COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 50 (2026) e70211

ISSN: 1551-6709 online
DOI: 10.1111/cogs.70211

Synchrony and Task Engagement in Virtual Reality: Temporal Dynamics, Predictors, and Psychological Outcomes of Collaborative Behaviors

Portia Wang,[a] Monique Santoso,[a] Eugy Han,[b] Jeremy N. Bailenson[a]

[a]*Department of Communication, Stanford University*
[b]*Department of Media Production, Management, and Technology, University of Florida*

Received 1 August 2025; received in revised form 5 February 2026; accepted 3 April 2026

Abstract

Collaborative behaviors provide useful signals for understanding how minds align through perception and actions. Virtual reality (VR) is a useful tool for studying these behaviors, as it enables fine-grained measurements of coordination in virtual social settings. In this work, we investigate collaborative behaviors in a large-scale classroom VR dataset of space-building activities ($N = 146$), focusing on dyadic synchrony and individual task engagement during the collaborative group activity. An analysis of collaborative behaviors over time revealed a U-shaped pattern in head and hand synchrony, with a turning point occurring approximately two-thirds into the activity. We found that the likelihood of dyads temporally aligning their object editing behaviors (i.e., nonzero vs. zero synchrony scores) and whether they actively created, edited, or deleted objects all followed an inverted-U shape over time, peaking around midway through the activity. We further analyzed synchrony and task engagement both as possible indicators of individual dispositions (i.e., previous extended reality [XR] and design experiences) and social context (i.e., group size), and also as behavioral signals for how individuals perceive their group members and collaborative outcomes. The findings revealed that collaborative behaviors such as object edit synchrony are shaped by previous XR experience, group size negatively predicted the frequency of object creation, and that the frequency of object deletion is positively associated with perception of group closeness. Taken together, this work advances the understanding of collaborative behavior by modeling its temporal dynamics, identifying predictors and psychological outcomes, thereby demonstrating how VR enables large-scale examination of its cognitive underpinnings.

Keywords: Behavioral tracking; Collaboration; Computer-mediated communication; Nonverbal behavior; Social interaction; Virtual reality

Correspondence should be sent to Portia Wang, Department of Communication, Stanford University, 450 Jane Stanford Way Building 120, Stanford, CA 94305, USA. E-mail: portiaw@stanford.edu

1. Introduction

Collaborative behaviors are tightly woven into the fabric of daily life (Montoya, Massey, & Lockwood, 2011). Whether it is passing along dishes across a table or assembling furniture with friends, social settings often require individuals to anticipate, respond to, and coordinate their actions with those of others. How we interact with others in collaborative settings and how these interactions are shaped by the social and spatial context offer meaningful signals to cognitive processes (Knoblich & Sebanz, 2006). These same collaborative behaviors are increasingly present in virtual environments, ranging from nonimmersive platforms such as Second Life (Andreas, Tsiatsos, Terzidou, & Pomportsis, 2010) to immersive ones like VRChat and Rec Room (Freeman, Acena, McNeese, & Schulenberg, 2022). Social virtual reality (VR), in particular, networks multiple users into the same virtual environment, allowing them to interact with their surroundings and one another. As a result, users regarded these applications as social hubs where they can build and create with friends (Maloney, Freeman, & Robb, 2021), educators have leveraged them to facilitate classroom activities (Han et al., 2023), and practitioners have utilized them to foster coexploration (Hong, Jeong, Kalay, Jung, & Lee, 2016).

VR also offers a rich tool for studying collaborative behaviors. In VR, users interact through shared virtual objects and embodied avatars that move in direct correspondence to their physical movements. As a result, VR interactions closely approximate face-to-face ones (Maloney, Freeman, & Wohn, 2020; Smith & Neff, 2018), suggesting that empirical insights on immersive collaborative behaviors may likely extend to real-world scenarios. The medium further affords fine-grained behavioral tracking of head and hand motion not easily accessible in face-to-face interactions, allowing for automatic extraction of behavioral markers such as visual attention and motion synchrony at scale (Han et al., 2023; Wang, Han, Queiroz, DeVeaux, & Bailenson, 2025). As such, social VR is spatially and socially immersive yet experimentally controllable, which offers a combination of ecological validity and experimental precision (Blascovich et al., 2002).

The medium's ability to extract group behaviors at scale motivated the present work. We examined a large-scale VR classroom dataset involving 146 students, organized into groups of two to seven, participating in a collaborative space-building activity. This work offers three main contributions. First, we outline and define immersive collaborative behaviors that can be quantitatively and automatically extracted from behavioral tracking data, specifically categorizing behaviors into dyadic synchrony and individual task engagement. Our second contribution is an understanding of how these behaviors change over time. The final contribution is an examination of the role of collaborative behaviors as reflections of individual differences and social contexts, as well as predictors of psychological outcomes. Accordingly, the present work is guided by the following research questions:

- **Research Question 1**: How do collaborative behaviors unfold in immersive environments? Specifically, does synchrony occur in collaborative space-building activities (a), and if so, how does synchrony change over time (b)? Furthermore, how does individual task engagement change over time (c)?

- **Research Question 2**: How are individual dispositions (i.e., prior extended reality [XR] and design experiences) and contextual differences (i.e., group size) related to collaborative behaviors in immersive environments?
- **Research Question 3**: How are collaborative behaviors related to the perception of group closeness and task outcomes in immersive environments?

By investigating these questions, we advance the theoretical understanding of how collaborative behaviors unfold in immersive environments and offer insights on how they meaningfully reflect group dynamics, context, and individual differences.

1.1. Collaborative behaviors through the lens of distributed cognition

One helpful theory for understanding collaborative behavior is distributed cognition (Hutchins, 2000), which posits that cognition is not only sustained within one's mind, but rather distributed and shaped by external structure such as one's environment and other members of a social group. Distributed cognition highlights three forms of distribution. First, the theory argues that cognitive processes can be distributed across social groups. That is, social actors in the same environment, potentially working on the same task, can serve as support to one's cognitive process. The second form of distribution notes that cognition is distributed across an individual's internal thinking and external structures, such as their environments and tools. Finally, cognition can also be distributed over time, with earlier actions and artifacts shaping later cognitive and behavioral trajectories.

The theory offers several guiding principles for how we should theorize and study collaborative behaviors in virtual environments. To start, the theory's emphasis that cognition is distributed over time highlights the need to study how behaviors change over time, as it can help to better understand how prior actions and intermediate task artifacts (e.g., a half-created prototype) contribute to the unfolding of later behaviors. Furthermore, the view that cognition is supported and shaped by social others, as well as the external environment and tools, suggests a need to examine how individuals' behaviors are coordinated with those around them and how they interact with the virtual environment and immersive platform more broadly.

Guided by these perspectives, we examined collaboration behaviors within a VR dataset in which groups collaboratively built out a virtual space. This virtual space acted as both the environment housing the activity, and as an evolving external structure that shapes and is shaped by individuals' actions. Leveraging VR and the tracking data it affords, and through the lens of distributed cognition, we examined and interpreted how individuals coordinated their behaviors with social others (e.g., synchrony) and their interactions with the virtual environment (e.g., creating, editing, and deleting objects).

1.2. Collaborative behaviors in virtual environments

Likely given the granular tracking data afforded by VR, research has focused on synchrony, defined as the temporal alignment of behavior and typically operationalized as the correlation between users' actions or motion. Synchrony offers a helpful measure to understand social dynamics. For example, Tarr, Slater, and Cohen (2018) investigated triad synchrony through a joint movement activity and found that participants who were paired with virtual avatars exhibiting higher levels of synchrony reported greater self-other overlap and social closeness. This finding is echoed by Sun, Shaikh, and Won (2019), who showed that head synchrony positively correlated with social closeness. This work also highlights nuances of synchrony, as this relationship between head synchrony and social closeness was not observed for hand or total synchrony. Synchrony can also be influenced by virtual context and social roles. In Miller, Sonalkar, Mabogunje, Leifer, and Bailenson (2021), researchers investigated head synchrony in triads who completed design tasks and found that participants exhibited higher head synchrony in a conference room compared to a garage, and that speaking roles predicted synchrony.

Prior work has examined task-specific behaviors in collaborative tasks. Montoya et al. (2011) examined the link between collaborative behaviors in dyads and triads and task performance during puzzle solving using Second Life, a desktop virtual environment platform. By extracting communication frequency and contributions (e.g., number of correctly positioned puzzle pieces), the authors clustered participant behaviors into two strategies, which differentially predicted performance measures such as the number of puzzles solved. More recently, Wang, Miller, Han, DeVeaux, and Bailenson (2024) proposed a bottom-up approach for describing how groups collaborate on 3D design in social VR and found that varying design behaviors differentially predicted final design characteristics.

We build on these works in several ways. First, unlike works that aggregated behaviors at the session level, we analyzed how immersive collaborative behaviors change over time. Additionally, while existing VR research on synchrony has focused on fixed-size interactions between dyads and triads (e.g., Miller et al., 2021, Sun et al., 2019), we considered synchrony and task engagement behaviors in groups with two to seven members. Finally, task-specific actions are fundamental to collaborative interactions, though prior works have largely focused on temporal alignment of head and hand movement during conversational and movement exercises (e.g., Sun et al., 2019; Tarr et al., 2018). Consequently, we analyze synchrony of task-related actions using a VR dataset of students engaging in an action-rich space-building task rather than the simple movement and verbal protocols used in prior studies.

1.3. Changes in collaborative behavior over time

In line with distributed cognition's emphasis on the temporal unfolding of cognition, scholars have studied how social interactions change over time. This research has considered language, body motion, and actions, typically extracting them by having human coders label behaviors using video recordings (Louwerse, Dale, Bard, & Jeuniaux, 2012; Zheng & Tversky, 2024). For example, Duran, Paige, and D'Mello (2024) analyzed language alignment in triads collaborating on a creative thinking task over the video-conferencing platform Zoom and found that lexical and syntactic alignment decreased over time, while semantic alignment increased. The authors interpreted these findings by suggesting a tradeoff in language alignment as participants converge on shared themes and topics over time. Changes in social behaviors are observed in gestures, with Louwerse et al. (2012) showing that nodding recurrences between dyads increased over time, which the authors believe reflected the mutual positive influence between synchronization and social affiliation. Zheng and Tversky (2024) studied how dependent joint actions occurred over time in a dyadic face-to-face task of assembling a TV cart. The authors found that over time, coordination efforts transitioned from explicit forms (i.e., speech, gesture) to implicit ones (i.e., actions that advance the task and communicate next steps), suggesting that actions alone were sufficient in establishing a common goal representation in latter parts of the task.

These findings illustrate how temporal changes in collaborative behaviors signal social and cognitive processes. That said, there are reasons to expect that collaborative behaviors during our VR space-building activity may unfold differently from past research. From a distributed cognition perspective, space-building places a high emphasis on how external environments could support collaborative thinking and behaviors. As objects begin to populate virtual spaces, the emerging environment can reduce the need for continuous monitoring of one another, as it can become an effective external scaffold that guides decisions. This shift can lessen behavioral alignment (i.e., synchrony), which Louwerse et al. (2012) argue helps partners narrow possible actions and reduce uncertainty and cognitive load. Furthermore, the temporal cadence in immersive world-building activities may differ from furniture assembly tasks and problem-solving games (Duran et al., 2024; Zheng & Tversky, 2024). While those tasks often unfold linearly (e.g., following step-by-step guidance), space-building can involve early exploration followed by focused construction. This can produce different temporal patterns of task-related actions, such as fluctuations in the frequency of object creation and editing. Space-building could also entail working with different object types that vary in the effort required to manipulate them, and allow for nonsequential actions and parallel work. Collectively, these features may uniquely influence collaborative temporal dynamics, motivating **RQ1**.

