. https://doi.org/10.1016/j.chb.2008.06.008

LaFrance, M. (1979). Nonverbal synchrony and rapport: Analysis by the cross-lag panel technique. *Social Psychology Quarterly*, 42(1), 66. https://doi.org/10.2307/3033875

Lanier, M., Waddell, T. F., Elson, M., Tamul, D., Ivory, J. D., & Przybylski, A. K. (2019). *Virtual reality check: Statistical power, reported results, and the validity of research on the psychology of virtual reality and immersive environments*. Center for Open Science. http://dx.doi.org/10.31234/osf.io/6hk89

Leach, C. W., van Zomeren, M., Zebel, S., Vliek, M. L. W., Pennekamp, S. F., Doosje, B., Ouwerkerk, J. W., & Spears, R. (2008). Group-level self-definition and self-investment: A hierarchical (multi-component) model of in-group identification. *Journal of Personality and Social Psychology*, 95(1), 144–165. https://doi.org/10.1037/0022-3514.95.1.144

Lederbogen, F., Kirsch, P., Haddad, L., Streit, F., Tost, H., Schuch, P., Wüst, S., Pruessner, J. C., Rietschel, M., Deuschle, M., & Meyer-Lindenberg, A. (2011). City living and urban upbringing affect neural social stress processing in humans. *Nature*, 474(7352), 498–501. https://doi.org/10.1038/nature10190

Lee, E.-J. (2006). Effects of visual representation on social influence in computer-mediated communication. *Human Communication Research*, 30(2), 234–259. https://doi.org/10.1111/j.1468-2958.2004.tb00732.x

Lee, M., Kim, E., Choe, J., Choi, S., Ha, S., & Kim, G. (2022). Psychological effects of green experiences in a virtual environment: A systematic review. *Forests*, 13(10), 1625. https://doi.org/10.3390/f13101625

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2). https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Loomis, J. M. (1992). Distal attribution and presence. *Presence: Teleoperators and Virtual Environments*, 1(1), 113–119. https://doi.org/10.1162/pres.1992.1.1.113

MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous race face effect. *Psychology, Public Policy, and Law*, 7(1), 98–118. https://doi.org/10.1037/1076-8971.7.1.98

Mael, F., & Ashforth, B. E. (1992). Alumni and their alma mater: A partial test of the reformulated model of organizational identification. *Journal of Organizational Behavior*, 13(2), 103–123. https://doi.org/10.1002/job.4030130202

Matz, S. C., & Harari, G. M. (2021). Personality–place transactions: Mapping the relationships between Big Five personality traits, states, and daily places. *Journal of Personality and Social Psychology*, 120(5), 1367–1385. https://doi.org/10.1037/pspp0000297

McCall, C., Bunyan, D. P., Bailenson, J. N., Blascovich, J., & Beall, A. C. (2009). Leveraging collaborative virtual environment technology for inter-population research on persuasion in a classroom setting. *Presence: Teleoperators and Virtual Environments*, 18(5), 361–369. https://doi.org/10.1162/pres.18.5.361

Meyers-Levy, J., & Zhu, R. (2007). The influence of ceiling height: The effect of priming on the type of processing that people use. *Journal of Consumer Research*, 34(2), 174–186. https://doi.org/10.1086/519146

Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, 67(4), 371–378. https://doi.org/10.1037/h0040525

Miller, M. R., Sonalkar, N., Mabogunje, A., Leifer, L., & Bailenson, J. (2021). Synchrony within triads using virtual reality. *Proceedings of*

the ACM on Human-Computer Interaction, *5*(CSCW2), 1–27. https://doi.org/10.1145/3479544

Monteiro, D., Liang, H.-N., Wang, J., Wang, L., Wang, X., & Yue, Y. (2018, December). *Evaluating the effects of a cartoon-like character with emotions on users' behaviour within virtual reality environments*. 2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). http://dx.doi.org/10.1109/aivr.2018.00053

Moustafa, F., & Steed, A. (2018, November 28). *A longitudinal study of small group interaction in social virtual reality*. Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology. http://dx.doi.org/10.1145/3281505.3281527

Newman, B. M., & Newman, P. R. (2020). Dynamic systems theory. In Newman, B. M., & Newman, P. R. (Eds.) *Theories of Adolescent Development* (pp. 77–112). Elsevier. http://dx.doi.org/10.1016/b978-0-12-815450-2.00004-8

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *12*(5), 481–494. https://doi.org/10.1162/105474603322761289

Nowak, K. L., Hamilton, M. A., & Hammond, C. C. (2009). The effect of image features on judgments of homophily, credibility, and intention to use as avatars in future interactions. *Media Psychology*, *12*(1), 50–76. https://doi.org/10.1080/15213260802669433