1.4. Internal dispositions and context shaping group behaviors

A long-standing interest in the field of cognitive science is how internal dispositions and contextual factors shape behavior and cognition (Duffy, Feist, & McCarthy, 2014; Griffin, Guillette, & Healy, 2015; Roberts & Yoon, 2022). In particular, one's prior experience with relevant technologies or tasks can meaningfully influence interpersonal dynamics and performance. For example, Montoya et al. (2011) found that participants with more prior experience with Second Life and higher frequency in using it were more likely to fall within the cluster of problem-solving strategies that yielded higher task performance. In VR, previous familiarity with the medium also predicts higher embodiment (Lehikko, Nykänen, & Ruokamo, 2025) and self-assessment of task performance (Sagnier, Loup-Escande, & Valléry, 2020), which could also shape downstream social behaviors. Prior domain-specific expertise also differentiates how individuals prepare for and navigate complex problems: novices often approach design tasks with a less defined problem scope, leading to more incoherent performance (Chen, Yan-Ting, & Chia-Han, 2022); experts tend to participate more in collaborative settings, report a greater sense of ownership toward their individual contribution, and be more critical toward contributions by others (Thom-Santelli, Cosley, & Gay, 2010). It is plausible then that collaborative behaviors reflect both technological fluency and domain-specific expertise, though it is unclear how these relationships manifest in a VR space-building task. Greater familiarity with the technology may, for example, increase ease of action and yield higher frequencies of task behaviors, yet experts might also require fewer actions, iterations, and time to accomplish the same goals (Ericsson & Lehmann, 1996). Likewise, while technological fluency can reduce friction in coordinating behaviors it may also support more efficient task delegation, allowing collaborators to work in parallel without continuously monitoring or aligning behaviors with one another (Nokes-Malach, Meade, & Morrow, 2012; Olson & Olson, 2000).

Contextual differences such as group size can also influence interaction dynamics. For example, studies in nonimmersive educational and collaborative settings have shown that larger groups often struggle with coordination, yielding lower cohesion, reduced participation, and less effective information exchange (Rannastu, Siiman, Mäeots, Pedaste, & Leijen, 2019; Saqr, Nouri, & Jormanainen, 2019). We posit that such relationships can manifest in VR collaborative behaviors: individuals in smaller groups may take on larger portions of the virtual task, whereas those in larger groups may be delegated less work yet face greater cognitive load and difficulty in aligning collective effort. Group size can also more broadly impact nonverbal and verbal behaviors. For instance, Hadley, Whitmer, Brimijoin, and Naylor (2021) found that, during face-to-face conversations, triads employed more optimal listening head orientations and took longer speaking turns compared to dyads. In VR, group size has also been found to saliently predict turn-taking behaviors in classroom activities such as discussions and design activities (Wang et al., 2025).

Collectively, these findings underscore the role of internal dispositions and contextual factors in shaping social and collaborative behavior, which Heerey (2015) argues is essential in understanding downstream social outcomes. This is echoed by Markowitz et al. (2025), who stressed the importance to consider both environmental and individual factors in studying VR social interactions. Yet, there remains a lack of research studying how individual differences and contextual factors are related to collaborative behaviors in immersive virtual environments. **RQ2**, therefore, aimed to understand the predictors of collaborative behaviors, focusing specifically on prior XR and design experiences, as well as group size.

1.5. Psychological outcomes associated with user behaviors

Collaborative behaviors have been linked to psychological states and evaluations of social experiences, with most works examining face-to-face or nonimmersive contexts. A meta-analysis by Mogan, Fischer, and Bulbulia (2017) found synchrony to be associated with enhanced collaboration, affect, social cohesion, and partner perception. However, there are exceptions to this relationship, as reduced synchrony may signal greater division of labor and more positive outcomes. For instance, Wallot, Mitkidis, McGraw, and Roepstorff (2016) found that less synchronized behavior could lead to higher-quality collaborative outcomes in a dyadic face-to-face production task and suggested that behavior alignment may hinder effective division of labor or initiative-taking. Relatedly, Howard, Di Eugenio, Jordan, and Katz (2017) facilitated dyadic collaboration of students interacting through a computer interface and demonstrated that shifts in task initiative predicted productive knowledge coconstruction and individual learning. In Duran et al.'s (2024) study of triads collaboratively completing an educational creative thinking game over Zoom, the authors further found that the timing of gameplay and the number of collaborative turns were associated with how well teams solved online problems. These studies collectively suggest that real-time collaboration patterns provide insights into both what groups achieve and how participants perceive their collaboration success and effectiveness.

That said, few studies have systematically examined how collaborative behaviors are related to the perception of group closeness and task outcomes in VR. Existing VR works have focused on motion-based synchrony measures (e.g., Sun et al., 2019) as a predictor of perception of others, while largely overlooking other behavioral traces of collaboration, particularly those expressed through virtual object actions. In this work, we address this gap by examining both action- and movement-based synchrony, individual task-engagement behaviors, and participants' subjective perceptions of their collaborative experience (**RQ3**). Through linking quantifiable behaviors to downstream outcomes, this work supports the practical deployment of real-time feedback systems and advances theoretical understandings of the psychological outcomes of collaborative behaviors.

2. Methods

2.1. Dataset and task

We examined a dataset collected during a university course about VR in 2024, where groups of students used the social VR platform ENGAGE to complete a collaborative space-building activity. Past work on the dataset has examined the effects of groups revisiting the virtual worlds they built together in a later session (Wang, Santoso, Han, Srirangarajan, & Bailenson, 2026), but no work has analyzed behavior during the building process.

An onboarding session preceded the space-building activity. During these 40- to 50-min sessions, a teaching staff member introduced participants to ENGAGE, guided them in creating an avatar that looks and feels like them, and walked them through a series of platform-specific exercises (e.g., virtual movement, importing and manipulating 3D objects). Participants were allowed to move freely within the virtual environment using physical motion (e.g., walking around their physical room) and virtual motion (e.g., teleportation).

In the week following onboarding, participants gathered with their group to complete the collaborative space-building activity in VR. For this task, participants worked in groups of two to seven ($M = 3.81$, $SD = 1.20$) and were instructed to construct a virtual replica of a real-world physical space built in an empty event space of 10 by 7 meters. Prior to the VR sessions, in a face-to-face class session of 70 min, participants worked in a larger group (i.e., two VR groups assigned to recreate the same space) and used their laptop computers and Meshy AI, an online generative AI platform that allows for text-to-model asset creation, to generate realistic 3D models of the physical space they were instructed to recreate. In the class session, participants collaboratively (1) brainstormed a list of items they would like to generate from the real-world physical space, (2) input and iterated on object prompts for generating 3D meshes of the objects, and (3) textured them. They repeated these steps until they finished generating all the objects. These models, which ranged between 18 and 25 objects per larger group ($M = 21.71$, $SD = 1.60$), were uploaded to ENGAGE as 3D assets by a researcher so that participants could import and manipulate them using the asset library. During the later VR session, each group was limited to using the 3D objects previously generated by their larger group, alongside the platform's default assets. This meant that the pairs of VR groups that made up each larger group accessed an identical asset library.

During the VR activity, participants gathered with their VR group in the same physical space and were placed into an empty virtual environment in ENGAGE, with a 360-degree photo imported as a sphere into the virtual environment as a reference. Participants were instructed to create a virtual replica "that is a one-to-one, true-to-scale mapping of the physical space you see in 360." They were given 20 min to complete the collaborative task, and instructions were given by the teaching staff at the start of the sessions. In practice, participants took up to 30 min to complete the exercise. All participants remained visible in their avatars and were allowed to chat as they were physically colocated. During the exercise, participants were instructed to utilize the platform's built-in tools, which enabled them to import, manipulate (i.e., translate, rotate, scale), and delete 3D objects from the asset library and draw using 3D pens. Participants were also allowed to manipulate and delete objects added by other users. Fig. 1 shows screenshots from the VR sessions. As the dataset was collected during a university course, some groups were made up of both nonconsenting and consenting students. Because our collaborative behavior measures were derived at both the dyadic and individual level, we analyzed data from consenting students and all-consenting dyads, yielding a sample size of 146 participants. In other words, we excluded self-report and behavioral analysis of dyads in which one or both of the dyadic members were nonconsenters. A total of 164 students participated in the activity, with 18 nonconsenting students. Supplementary Materials detail the descriptive statistics of the dataset.

2.2. *Motion tracking data*

From the tracking data collected during the collaborative activity using ENGAGE's recording feature, we extracted two types of behavioral measures: those related to body motion and those concerning actions.

2.2.1 Motion behavior

Head and hand motion were tracked at 30 Hz and contained measures of both abstract (i.e., virtual teleportation) and physical movement. Consistent with prior works on VR synchrony (e.g., Han et al., 2023), we aggregated abstract and physical motion into a single measure of visible motion. For the three tracked points (i.e., headset, left, and right controller), we extracted their position (i.e., x, y, z) and orientation (i.e., pitch, yaw, roll).

2.2.2 Action-level behavior

From the recordings, we extracted individual action logs. We analyzed four types of individual-level actions: object creation, object deletion, object manipulation, and draw-

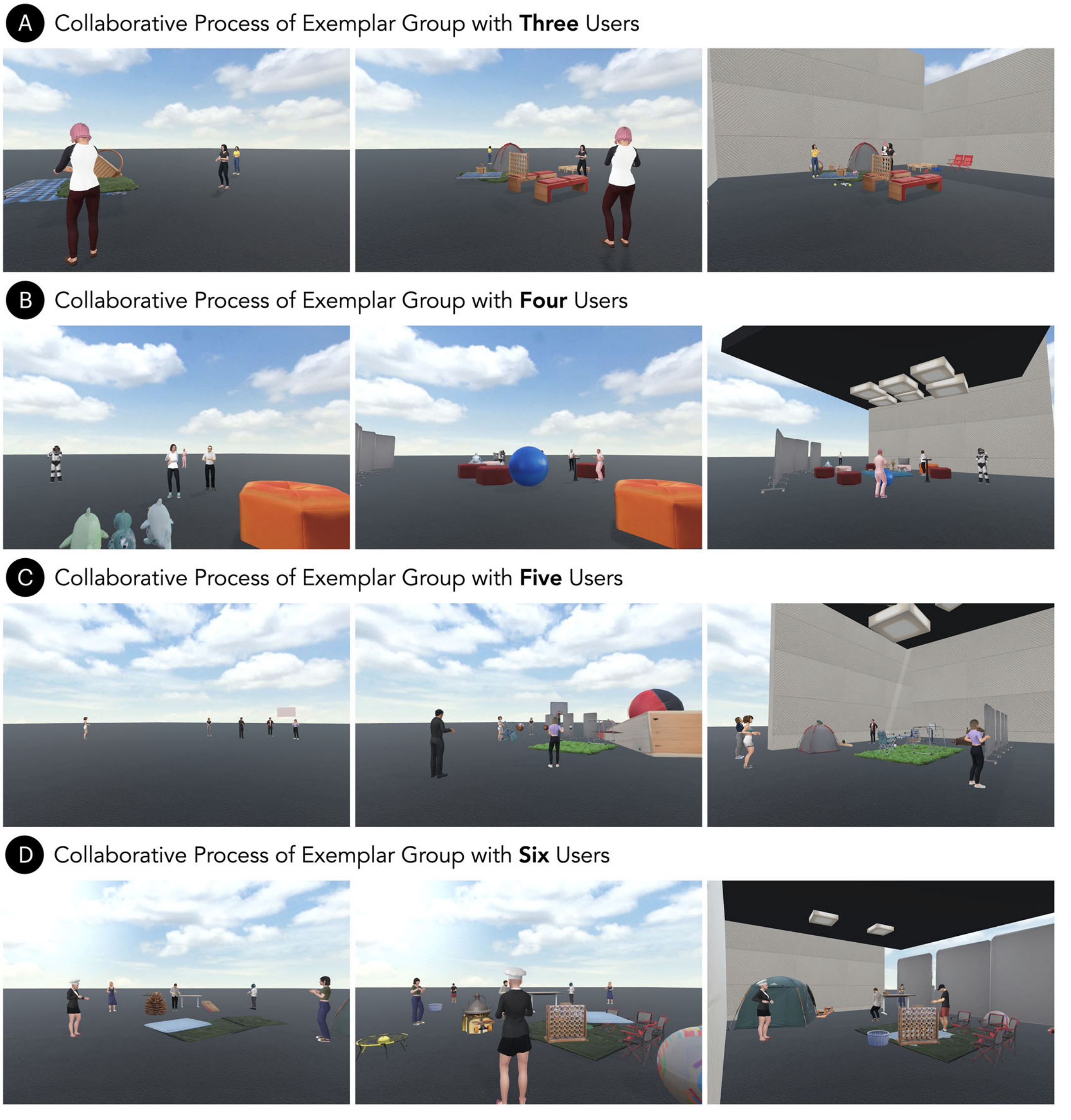

Fig. 1. Screenshots of collaborative processes in ENGAGE. Subfigures (A–D) show the collaborative process of four groups, with the three panels within each subfigure showing snapshots taken at the same virtual position over time from left to right.

ing. Given the structure of the recording files, object creation, deletion, and editing are tracked at 30 Hz, and drawing actions were recorded at around 0.21 Hz. The lower recording frame rate of drawing actions reflects the platform's recording settings, which logged changes in 3D drawings at a coarser temporal resolution. Object creation and deletion are logged whenever a user creates or deletes an object, and object editing was recorded when-

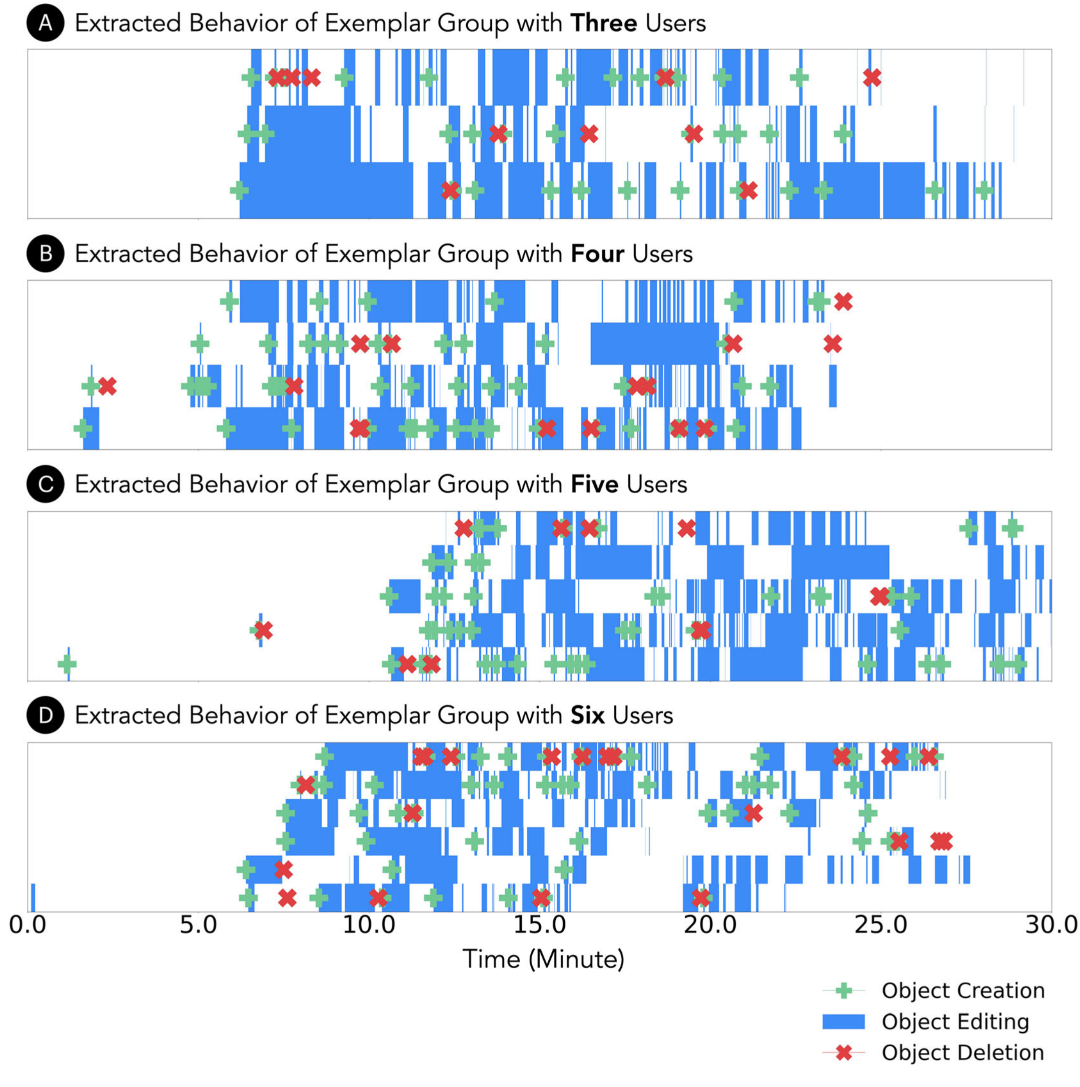

Fig. 2. Examples of extracted action sequences. Each subfigure shows an example of action sequences extracted from an ENGAGE session. Groups shown here map to those in Fig. 1, with the same lettered subfigures referring to the same group (e.g., Figs. 1A and 2A). Each row in a given subfigure represents the action sequence of a single user. Adjacent actions are smoothed based on the protocol outlined in Section 2.3.1.

ever a user translated, rotated, or scaled an existing object. We further extracted drawing behaviors using 3D pens. In ENGAGE, a 3D object of the drawing was created whenever a user marked a set of drawing strokes as complete. Since the created drawings can be manipulated and deleted, similar to other assets, we grouped drawing and object manipulation actions under editing behavior. Fig. 2 shows examples of the extracted action-level behavior.

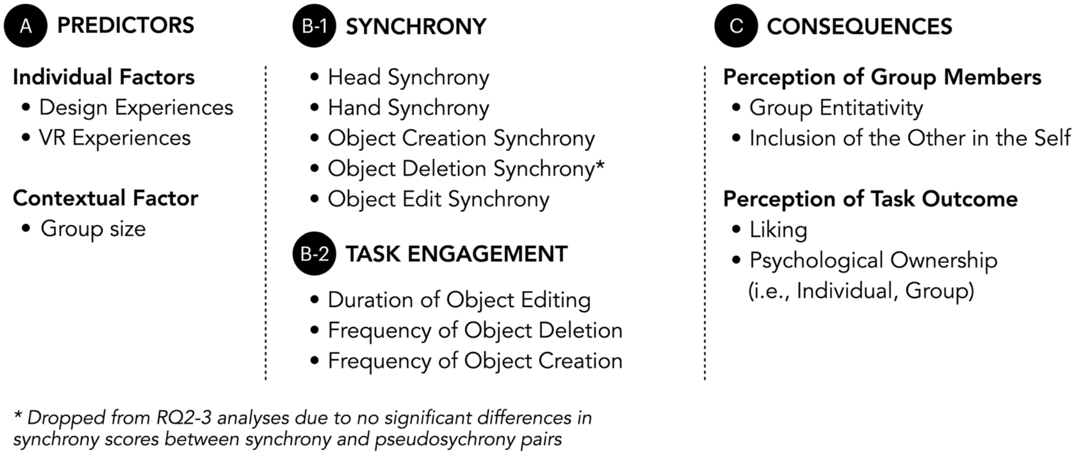

Fig. 3. Summary of analyzed measures.

2.3. Formulating and extracting synchrony and task engagement

From the extracted behavioral data, we derived two groups of measures concerning collaborative behaviors: dyadic synchrony and individual task engagement (Fig. 3B). The two differ in several ways. The first is in their time scale. While synchrony emphasizes temporal coordination of body movements and actions, task engagement measures capture collaboration dynamics across time, where related actions may be separated by substantial temporal gaps. Furthermore, while task engagement (actions on task objects) is initiated by individuals and reflects intentional goal-directed behavior, synchrony directly reflects coordination across multiple individuals and could emerge through conscious and subconscious processes (Sun et al., 2019). Finally, from a distributed cognition perspective, synchrony provides insight into how cognition is distributed across individuals, whereas individual task engagement reflects how people interact with the virtual environment, and by extension, how cognition is supported by external structures and tools. The following subsections define the different measures. Table 1 presents their descriptive statistics.