Nukarinen, T., Rantala, J., Korpela, K., Browning, M. H. E. M., Istance, H. O., Surakka, V., & Raisamo, R. (2022). Measures and modalities in restorative virtual natural environments: An integrative narrative review. *Computers in Human Behavior*, *126*, 107008. https://doi.org/10.1016/j.chb.2021.107008

Oh, C., Herrera, F., & Bailenson, J. (2019). The effects of immersion and real-world distractions on virtual social interactions. *Cyberpsychology, Behavior, and Social Networking*, *22*(6), 365–372. https://doi.org/10.1089/cyber.2018.0404

Oh Kruzic, C., Kruzic, D., Herrera, F., & Bailenson, J. (2020). Facial expressions contribute more than body movements to conversational outcomes in avatar-mediated virtual environments. *Scientific Reports*, *10*(1), 1–23. https://doi.org/10.1038/s41598-020-76672-4

Okken, V., van Rompay, T., & Pruyn, A. (2012). Room to move. *Environment and Behavior*, *45*(6), 737–760. https://doi.org/10.1177/0013916512444780

Praetorius, A. S., & Görlich, D. (2020, September 15). *How avatars influence user behavior*. International Conference on the Foundations of Digital Games. http://dx.doi.org/10.1145/3402942.3403019

Ratan, R., Beyea, D., Li, B. J., & Graciano, L. (2020). Avatar characteristics induce users' behavioral conformity with small-to-medium effect sizes: A meta-analysis of the proteus effect. *Media Psychology*, *23*(5), 651–675. https://doi.org/10.1080/15213269.2019.1623698

Roth, D., Klelnbeck, C., Feigl, T., Mutschler, C., & Latoschik, M. E. (2018, March). *Beyond replication: Augmenting social behaviors in multi-user virtual realities*. 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). http://dx.doi.org/10.1109/vr.2018.8447550

Rydell, R. J., & McConnell, A. R. (2005). Perceptions of entitativity and attitude change. *Personality and Social Psychology Bulletin*, *31*(1), 99–110. https://doi.org/10.1177/0146167204271316

Schoenherr, D., Paulick, J., Worrack, S., Strauss, B. M., Rubel, J. A., Schwartz, B., Deisenhofer, A.-K., Lutz, W., Stangier, U., & Altmann, U. (2018). Quantification of nonverbal synchrony using linear time series analysis methods: Lack of convergent validity and evidence for facets of synchrony. *Behavior Research Methods*, *51*(1), 361–383. https://doi.org/10.3758/s13428-018-1139-z

Sherif, M. (1937). An experimental approach to the study of attitudes. *Sociometry*, *1*(1/2), 90. https://doi.org/10.2307/2785261

Simmons, J., Nelson, L., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *PsycEXTRA Dataset*. https://doi.org/10.1037/e519702015-014

Sun, Y., Shaikh, O., & Won, A. S. (2019). Nonverbal synchrony in virtual reality. *PLoS One*, *14*(9), e0221803. https://doi.org/10.1371/journal.pone.0221803

Szolin, K., Kuss, D. J., Nuyens, F. M., & Griffiths, M. D. (2022). Exploring the user-avatar relationship in videogames: A systematic review of the Proteus effect. *Human–Computer Interaction*, *37*(6), 1–26. https://doi.org/10.1080/07370024.2022.2103419

Tanis, M., & Postmes, T. (2003). Social cues and impression formation in CMC. *Journal of Communication*, *53*(4), 676–693. https://doi.org/10.1111/j.1460-2466.2003.tb02917.x

Tarr, B., Slater, M., & Cohen, E. (2018). Synchrony and social connection in immersive virtual reality. *Scientific Reports*, *8*(1), 1–8. https://doi.org/10.1038/s41598-018-21765-4

Ulrich, R. S. (1983). Aesthetic and affective response to natural environment. In I. Altman & J. F. Wohlwill (Eds.), *Behavior and the Natural Environment. Human Behavior and Environment* (pp. 85–125). Springer. https://doi.org/10.1007/978-1-4613-3539-9_4

Ulrich, R. S., Simons, R. F., Losito, B. D., Fiorito, E., Miles, M. A., & Zelson, M. (1991). Stress recovery during exposure to natural and urban environments. *Journal of Environmental Psychology*, *11*(3), 201–230. https://doi.org/10.1016/s0272-4944(05)80184-7

van den Berg, A. E., Jorgensen, A., & Wilson, E. R. (2014). Evaluating restoration in urban green spaces: Does setting type make a difference? *Landscape and Urban Planning*, *127*, 173–181. https://doi.org/10.1016/j.landurbplan.2014.04.012