2.3.1. Synchrony

We extracted five dyadic synchrony measures. Two measures concerned motion synchrony and are derived from synchrony calculations from prior VR works on synchrony (Miller et al., 2021; Sun et al., 2019). For each individual, we first computed a time series of head and hand speed, where head speed reflects the translational speed of the headset in virtual space and hand speed captures the combined translational speed of the left and right controllers, both in meters per second. The decision to aggregate hand motion, as opposed to examining each hand separately, served two purposes: to account for handedness and avoid collinearity between left- and right-hand synchrony measures. The head and hand synchrony scores for a dyadic user pair with users P_i and P_j are defined as:

$$MotionSync\left(P_i, P_j\right) = cor\left(m_i, m_j\right), \tag{1}$$

Table 1
Descriptive statistics of extracted collaborative behavior measures

Collaborative behavior measure	Whole session		120-s segments	
	Summary statistics	Percentage of zero values (%)	Summary statistics	Percentage of zero values (%)
Synchrony (dyadic)				
Head synchrony	$M = 0.038$ $SD = 0.043$	N/A	$M = 0.031$ $SD = 0.11$	N/A
Hand synchrony	$M = 0.055$ $SD = 0.067$	N/A	$M = 0.052$ $SD = 0.11$	N/A
Object creation synchrony	$M = 0.065$ $SD = 0.043$	16.00	$M = 0.28$ $SD = 0.20$	84.03
Object edit synchrony	$M = 0.23$ $SD = 0.083$	0.00	$M = 0.31$ $SD = 0.21$	37.54
Object deletion synchrony	$M = 0.085$ $SD = 0.072$	83.48	$M = 0.38$ $SD = 0.25$	96.52
Task engagement (individual)				
Object creation (occurrence per minute)	$M = 0.56$ $SD = 0.34$	0.68	$M = 1.01$ $SD = 0.80$	52.76
Object editing (seconds per minute)	$M = 17.63$ $SD = 5.67$	0.00	$M = 23.74$ $SD = 13.71$	36.52
Object deletion (occurrence per minute)	$M = 0.17$ $SD = 0.12$	8.90	$M = 0.69$ $SD = 0.39$	81.28

Note: Summary statistics were calculated based on nonzero values, and the percentages of zero values are shown in adjacent columns. The percentage of zero values contextualizes statistical approaches outlined in Section 2.5.2.

where m_i represents the time series of either head and hand speed for user P_i. Similar to past works (Miller et al., 2021; Sun et al., 2019), we used Spearman's ranking correlation. With this formulation of behavioral synchrony, we extracted head and hand synchrony scores for all unique dyads within each VR session, with time series extracted for the entire session, as well as scores for time series extracted at nonoverlapping windows of consecutive 120 s (e.g., 0–120 s, 120–240 s). The 120 s was selected based on testing window sizes to balance temporal resolution with data density: it is fine-grained enough to capture meaningful changes over time while still containing sufficient behavioral signals to compute action-related measures without excessive sparsity. Sampling synchrony scores over time allows us to study changes of behavioral synchrony over time (**RQ1b**). Three measures are related to action synchrony. Our formulation of action synchrony bears similarities with both cross-recurrence analysis and the motion synchrony presented previously. Specifically, action synchrony is defined as the average pairwise Jaccard index between a user and all remaining users in their group. For a dyadic pair of users P_i and P_j, action synchrony scores are calculated as:

$$ActionSync\left(P_i,\ P_j\right) = \frac{\left|A_i \cap A_j\right|}{\left|A_i \cup A_j\right|}, \tag{2}$$

where A_i represents the entire set of time bins during which user P_i performed a specific action. When the denominator $|A_i \cup A_j|$ is zero, that is, neither user P_i or P_j has engaged in any actions over the course of the VR session or the considered time duration, we drop the dyadic pair from analysis. Like cross-recurrence analysis, our formulation focuses on quantifying the extent to which events co-occur in dyadic pairs and further aligns with the established measures of synchrony of continuous body motion. With this definition, we can derive, for any dyad, their object creation, edit, and deletion synchrony scores. In addition to deriving action synchrony scores for the entire VR session to confirm the presence of synchrony (**RQ1a**) as well as its predictors and psychological outcomes (**RQ2−3**), we computed more fine-grained action synchrony scores to study how they change over time (**RQ1b**). For this, we again used nonoverlapping 120-s sampling windows and calculated the action synchrony using time bins that fall within sampled windows.

Since object creation and deletion events are logged as instantaneous time points, we extended each event to a 10-s window, spanning from 5 s before to 5 s after the recorded timestamp. Because users tend to edit 3D objects in relatively brief bursts, such as rotating an object, pausing to inspect it, then resuming edits, we applied a smoothing procedure that merges adjacent editing actions when they targeted the same object and occurred within 5 s apart. This approach reduces noise and produces more coherent signals, capturing the editing process.

2.3.2. Individual task engagement

We extracted three measures related to individual task engagement: *object editing*, *object creation*, and *object deletion*. Object editing quantifies the amount of time an individual spent editing virtual objects, in seconds per minute. Object creation records the frequency of object creation, with a unit of occurrence per minute. Finally, object deletion captures an individual's frequency of object deletion, also in occurrence per minute.

2.4. *Self-report and contextual measures*

To investigate **RQ2–3**, we first detail measures related to individual dispositions and contextual differences (Fig. 3A), as well as post-VR measures on the perception of group members and task outcome (Fig. 3C). See Supplementary Materials for additional information on the self-report scales.

2.4.1. *Individual and contextual factors*

2.4.1.1. Previous design experience: In a prestudy questionnaire administered prior to the study activities, participants self-reported their prior experiences in design (e.g., 3D modeling, user interface design) on a 5-point Likert scale (1 = None, 5 = 4+ years; $M = 1.73$, $SD = 1.00$).

2.4.1.2. Previous XR experience[1]*:* Participants reported their prior XR experience on a 5-point Likert scale by approximating the number of times they have used XR prior to the university course (1 = I have never experienced XR, 5 = More than 10 times; Mean = 2.62, $SD = 1.28$).

2.4.1.3. Group size: For contextual differences, we extracted the group size for the collaborative VR activities, which ranged from two to seven ($M = 3.81$, $SD = 1.20$).[2]

2.4.2. *Perception of group closeness and task outcomes*

We analyzed five measures related to the perception of group closeness (i.e., affiliation) and task outcomes. Measurements were collected in a questionnaire after the space-building activity.

2.4.2.1. Group entitativity: The first measure of group closeness and affiliation captures the degree of common behaviors and intentions toward a group's shared goal (Hamilton, Sherman, & Castelli, 2002). Entitativity was measured using an 8-item 7-point Likert scale (1 = Strong disagree, 7 = Strongly agree) following prior research (Han et al., 2023; Rydell & McConnell, 2005). A composite score was created by averaging the items ($r_\alpha = .87$; $M = 3.06$, $SD = 0.64$).

2.4.2.2. Inclusion of the other in the self: The second measure of group closeness and affiliation captures a more subjective evaluation of relationship closeness (Gächter, Starmer, & Tufano, 2015). Following Aron et al. (1992), participants reported their levels of self-other overlap with their activity group by selecting one of seven drawings with different degrees of overlapping circles (1 = Not overlapping, 7 = Almost fully overlapping; $M = 3.75$, $SD = 1.50$).

2.4.2.3. Virtual space liking: Following past works (e.g., Harmon-Jones & Allen, 2001), the dataset assessed liking toward the virtual space built by the participants using a single-item 7-point Likert scale (-3 = Dislike, 0 = Neutral, 3 = Like; $M = 1.00$, $SD = 1.41$).

2.4.2.4. Virtual space psychological ownership: Psychological ownership is defined as the "possessive feeling that some object is mine or ours" (Van Dyne & Pierce, 2004, p. 439). The dataset measured the degree to which individuals feel that they themselves or their group owns the virtual space they built using a 4-item 5-point Likert scale (1 = Not at all, 5 = Extremely) following Van Dyne and Pierce (2004). The dataset further distinguished between individual psychological ownership (i.e., the space is mine) and collective psychological ownership (i.e., the space is ours) as the two draw on different aspects of self- and group-based identity (Pierce & Jussila, 2010). The items for individual psychological ownership ($r_\alpha = .84$; $M = 2.97$, $SD = 0.92$) and collective psychological ownership ($r_\alpha = .79$; $M = 3.42$, $SD = 0.81$) are averaged to create composite scores.

2.5. Data analysis

In this section, we detail the analysis plan for answering **RQ1–3**. All models were built in R, with linear mixed-effects models built using the lmerTest package, and logistic models were built using the Stats package. Effect sizes and confidence intervals were computed with bootstrapping using the performance package. The effect size and confidence intervals for the Wilcoxon signed rank test in Section 2.5.1 are calculated using the rstatix package, and those for the linear model are estimated using the boot package. Significance level is evaluated at $\alpha = 0.05$.

2.5.1. Synchrony and pseudosynchrony

To confirm that synchrony exists (**RQ1a**), we compared synchrony scores computed from dyads within the same group with those generated from pseudosynchronous dyads (i.e., participants pairs who did not interact with each other) (Bernieri, Reznick, & Rosenthal, 1988; Sun et al., 2019). This allows us to investigate whether synchrony scores are different from correlations from motion and action time series by chance. After extracting synchrony scores for all dyads in each VR session, we sought to generate an equal number of unique dyads who were not in the same group. From these dyads, we computed pseudosynchrony using the same approach for calculating synchrony. Unlike prior works, where synchrony and pseudosynchrony are compared across groups interacting in similar spatial and social context, our dataset presents a unique challenge where group sizes varied from two to seven, and as did the environment (e.g., what 3D assets are made available and created by the groups).

For example, as group size can influence levels of motion and action frequency, pseudosynchronous pairs drawn from different-sized groups may exhibit lower baseline synchrony. Relatedly, while all participants worked on building virtual replicas, the specific objects available to them varied across groups, particularly because users uploaded AI-generated assets via Meshy AI. When pseudosynchronous pairs are sampled from groups with differing "action affordances" due to these varying asset libraries, their pseudosynchrony scores may differ from those of synchronous pairs for reasons unrelated to whether the two users are interacting with each other. That is, if pseudosynchronous pairs are naively selected from different groups, the resulting differences in correlations may stem from variations in social context and task constraints, rather than the nature of the social interaction itself. With this in mind, we sampled pseudosynchrony pairs by first matching groups with identical access to 3D models and action affordances during the VR activity, as noted in Section 2.1, and then selecting pseudosynchronous pairs across these matched groups. In practice, the possible number of pseudosynchronous pairs ($N = 223$) is slightly lower than that of synchronous pairs ($N = 225$). We report a sensitivity analysis with two additional approaches in sampling pseudosynchronous pairs in the Supplementary Materials. Both produced results consistent with the present approach.

To determine the presence of synchrony, we built linear mixed-effects models predicting each of the synchrony scores with the dyad type (i.e., synchrony, pseudosynchrony). We included random effects for the groups to which the dyad members belonged, specifying crossed random intercepts to account for dyads sampled from varying groups. Since synchrony scores for object deletion violated the normality assumptions of linear models, we used the Wilcoxon signed rank test to model the effects of dyad type on synchrony.

2.5.2. Synchrony and task engagement over time

We investigated how dyadic synchrony and individual task engagement varied over time (**RQ1b–c**) by predicting synchrony and task engagement scores sampled at 120-s intervals. For synchrony and task engagement scores that contained a considerable ratio of zeros, we built two-part hurdle-type models. That is, for each measure, we built two separate models, one logistic model that predicts whether a data point sampled at a particular time point is zero or not, and another that models the nonzero portions of the outcome measures. This approach aligns with prior works (e.g., Atkins et al., 2013) that leverage two independent models (i.e., one for modeling zeros and nonzero values, another for modeling nonzero values) to address data distributions with large ratios of zeros. During the modeling process, we also observed that many of the data distributions do not follow a linear trend over time but instead appeared to follow U shapes. Therefore, for all measures, we compared the model fit of (1) a mixed-effects model where the outcome is modeled only as a first-order linear function of time, and (2) a more complex model with both a linear and quadratic term for time.

To identify the appropriate random-effects structure for models predicting dyadic synchrony, we compared two mixed-effects models of increasing complexity across both types of models. The first model included only a random intercept for the dyadic pair. The second model included a random effect of the dyadic pair nested within the group. Likelihood ratio tests indicated that the second model provided the optimal balance of model fit and parsimony for the majority of the dyadic outcomes. Group size was included as a predictor to control for the large variation in size across the collaborative groups for the dyadic measures. The two potential models for dyadic pair d, sampled from groups g, at time t, are, therefore, formulated as follows:

$$outcome_{gdt} = b_0 + b_1 t + b_3 group\ size + u_{0g0} + v_{0gd} + \epsilon_{gdt}, \tag{3.1}$$

$$outcome_{gdt} = b_0 + b_1 t + b_2 t^2 + b_3 group\ size + u_{0g0} + v_{0gd} + \epsilon_{gdt}, \tag{3.2}$$

where b_0 represents the prototypical starting intercept, b_1 and b_2 capture the linear and quadratic effects of time, b_3 accounts for the effect of group size, u_{0g0} and v_{0gd} are the random intercepts for dyadic pair d nested within group g, respectively, and ϵ_{gdt} denotes the residual error for dyad d in group g at time t. When models yielded singular fits, we report results from the more parsimonious models with simpler random effect structures and include details of the alternative models and their parameter estimates in the Supplementary Materials.

For models built to predict individual task engagement, we included random intercept for the individual user. To decide on the appropriate random-effect structure, we compared these models with a more complex model with a random effect of individual users nested within their respective group. Likelihood ratio tests indicated that the nested models did not provide a significantly better fit; therefore, we retained the more parsimonious models with individual users modeled as random intercepts. The resulting models for user i in group g at time t are, therefore, formulated as follows:

$$outcome_{git} = b_0 + b_1 t + u_{0i} + \epsilon_{git}, \tag{4.1}$$

$$outcome_{git} = b_0 + b_1 t + b_2 t^2 + u_{0i} + \epsilon_{git}, \tag{4.2}$$

where b_0 represents the prototypical starting value of the outcome at $t = 0$, b_1 and b_2 capture the linear and quadratic effects of time, respectively, u_{0i} is the random intercept accounting for individual variability, and ϵ_{git} denotes the residual error for user i in group g at time t.

For each measure, we compared the model fit between the linear and quadratic models using a likelihood ratio test to investigate whether the added parameterization in the quadratic models significantly improved model fit. We report the results of the likelihood ratio test and the findings based on the model that better fit the data.

2.5.3. Predictors of synchrony and task engagement

To study **RQ2**, we built linear-mixed effects models predicting each of the synchrony and task engagement measures, with fixed effects of group size, previous XR experiences, and previous design experiences. Predictors for the dyad-level synchrony scores were computed by averaging the individual values for the dyadic pairs. We tested two instantiations of the models, one with a random intercept of group, and another with additional random intercepts for each of two members within the group. Likelihood ratio tests confirmed that the more complicated model did not significantly improve model fit for any of the outcome variables, hence the more parsimonious model with only a random effect of group is used. For the individual measures of task engagement, we built the linear-mixed effects models with group as the random effect.

2.5.4. Psychological outcomes of synchrony and task engagement

To investigate the psychological outcomes of collaborative behaviors (**RQ3**), we built models predicting the perception of group closeness (i.e., group entitativity, inclusion of the other in the self) and perception of task outcome (i.e., liking of built space, individual and group psychological ownership). For each outcome, we built (1) a model using the four dyadic synchrony measures to predict the average outcome of the two members of the dyad and (2) a model using the three task engagement measures to predict the individual outcomes. We modeled the effects of synchrony and task engagement separately to maintain consistency in the level of analysis across fixed effects (i.e., dyadic for synchrony, individual for task engagement). In other words, each outcome was modeled twice: once with dyadic-level predictors and once with individual-level predictors. We, therefore, controlled the familywise error rate by evaluating statistical significance after applying a Bonferroni correction ($\alpha = 0.05/2 = 0.025$). Consistent with earlier modeling approaches detailed in Section 2.5.3, all models included group as a random intercept. When a mixed-effects model produced a singular fit, we report results from the corresponding linear model and detail model parameter estimates for the alternative models in the Supplementary Materials.

3. Results

We present results by research question: Section 3.1 compares synchrony with pseudosynchrony (**RQ1a**); Section 3.2 examines temporal changes in synchrony and task engagement (**RQ1b−c**); and Sections 3.3 and 3.4 report predictors (**RQ2**) and psychological outcomes (**RQ3**) of synchrony and task engagement.

3.1. Synchrony versus pseudosynchrony

Table 2 summarizes our comparisons between synchrony and pseudosynchrony scores. The models revealed significant differences between synchrony and pseudosynchrony pairs for head synchrony ($b = 0.033$, $SE = 0.0046$, $p < .001$), hand synchrony ($b = 0.053$, $SE = 0.010$, $p < .001$), object edit synchrony ($b = 0.023$, $SE = 0.011$, $p = .040$), and object creation synchrony ($b = 0.014$, $SE = 0.0051$, $p = .008$). All four measures were significantly higher than their pseudosynchrony counterparts, confirming the presence of synchrony (**RQ1a**). There was no evidence of synchrony in object deletion as it did not differ significantly across pair type ($W = 24{,}104$, $p = .53$). We, therefore, excluded object deletion synchrony in subsequent analyses. As noted in the Supplementary Materials, additional approaches in sampling pseudosynchronous pairs (i.e., random, by group size) yielded the same findings as those presented here, suggesting that the differences between synchrony and pseudosynchrony are not due to differences in access to 3D objects or group size, but is rather caused by whether users in the selected dyads are interacting with each other in a collaborative task.

3.2. Synchrony and task engagement changes over time

3.2.1. Synchrony

Fig. 4 summarizes findings on how dyadic synchrony changes over time (**RQ1b**). Table 3 presents model summaries. Head and hand synchrony did not violate the assumption of linear models, so we built linear mixed-effects models following Section 2.5.2 (Fig. 4A,B). For head synchrony, the model with the nested random effect resulted in a singular fit,

Table 2
Comparisons of synchrony and pseudosynchrony scores

Measure	Synchrony	Pseudosynchrony	Parameter estimates for pair type	Effect size and confidence interval
Head	$M = 0.038$	$M = 0.0033$	$b = 0.033$, $SE = 0.0046$, $p < .001$	R^2c = .31, CI = [.22, .43]
	$SD = 0.043$	$SD = 0.037$		R^2m = .14, CI = [.071, .22]
Hand	$M = 0.055$	$M = 0.0019$	$b = 0.053$, $SE = 0.010$, $p < .001$	R^2c = .75, CI = [.66, .82]
	$SD = 0.067$	$SD = 0.033$		R^2m = .14, CI = [.054, .26]
Object edit	$M = 0.23$	$M = 0.21$	$b = 0.023$, $SE = 0.011$, $p = .040$	R^2c = .27, CI = [.15, .39]
	$SD = 0.083$	$SD = 0.090$		R^2m = .018, CI = [.000053, .074]
Object	$M = 0.054$	$M = 0.042$	$b = 0.014$, $SE = 0.0051$, $p = .0087$	R^2c = .17, CI = [.086, .25]
creation	$SD = 0.046$	$SD = 0.039$		R^2m = .025, CI = [.003, .064]
Object	$M = 0.014$	$M = 0.011$	$W = 24{,}104$, $p = .53$	$r = .030$, CI = [.0021, .13]
deletion	$SD = 0.043$	$SD = 0.032$		

Note: We present means and standard deviations of scores calculated from synchrony and pseudosynchrony pairs, the model parameter estimates for pair type, and their effect sizes and confidence intervals.

Table 3
Models predicting synchrony scores over time

Synchrony outcome	Model type	Parameter estimates for time	Effect size and confidence interval
Head (conti.)	Quad.	$b_0 = 0.043$, $SE = 0.013$, $p < .001$ $b_1 = -0.0043$, $SE = 0.0011$, $p < .001$ $b_2 = 0.00012$, $SE = 0.000039$, $p = .0020$ $b_3 = 0.0037$, $SE = 0.0023$, $p = .12$	R^2c = .037, CI = [.014, 0.059] R^2m = .008, CI = [.004, .019]
Hand (conti.)	Quad.	$b_0 = 0.15$, $SE = 0.020$, $p < .001$ $b_1 = -0.0031$, $SE = 0.00098$, $p = .0016$ $b_2 = 0.000078$, $SE = 0.000034$, $p = .021$ $b_3 = -0.016$, $SE = 0.0039$, $p < .001$	R^2c = .28, CI = [.23, .33] R^2m = .028, CI = [.012, .050]
Nonzero versus zero object creation (logit)	Linear	$b_0 = -1.10$, $SE = 0.35$, $p = .0019$ $b_1 = 0.011$, $SE = 0.0076$, $p = .15$ $b_3 = -0.18$, $SE = 0.075$, $p = .014$	R^2c = .077, CI = [.031, .097] R^2m = .014, CI = [.004, .027]
Nonzero object creation (conti.)	Linear	$b_0 = 0.27$, $SE = 0.043$, $p < .001$ $b_1 = -0.0020$, $SE = 0.0012$, $p = .11$ $b_3 = 0.010$, $SE = 0.0080$, $p = .19$	R^2 = .013, CI = [.0010, .034]
Nonzero versus zero object edit (logit)	Quad.	$b_0 = 2.06$, $SE = 0.38$, $p < .001$ $b_1 = 0.70$, $SE = 0.056$, $p < .001$ $b_2 = -0.45$, $SE = 0.045$, $p < .001$ $b_3 = -0.29$, $SE = 0.085$, $p < .001$	R^2c = .22 CI = [.17, .25] R^2m = .075, CI = [.050, .11]
Nonzero object edit (conti.)	Linear	$b_0 = 0.43$, $SE = 0.032$, $p < .001$ $b_1 = -0.0053$, $SE = 0.00068$, $p < .001$ $b_3 = -0.0068$, $SE = 0.0067$, $p = .32$	R^2c = .066, CI = [.047, .094] R^2m = .026, CI = [.017, .040]

Note: We present model type (i.e., linear, quadratic) and parameters, as well as their effect size and confidence intervals. For the logistic quadratic model for object edit synchrony, time was centered and scaled to facilitate convergence; the model's time-related coefficients represent the effect of time per one unit of standard deviation.

Abbreviations: Conti., continuous; Quad, quadratic.

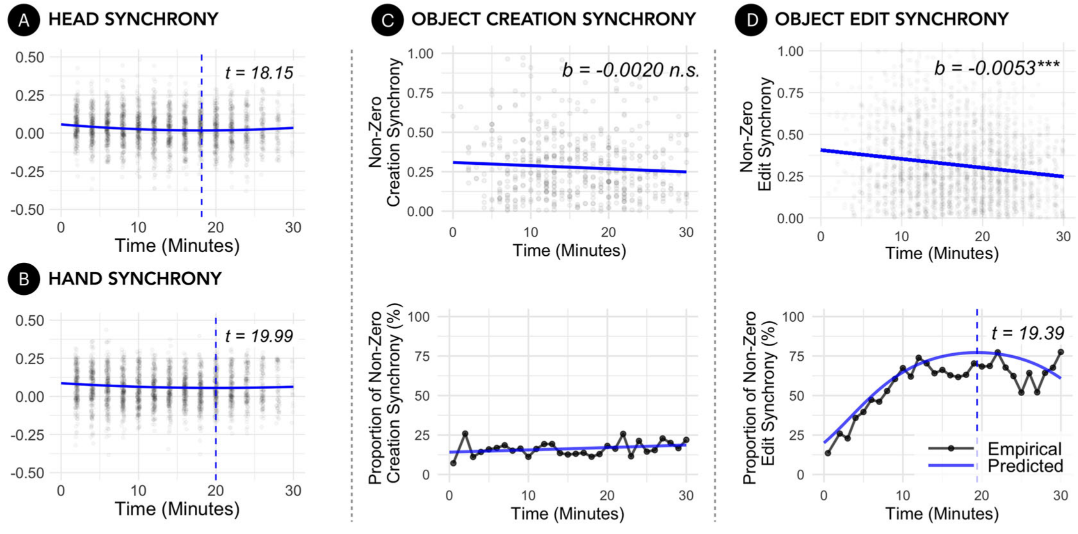

Fig. 4. Synchrony over time. Black dots and line segments represent empirical data, and blue curves represent predictions from fitted models assuming the group size of the sample mean across all groups, 3.81. Vertical dotted lines visualize the predicted peaks of quadratic models.
(* $p < .05$, ** $p < .01$, ** $p < .001$).

so a random effect of dyadic pair is used. The model with the added quadratic term for time resulted in better fit compared to the model without it ($\chi^2(1) = 9.52$, $p = .0020$) and revealed a significant negative linear effect of time ($b_1 = -0.0043$, $SE = 0.0011$, $p < .001$), a significant positive quadratic effect of time ($b_2 = 0.00012$, $SE = 0.000039$, $p = .0020$), and a nonsignificant effect of group size ($b_3 = 0.0037$, $SE = 0.0023$, $p = .12$). This is consistent with a U-shaped pattern with a turning point at 18.15 min. The quadratic model for predicting hand synchrony yielded a better fit compared to the linear model ($\chi^2(1) = 4.27$, $p = .039$) and showed significant effects of the linear term for time ($b_1 = -0.0031$, $SE = 0.00098$, $p = .0016$), the quadratic term for time ($b_2 = 0.000078$, $SE = 0.000034$, $p = .021$), and group size ($b_3 = -0.016$, $SE = 0.0039$, $p < .001$). This reflected a U-shaped trend over time, with a turning point at 19.99 min.