Waltemate, T., Gall, D., Roth, D., Botsch, M., & Latoschik, M. E. (2018). The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Transactions on Visualization and Computer Graphics*, *24*(4), 1643–1652. https://doi.org/10.1109/tvcg.2018.2794629

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5), 2020–2045. https://doi.org/10.1037/xge0000014

White, M. P., Alcock, I., Wheeler, B. W., & Depledge, M. H. (2013). Would you be happier living in a greener urban area? A fixed-effects analysis of panel data. *Psychological Science*, *24*(6), 920–928. https://doi.org/10.1177/0956797612464659

Won, A. S., Bailenson, J. N., Stathatos, S. C., & Dai, W. (2014). Automatically detected nonverbal behavior predicts creativity in collaborating dyads. *Journal of Nonverbal Behavior*, *38*(3), 389–408. https://doi.org/10.1007/s10919-014-0186-0

Wu, X., Law, S., Heath, T., & Borsi, K. (2017, July 7). *Spatial configuration shapes student social and informal learning activities in educational complexes*. Proceedings - 11th International Space Syntax Symposium. https://discovery.ucl.ac.uk/id/eprint/10107148/

Yee, N., Ducheneaut, N., Yao, M., & Nelson, L. (2011, May 7). *Do men heal more when in drag?* Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. http://dx.doi.org/10.1145/1978942.1979054

Zhang, M., Beetle, C., Kelso, J. A. S., & Tognoli, E. (2019). Connecting empirical phenomena and theoretical models of biological coordination across scales. *Journal of The Royal Society Interface*, *16*(157), 20190360. https://doi.org/10.1098/rsif.2019.0360

Appendix A: Uniform avatar pre-test

The following questions were asked about five different avatars. Based on the results, different features (e.g., technically female or male and skin color) were selected from each avatar option. The features that were closest to neutral in terms of perceived gender, race, and comfort in representation were selected.

Gender perception

Please answer the following about this avatar's gender presentation (7-point Likert scale, $1 =$ Not at all, $4 =$ Neutral, $7 =$ Very much).

1) This avatar is feminine
2) This avatar is masculine
3) I am easily able to identify this avatar's gender

Racial perception

Please rate the degree to which the avatar fits in the following racial categories (5-point Likert scale, $1 =$ Extremely, $5 =$ Not at all).

1) African, African-American, or Black
2) Asian or Asian-American
3) Hispanic or LatinX
4) Indigenous/Native American, Alaska Native, First Nations
5) Middle Eastern
6) Native Hawaiian or other Pacific Island
7) White
8) More than one race

Representation comfort

Please rate how you would feel if this avatar were to visually represent you in a virtual environment (7-point Likert scale, $1 =$ Extremely uncomfortable, $4 =$ Neither uncomfortable nor comfortable, $7 =$ Extremely comfortable).

Table A1. Avatar pre-test means and standard deviations (in parentheses) of each avatar

		Avatar 1 (Female)	Avatar 2 (Female)	Avatar 3 (Male)	Avatar 4 (Female)	Avatar 5 (Male)
Gender perception	Feminine	4.48 (1.05)	3.81 (1.33)	1.52 (0.80)	3.30 (1.59)	1.70 (0.99)
	Masculine	3.59 (0.93)	4.15 (1.23)	6.11 (1.05)	4.81 (1.57)	5.96 (1.13)
	Identifiability	3.07 (1.66)	3.22 (1.72)	5.70 (1.44)	4.30 (1.79)	5.48 (1.72)
Racial perception	African, African-American, or Black	4.58 (0.64)	3.56 (1.01)	4.22 (0.93)	1.88 (0.71)	4.58 (0.76)
	Asian or Asian-American	3.78 (1.05)	3.89 (1.01)	3.70 (1.17)	4.44 (0.82)	4.04 (1.11)
	Hispanic or LatinX	4.00 (0.68)	3.44 (0.80)	3.89 (1.01)	3.63 (0.93)	3.88 (0.99)
	Indigenous/Native American, Alaska Native, First Nations	4.27 (0.67)	3.78 (0.97)	3.88 (0.95)	3.81 (0.80)	4.13 (0.90)
	Middle Eastern	4.42 (0.81)	3.72 (0.84)	3.88 (0.86)	3.85 (1.01)	4.04 (0.87)
	Native Hawaiian or other Pacific Island	4.31 (0.79)	3.88 (0.95)	3.92 (0.84)	4.31 (0.74)	4.16 (0.8)
	White	2.58 (1.10)	3.85 (0.99)	3.15 (1.38)	4.78 (0.64)	2.76 (1.27)
	More than one race	3.00 (1.04)	2.74 (0.98)	3.15 (1.17)	3.17 (1.05)	3.48 (1.12)
Representation comfort		4.19 (1.47)	4.19 (1.44)	2.74 (1.38)	3.26 (1.61)	2.96 (1.31)