The distribution object creation synchrony contained a considerable number of zeros, so we followed the procedure for two-part hurdle-type models outlined in Section 2.5.2 (Fig. 4C). For the logistic model predicting whether or not object creation synchrony is nonzero, the quadratic model did not result in significantly better fit compared to a linear model ($\chi^2(1) = 1.13$, $p = .29$). The linear model revealed a significant main effect of group size ($b_3 = -0.18$, $SE = 0.075$, $p = .014$) and a nonsignificant effect of time ($b_1 = 0.011$, $SE = 0.0076$, $p = .15$). In predicting nonzero object creation synchrony, the more complicated quadratic model did not result in a better fit ($\chi^2(1) = 0.16$, $p = .69$). The model with random intercepts for the group of each member and the dyadic pair resulted in a singular fit, indicating minimal variance explained by the random effects. A simpler model with only the dyadic pair as a

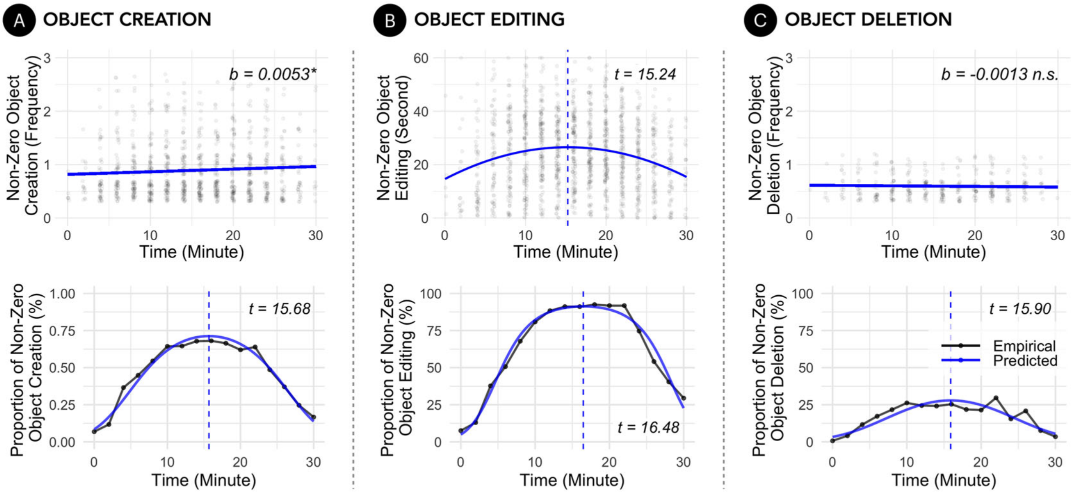

Fig. 5. Task engagement over time. Black dots and line segments represent empirical data, and blue curves represent predictions from fitted models. Vertical dotted lines visualize the predicted peaks of quadratic models. (* $p < .05$, ** $p < .01$, ** $p < .001$).

random effect also yielded a singular fit. We, therefore, used a linear model with time as the fixed effect to predict nonzero creation synchrony scores, which showed nonsignificant effect for both time ($b_1 = -0.0020$, $SE = 0.0012$, $p = .11$) and group size ($b_3 = 0.010$, $SE = 0.0080$, $p = .19$).

Object edit synchrony had a considerable number of zeros, so we again modeled it using the two-part approach (Fig. 4D). For predicting whether object edit synchrony was nonzero or not, the quadratic model yielded a significantly better fit compared to the linear model ($\chi^2(1) = 105$, $p < .001$). The model revealed a significant positive linear effect of time ($b_1 = 0.70$, $SE = 0.056$, $p < .001$), a significant negative quadratic effect of time ($b_2 = -0.45$, $SE = 0.045$, $p < .001$), and a significant effect of group size ($b_3 = -0.29$, $SE = 0.085$, $p < .001$).[3] The model suggests that the likelihood of any temporal overlap in object editing between dyadic pairs increased at the start of the activity, peaked at $t = 19.39$ min, and decreased thereafter. For predicting the nonzero object edit synchrony over time, we used a linear model as a quadratic model did not yield a significantly better fit ($\chi^2(1) = 1.88$, $p = .17$). The model revealed a nonsignificant effect of group size ($b_3 = -0.0068$, $SE = 0.0067$, $p = .32$), and that nonzero object edit synchrony significantly decreased over time ($b_1 = -0.0053$, $SE = 0.00068$, $p < .001$).

3.2.2. *Task engagement*

Fig. 5 summarizes changes in individual task engagement over time (**RQ1c**). Table 4 presents model summaries. All three measures of task engagement had a considerable number of zeros, so we applied the two-part modeling approach in Section 2.5.2. For predicting nonzero versus zero frequency of object creation, we chose a quadratic model over a linear

Table 4
Summary of models predicting task engagement over time

Task engagement outcome	Model type	Parameter estimates for time	Effect size and confidence interval
Nonzero versus zero object creation (logit)	Quad.	$b_0 = -2.38$, $SE = 0.20$, $p < .001$ $b_1 = 0.42$, $SE = 0.023$, $p < .001$ $b_2 = -0.013$, $SE = 0.00073$, $p < .001$	R^2c = .28, CI = [.24, .30] R^2m = .23, CI = [.20, .26]
Nonzero object creation (cont.)	Linear	$b_0 = 0.81$, $SE = 0.041$, $p < .001$ $b_1 = 0.0053$, $SE = 0.0023$, $p = .020$	R^2c = .11, CI = [.059, .16] R^2m = .0050, CI = [.00041, .17]
Nonzero versus zero object editing (logit)	Quad.	$b_0 = -2.96$, $SE = 0.23$, $p < .001$ $b_1 = 0.64$, $SE = 0.026$, $p < .001$ $b_2 = -0.020$, $SE = 0.00085$, $p < .001$	R^2c = .44, CI = [.40, .48] R^2m = .42, CI = [.39, .47]
Nonzero object editing (cont.)	Quad.	$b_0 = 14.62$, $SE = 1.60$, $p < .001$ $b_1 = 1.56$, $SE = 0.20$, $p < .001$ $b_2 = -0.051$, $SE = 0.0061$, $p < .001$	R^2c = .14, CI = [.10, .18] R^2m = .044, CI = [.026, .064]
Nonzero versus zero object deletion (logit)	Quad.	$b_0 = -3.37$, $SE = 0.26$, $p < .001$ $b_1 = 0.30$, $SE = 0.030$, $p < .001$ $b_2 = -0.0096$, $SE = 0.00096$, $p < .001$	R^2c = .18, CI = [.14, .24] R^2m = .14, CI = [.10, .18]
Nonzero object deletion (cont.)	Linear	$b_0 = 0.62$, $SE = 0.025$, $p < .001$ $b_1 = -0.0013$, $SE = 0.0014$, $p = .36$	R^2c = .008, CI = [.005, .11] R^2m = .0020, CI = [.0000057, .024]

Note: We present model type (i.e., linear, quadratic) and parameters, as well as their effect size and confidence intervals.
Abbreviations: Conti., continuous; Quad., quadratic.

one as it yielded a better fit ($\chi^2(1) = 432.55$, $p < .001$). The prototypical user has a starting log-odds of −2.38 ($SE = 0.20$, $p < .001$). There is a significant linear effect of time ($b_1 = 0.42$, $SE = 0.023$, $p < .001$) and a significant quadratic effect of time ($b_2 = -0.013$, $SE = 0.00073$, $p < .001$), suggesting that the likelihood of a user engaging in any amount of object creation follows an inverted-U shape peaking at $t = 15.68$ min. For predicting nonzero values of object creation, a linear model was sufficient for capturing the relationship between time and the outcome ($\chi^2(1) = 0.50$, $p = .48$). The prototypical individual exhibited a starting frequency of object creation of 0.81 ($SE = 0.041$, $p < .001$), which increased significantly over time ($b_1 = 0.0053$, $SE = 0.0023$, $p = .020$). In modeling whether or not an individual spends any time editing objects, a quadratic model yielded significantly a better fit compared to a linear one ($\chi^2(1) = 806.12$, $p < .001$). The prototypical individual had a starting log-odds of −2.96 ($SE = 0.23$, $p < .001$), with a significant positive linear effect of time ($b_1 = 0.64$, $SE = 0.026$, $p < .001$) and a significant negative quadratic effect of time ($b_2 = -0.020$, $SE = 0.00085$, $p < .001$). This revealed an inverted-U shape peaking at $t = 16.48$ min. We separately modeled the nonzero values of object editing and found the quadratic model to be a better fit compared to a linear one ($\chi^2(1) = 67.48$, $p < .001$). The prototypical individual begins the activity spending 14.62 s per minute editing objects ($SE = 1.60$, $p < .001$). There is a significant positive linear effect of time ($b_1 = 1.56$, $SE = 0.20$, $p < .001$) and a significant negative quadratic effect of time ($b_2 = -0.051$, $SE = 0.0061$, $p < .001$), reflecting an inverted-U curve peaking at 15.24 min.

Finally, we modeled individual object deletion frequency. For predicting whether or not individuals engaged in a nonzero amount of object deletion, a quadratic model was chosen as it yielded a better fit compared to a linear model ($\chi^2(1) = 124.06$, $p < .001$). The prototypical individual has a starting log-odds of −3.37 ($SE = 0.26$, $p < .001$), with a significant positive linear effect of time ($b_1 = 0.30$, $SE = 0.030$, $p < .001$) and a significant negative quadratic effect of time ($b_2 = -0.0096$, $SE = 0.00096$, $p < .001$). This reflects an inverted-U shape relationship with a peak at $t = 15.90$ min. For predicting the nonzero values of individual object creation, we used a linear model as the quadratic one did not improve model fit ($\chi^2(1) = 0.43$, $p = .51$). The prototypical individual had a starting object deletion frequency of 0.62 ($SE = 0.025$, $p < .001$), which decreases by a nonsignificant rate of 0.0013 ($SE = 0.0014$, $p = .36$).