Sample avatars
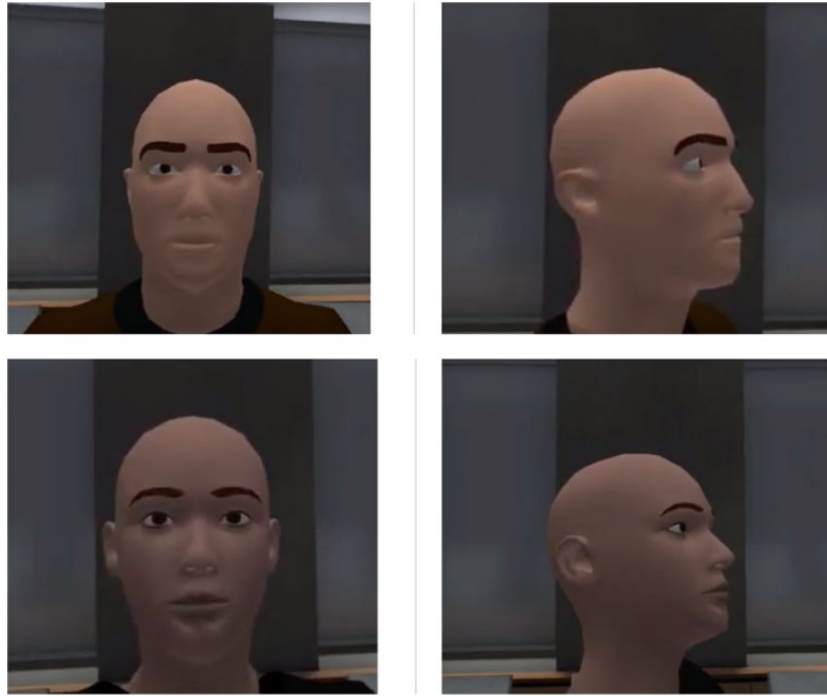
Appendix B: Synchrony measurement

Motion synchrony

Due to the long history and various technologies used to capture motion synchrony, there is a proliferation of methods to calculate it (Delaherche et al., 2012; Schoenherr et al., 2018, "Quantification of Nonverbal Synchrony"). While this range of methods may demonstrate the robustness of motion synchrony, it also may lead to researchers having many degrees of freedom to select a favorable outcome, reducing the trustworthiness of a reported result (Simmons et al., 2011). In this study, we pre-registered our measure of motion synchrony (pre-registration at https://osf.io/3c4aj/) as Spearman correlation of head speed over the whole time in an interaction. This pathway is based on the previous methods used to detect synchrony in VR (Miller et al., 2021; Sun et al., 2019).

One unique aspect in this dataset relative to the previous two is whether to separate out different variations of virtual motion. Due to the recording capabilities of ENGAGE, position data are separated into *physical* motion tracked by the headset and *abstract* motion produced by the interface (e.g., teleporting, joystick motion). These two feeds provide three options for defining head motion: only considering physical motion, only considering abstract motion, or considering motion as visible to the other participants in the environment, which we term *visible* motion. In the application we used, visible motion is the vector sum of abstract motion and a rotation of physical motion. We selected visible motion in this analysis due to the prevailing theory of synchrony as a response or anticipation to what is perceived.

There is a concern that the large magnitude of virtual motions with teleporting will influence results. We believe this is effectively addressed by the use of Spearman correlation rather than Pearson correlation.

We did encounter one issue with motion that required a change to the measure of synchrony. Due to the recording software, minor variations in the duration of time between samples of data can cause the speed of all participants in the frame to fluctuate in the same direction together. This would naturally artificially inflate synchrony. However, this effect only inflates samples that match perfectly—as soon as there is at least one frame of offset, the dependence is broken. Because of this artifact, we opted to use the average of synchrony with an offset of ± 2.5 s, removing the value where the offset is zero. In relation to the figure, only the portion between -2.5 s and $+2.5$ s, with the zero point removed, was included in the measurement of synchrony. The zero point as well as all values from -60 s to -2.5 s and from $+2.5$ s to $+60$ s were not included.

This process produces synchrony values for each pair of participants in a section. It is unclear at what level synchrony occurs (person, pair, full group) as the dominant trend in synchrony research is a single participant (see Miller et al., 2021 and Zhang et al., 2019 for more information). For symmetry with the other analyses, we chose to work with synchrony where the unit of analysis is the person-per-session level, indicating what might be termed a "tendency to synchronize." The calculation of this value is simply the average of synchrony scores in all pairs that contain the participant in question.