3.3. Predictors of synchrony and task engagement

Table 5 summarizes our findings on predictors of collaborative behaviors. Prior levels of design experiences did not significantly predict any of the collaborative behavior measures ($ps > .05$). Prior XR experiences positively predicted object creation synchrony ($b = 0.0087$, $SE = 0.0040$, $p = .032$), object edit synchrony ($b = 0.025$, $SE = 0.0067$, $p < .001$), and frequencies of object creation ($b = 0.058$, $SE = 0.023$, $p = .013$) and object deletion ($b = 0.028$, $SE = 0.0083$, $p < .001$). Prior XR experiences did not significantly predict head synchrony, hand synchrony, or duration of object editing ($ps > .05$). Group size negatively predicted the frequency of object creation ($b = -0.058$, $SE = 0.027$, $p = .039$) and was nonsignificant in predicting all other collaborative behaviors.

Table 5
Summary of models predicting collaborative behavior measures using individual and contextual differences predictors

Collaborative behavior measure	Parameter estimates for individual and contextual differences predictors	Effect size and confidence interval
Head synchrony (dyadic)	Design Exp. ($b = 0.0025$, $SE = 0.0050$, $p = .62$) XR Exp. ($b = -0.0056$, $SE = 0.0038$, $p = .14$) Group size ($b = 0.00046$, $SE = 0.0042$, $p = .91$)	R^2c = .26, CI = [0.15, 0.45] R^2m = .012, CI = [.0050, .088]
Hand synchrony (dyadic)	Design Exp. ($b = -0.00070$, $SE = 0.0056$, $p = .90$) XR Exp. ($b = -0.0021$, $SE = 0.0042$, $p = .61$) Group size ($b = -0.012$, $SE = 0.0092$, $p = .19$)	R^2c = .73, CI = [.59, .83] R^2m = .033, CI = [.0020, .19]
Object creation synchrony (dyadic)	Design Exp. ($b = 0.00079$, $SE = 0.0054$, $p = .88$) **XR Exp. ($b = 0.0087$, $SE = 0.0040$, $p = .032$)** Group size ($b = -0.0060$, $SE = 0.0041$, $p = .16$)	R^2c = .23, CI = [.10, .40] R^2m = .063, CI = [.011, .19]
Object edit synchrony (dyadic)	Design Exp. ($b = -0.0029$, $SE = 0.0093$, $p = .75$) **XR Exp. ($b = 0.025$, $SE = 0.0067$, $p < .001$)** Group size ($b = -0.011$, $SE = 0.0077$, $p = .17$)	R^2c = .32, CI = [.19, .47] R^2m = .11, CI = [0.047, 0.23]
Object creation (individual)	Design Exp. ($b = 0.033$, $SE = .029$, $p = .25$) **XR Exp. ($b = 0.058$, $SE = 0.023$, $p = .013$)** **Group size ($b = -0.058$, $SE = 0.027$, $p = .039$)**	R^2c = .24, CI = [.12, .43] R^2m = .13, CI = [.050, .27]
Object editing (individual)	Design Exp. ($b = 0.57$, $SE = 0.47$, $p = .23$) XR Exp. ($b = 0.69$, $SE = 0.38$, $p = .071$) Group size ($b = -0.62$, $SE = 0.51$, $p = .23$)	R^2c = .29, CI = [.13, .48] R^2m = .070, CI = [.018, .19]
Object deletion (individual)	Design Exp. ($b = -0.015$, $SE = 0.011$, $p = .15$) **XR Exp. ($b = 0.028$, $SE = 0.0083$, $p < .001$)** Group size ($b = -0.0030$, $SE = 0.0084$, $p = .72$)	R^2c = .084, CI = [.045, .28] R^2m = .082, CI = [.024, .19]

Note: We present model parameters, as well as their effect size and confidence intervals. Significant predictors are bolded.
Abbreviation: Exp., experience.

3.4. Psychological outcomes of synchrony and task engagement

3.4.1. Synchrony

We present a summary of models using dyadic synchrony scores to predict perception of group closeness and task outcome in Table 6. No synchrony measures significantly predicted the outcome measures after Bonferroni correction (ps $> .025$). Object creation synchrony positively predicted virtual space psychological ownership ($b = 1.60$, $SE = 0.80$, $p = .046$) but was nonsignificant after correction.

3.4.2. Task engagement

Table 7 presents a summary of model parameters predicting psychological outcomes using individual task engagement measures. Frequency of object deletion was significant in predicting the two measures of perception of group closeness, namely, group entitativity ($b = 1.20$, $SE = 0.50$, $p = .017$) and inclusion of the other in the self ($b = 2.69$, $SE = 1.14$, $p = .020$). For both measures, individuals who more frequently deleted task objects were modeled to report higher levels of group closeness. No other individual task engagement measures significantly predicted the psychological outcomes (ps $> .025$).

4. Discussion

4.1. Presence and temporal dynamics of dyadic synchrony

Head, hand, object creation, and object edit synchrony user pairs in the same group were significantly higher than pairs from different groups, indicating alignment in collaborators' movements and actions greater than what would be expected by chance, confirming the presence of synchrony (**RQ1a**). In contrast, object deletion synchrony did not differ significantly between real and artificial pairs, likely because deletion events were rare and provided an overly sparse signal (Table 1). These findings differ from those in Sun et al. (2019), who observed lower scores in synchronous compared to pseudosynchronous pairs. We attribute these discrepancies to the different social and task contexts: Sun et al. (2019) interpreted lower synchrony as evidence of turn-taking in dyadic conversations, whereas in our space-building task, synchrony reflects how much groups temporally align their actions as they construct shared environments. Turn-taking is central to conversation but is arguably less relevant in space-building tasks, where overlapping actions and coordinated movement are often necessary. Further, compared to Sun et al. (2019) dyadic task conducted in an empty virtual environment, our larger group sizes (two to seven) and use of a more complex social VR platform likely heightened the temporal coordination required to complete the tasks.

The results on synchrony over time (**RQ1b**) showed that head and hand synchrony followed a U-shaped pattern over time. We interpret this as a shift in collaboration dynamics: early on, participants were more tightly aligned in their bodily movements, likely engaging in actions such as looking around the space together, gesturing toward, or manipulating 3D objects at similar moments. As the activity unfolded, group members likely engaged in more independent sets of actions and behaviors that entailed less temporal alignment across mem-

Table 6
Summary of models predicting psychological outcomes with dyadic synchrony

Psychological outcomes	Parameter estimates for synchrony measures	Effect size and confidence interval
Group entitativity	Head ($b = -0.12$, $SE = 0.64$, $p = .86$)	R^2c = .54, CI = [.38, .68]
	Hand ($b = -0.25$, $SE = 0.58$, $p = .67$)	R^2m = .070, CI = [.0030, .053]
	Object creation ($b = 0.81$, $SE = 0.59$, $p = .17$)	
	Object edit ($b = 0.048$, $SE = 0.35$, $p = .89$)	
Inclusion of the other in the self	Head ($b = 0.44$, $SE = 1.45$, $p = .77$)	R^2c = .62, CI = [.48, .73]
	Hand ($b = 0.10$, $SE = 1.34$, $p = .94$)	R^2m = .0090, CI = [.0030, .060]
	Object creation ($b = 1.53$, $SE = 1.33$, $p = .25$)	
	Object edit ($b = -1.19$, $SE = 0.79$, $p = .13$)	
Virtual space liking	Head ($b = 1.24$, $SE = 1.36$, $p = .36$)	R^2c = .61, CI = [.46, .74]
	Hand ($b = 1.40$, $SE = 1.25$, $p = .27$)	R^2m = .017, CI = [.0040, .075]
	Object creation ($b = 0.52$, $SE = 1.25$, $p = .68$)	
	Object edit ($b = 0.94$, $SE = 0.74$, $p = .20$)	
Virtual space individual PO	Head ($b = 0.028$, $SE = 1.00$, $p = .98$)	R^2c = .24, CI = [.12, .41]
	Hand ($b = 0.72$, $SE = 0.80$, $p = .37$)	R^2m = .0090, CI = [.0030, .076]
	Object creation ($b = 0.67$, $SE = 0.92$, $p = .47$)	
	Object edit ($b = 0.13$, $SE = 0.54$, $p = .81$)	
Virtual space group PO	Head ($b = 0.58$, $SE = 0.87$, $p = .51$)	R^2c = .27, CI = [.15, .42]
	Hand ($b = 1.13$, $SE = 0.69$, $p = .10$)	R^2m = .045, CI = [.016, .15]
	Object creation ($b = 1.60$, $SE = 0.80$, $p = .046$)	
	Object edit ($b = 0.32$, $SE = 0.47$, $p = .50$)	

Note: We present model parameters, as well as their effect size and confidence intervals. No synchrony measure was significant in predicting the outcomes.

Abbreviation: PO, psychological ownership.

Table 7
Summary of models predicting psychological outcomes with individual task engagement

Psychological outcomes	Parameter estimates for individual task engagement	Effect size and confidence interval
Group entitativity	Object creation ($b = -0.23$, $SE = 0.19$, $p = .24$)	R^2c = .27, CI = [.15, .42]
	Object editing ($b = 0.0081$, $SE = 0.010$, $p = .43$)	R^2m = .045, CI = [.016, .15]
	Object deletion ($b = 1.20$, $SE = 0.50$, $p = .017$)	
Inclusion of the other in the self	Object creation ($b = -0.47$, $SE = 0.45$, $p = .30$)	R^2c = .23, CI = [.091, .44]
	Object editing ($b = -0.017$, $SE = 0.024$, $p = .47$)	R^2m = .041, CI = [.0080, .14]
	Object deletion ($b = 2.69$, $SE = 1.14$, $p = .020$)	
Virtual space liking	Object creation ($b = -0.30$, $SE = 0.43$, $p = .48$)	R^2c = .23, CI = [.078, .42]
	Object editing ($b = 0.043$, $SE = 0.023$, $p = .064$)	R^2m = .028, CI = [.0040, .12]
	Object deletion ($b = 0.79$, $SE = 1.08$, $p = .47$)	
Virtual space individual PO	Object creation ($b = -0.045$, $SE = 0.27$, $p = .87$)	$R^2 = .054$, CI = [.015, .15]
	Object editing ($b = 0.027$, $SE = 0.014$, $p = .066$)	
	Object deletion ($b = 1.19$, $SE = 0.73$, $p = .10$)	
Virtual space group PO	Object creation ($b = 0.012$, $SE = 0.24$, $p = .67$)	$R^2 = .046$, CI = [.011, .15]
	Object editing ($b = 0.012$, $SE = 0.013$, $p = .37$)	
	Object deletion ($b = 1.06$, $SE = 0.64$, $p = .10$)	

Note: We present model parameters, as well as their effect size and confidence intervals. Significant predictors after Bonferroni correction are bolded.
Abbreviation: PO, psychological ownership.

bers. Nearing the two-thirds mark of the activity, head and hand synchronies increased, which we interpret as participants likely entering the final stages of collective critiquing and refinement. From a distributed cognition perspective, this suggests that during the early stages of the activity, groups increasingly relied on the task environment itself, rather than on continuous monitoring of others, as a shared external structure and cognitive artifact for behavioral coordination. Collaborators may have used the partially built virtual spaces as a scaffold for reasoning about and supporting their spatially distributed tasks, reducing the need for moment-to-moment bodily alignment. The increase in behavioral synchrony near the end signaled a possible shift in collaborative dynamics, likely guided by a more concerted and intentional group effort in completing the task. We posit that this pattern may extend to other collaborative tasks, where the progression of work reduces ambiguity and streamlines decision-making.

Nonzero object edit synchrony showed a significant declining trend, mirroring the broader shift noted earlier toward more independent and spatially distributed work once groups made progress on the spatial task and established a shared task understanding. The logistic model predicting the presence of object edit overlap revealed a more nuanced pattern that followed an inverted-U shape. This meant that early in the task, temporally overlapping editing actions were relatively rare, which we attribute to participants exploring the virtual environment, immersive platform, and its tools individually or in smaller subgroups, both of which can produce a high ratio of synchrony scores of zero. As time progressed, the likelihood of overlapping editing behaviors increased and peaked at approximately two-thirds of the way through the activity, reflecting a phase where more participants were simultaneously engaged and contributing to the collaborative outcome. Toward the end of the activity, this likelihood declined, possibly signaling that groups shifted into more sequential or fine-grained refinements where object edits became less overlapped temporally. For object creation synchrony, we observed nonsignificant trends for the nonzero values and for predicting whether there was any overlap in object creation across dyads. One explanation is that temporal overlap in creation behaviors was sparse (Fig. 4C), and examining it within a 120-s window further diluted its saliency. Furthermore, unlike the other synchrony measures, object creation is an instantaneous action that we extended to a 10-s duration. This analysis framework may be less well-suited for capturing changes in such momentary collaborative behaviors, particularly when compared to analyses that consider object creation as an individual-level behavior, which we examine next.

4.2. *Patterns of task engagement over time*

The analysis on task engagement over time (**RQ1b**) revealed that the likelihood of individuals engaging in any object creation, editing, and deletion followed an inverted-U shape over time, with peaks near the midpoint of the activity. We interpret this again as a reflection of changes in collaborative dynamics, where groups initially engage in a few object-related actions as they explored the environment, clarified the task, and decided how to begin. As time progressed, object-related actions increased as participants began to actively construct the space, yielding a mid-task phase in which object-related actions are most frequent. The presence of object-related participation decreased near the end of the activity as groups worked to complete the activity. While the nonzero duration of object editing also followed the same rise-and-fall pattern over time, nonzero object creation frequency significantly increased over time. This suggests that although object editing activity peaked mid-task and tapered off as the space gained structure, users who chose to create new objects may have continued to do so at a growing rate throughout the activity. The rate of nonzero object deletion did not change significantly over time, which we attribute to the sparsity of deletion (Table 1).

One interpretation for the findings on nonzero object creation and object editing is that as the built environment accrued structure and individuals developed clearer plans and gained familiarity with the platform, object creation became easier, more targeted, and less dependent on extended editing. From a distributed cognition perspective, the virtual space may have provided increasingly rich cues that supported efficient object placement, allowing users to add new elements without prolonged manipulation. This finding also suggests that there may be two behavioral types for participants regarding object creation, namely, those who created most extensively in the middle of the activity, and those who continued and increased their creation efforts throughout the activity. Post-hoc, qualitative observations from the recorded ENGAGE interactions corroborate this, in that while some participants slowed down object creation near the end of the activity to examine and debrief on their built spaces, others rushed to the end to put in objects such as walls, ceilings, and lights.

It is also likely that these two types of strategies are adopted based on the object's characteristics. For example, it is possible that singular distinctive objects, such as a tent, are heavily edited, whereas repetitive or modular objects, such as ceiling lights and walls, can be rapidly produced, require less editing, and are added once the room's structure is more fully defined. The results here could suggest that groups may have elected to work first on objects that required more manipulation, and, over time, focused more on creating and editing repetitive modular objects such as the ceiling lights and walls.

4.3. Prior experiences and group size in shaping collaborative behaviors

Individual differences in prior experiences in XR and design and group size differentially shaped collaborative behaviors (**RQ2**). To start, head and hand synchrony were not significantly predicted by any of the individual and contextual factors, suggesting that the degree of temporal alignment in how group members move is not reflective of their prior experiences and the number of individuals in their group. That said, prior XR experience positively predicted object creation and edit synchrony, as well as frequencies of object creation and deletion.

One possible interpretation for the effects on object creation and deletion frequencies is that participants with greater XR experience are more fluent in navigating the immersive interfaces and their design tools and, therefore, are more comfortable creating and deleting objects. The finding on object deletion frequency also suggests that those with more extensive XR experience exhibited greater intent to experiment with, revise, and refine their built spaces (Han et al., 2025). Importantly, this pattern was not observed for previous design experience, suggesting that collaborative behaviors in our context, although involving a task resembling design activities such as 3D prototyping, are shaped more by familiarity with immersive technology than by design background.

The positive relationship between previous XR experiences and individual task engagement offers one possible explanation for the effects previous XR experiences have on object creation and object edit synchrony. Namely, with higher frequencies of object-related actions for those more familiar with immersive technologies, the chances of greater overlap and temporal alignment in object-related actions also increase. Another explanation for these findings is that XR familiarity may enhance how smoothly individuals are able to coordinate their actions with others. When participants manipulate objects with less friction and navigate the virtual environment more fluidly, it might be easier for them to coordinate and align their behaviors with the unfolding behaviors of their collaborators. In other words, greater technological fluency and higher action density for participants with more previous XR experience could have jointly elevated object creation and edit synchrony scores. Analysis presented in the Supplementary Materials further found that participant pairs who exhibit greater object edit synchrony also stood closer to each other and yielded a lower division of labor by action category (i.e., greater differences in object edit durations). This pattern is compatible with the notion that task-related coordination is linked to how collaborators are situated within a shared environment. It is, however, unclear the causal relationship between the two, namely, whether spatial closeness elicits higher edit synchrony, whether higher synchrony prompts collaborators to work in closer proximity, or whether there are other factors that jointly influenced both. Broadly speaking, this exploratory analysis suggests that task-related synchronization is likely linked to both how collaborators are spatially situated within a shared environment and the extent to which their roles are differentiated.

Finally, group size negatively predicted the individual task engagement of object creation frequency. From a distributed cognition perspective, larger groups likely changed how collaboration occurs across group members and their shared virtual environment. As more collaborators contribute to the same outcome, opportunities for new object creation decrease as more aspects of the spatial environment are already being acted on or pre-emptively claimed by others. It is also possible that space became more spatially partitioned in larger groups, constraining the areas in which individuals can meaningfully contribute and create new objects for. Larger groups also introduce more concurrent actions for a member to monitor, which can increase the cognitive load of coordinating contributions, slow decision-making, ultimately resulting in less object creation.

4.4. Psychological outcomes of collaborative behaviors

There was little evidence that synchrony and task engagement are predictive of one's perception of group closeness and task outcome (**RQ3**). No synchrony measures significantly predicted the perception of group closeness and task outcome. This contrasts prior works linking synchrony to increased perception of others and affiliation (e.g., Hove & Risen, 2009; Mogan et al., 2017). One possible explanation is that our space-building task is more collaboratively complex and demands less explicit social attention compared to finger-tapping tasks (Hove & Risen, 2009) and joint movement activities (Mogan et al., 2017). The likely higher cognitive load and attention given toward the task as opposed to social others could have contributed to the null findings in linking synchrony and affiliation. Additional results in the Supplementary Materials revealed that the relationship between synchrony and perception of group closeness differed across group size. After controlling for group size and its interaction with synchrony, head synchrony significantly predicted inclusion of the other in the self, suggesting that group size may have also contributed to the discrepant results.

Object deletion frequency is positively associated with both measures of group closeness: group entitativity and inclusion of the other in the self. We posit that deletion behaviors may reflect a more collaborative mindset, where individuals more regularly revised their work to support task progression, thereby strengthening their perceived connection to the group. From a distributed cognition perspective, deletion also restructures the external environment that the group uses to think, coordinate, and manage the space-building activity. By potentially reducing visual clutter and cognitive load for processing external stimuli, those who frequently delete objects also help reduce coordination and cognitive demands placed on the group. Engaging in this form of environmental and task maintenance may reflect and heighten one's sense of engagement in the collaborative process, thereby strengthening their feelings of group closeness. Another interpretation of the finding is that one's willingness to delete task objects reflects a sense of psychological comfort with the group; individuals who feel able to critique and delete parts of their shared task outcome may experience greater group closeness.

4.5. Limitations and future work

This work has several limitations. To start, we engaged in a secondary analysis of a dataset of university students completing a space-building task. It is crucial to conduct preregistered studies replicating the relationships uncovered in the present work and propose extensions to theories such as distributed cognition. Future work can also explore a wider range of demographics and contexts to test whether the observed patterns extend beyond student populations and spatial-building tasks. Researchers should specifically evaluate how synchrony is affected by contexts, as the implications of synchrony may differ considerably across task types (e.g., motor, conversational). It is also crucial to propose quantitative methods for understanding the added complexities in interpreting their relative and absolute values when changes in task and spatial configurations may alter the autocorrelational properties of behavioral data. Relatedly, group composition, such as the mix of expertise levels and prior group familiarity, may influence interaction patterns. Examining varied group configurations would help determine the robustness of our findings. Future research can also explore using synchrony to predict dyad-specific self-report measures, such as each dyad's differences in perception of task and group. While we chose a sampling window of 120 s, researchers should examine other window sizes to understand the tradeoff between granularity and stability in behavioral markers, seeing as how the sparsity of object-related actions contributed considerably to the zero-inflated distributions for our temporal analysis.

Moreover, as groups spend more time together, collaboration strategies may shift, as new patterns of turn-taking and team roles may emerge. As such, longitudinal analyses of collaborative dynamics could determine how groups develop shared routines and rapport. Works can also systematically examine a broader range of possible antecedents, such as room size, spatial layout, and ambient mood. Developing a research agenda for parameterizing such environmental attributes would improve comparability across studies and contribute to a stronger understanding of how context shapes collaboration. Relatedly, coding schemes that capture collaborative dynamics such as division of labor, which can be derived from verbal and retrospective protocols (e.g., Jiang & Yen, 2009; Santoso, Wang, Han, & Bailenson, 2025), can offer additional insight into how cognition is distributed in immersive collaborative contexts. Furthermore, while our analysis centers on individual- and dyadic-level behaviors and self-report, research can also apply an object-centered approach to collaborative tasks. We see particular benefits to applying this approach toward tasks centered on fewer objects (e.g., virtual assembly), where analyses of the interdependencies of actions on the same object can inform our understanding of joint actions (Sebanz, Bekkering, & Knoblich, 2006; Sebanz & Knoblich, 2009). Finally, we did not include nonconsenting students in our analysis, despite them participating in the activities; the differences in analytical and experiential group size, as well as selection bias in participant consent, may have lowered the robustness of our findings. Comparison of results between our work and controlled in-lab studies with all consenting group members would help shed light on this shortcoming.

5. Conclusions

Collaborative behaviors are pervasive in everyday social interactions. We leveraged a large-scale classroom VR dataset to study how immersive collaborative behaviors unfold over time, how it is shaped by individual and contextual differences, as well as its relationship with the perception of group closeness and collaborative outcomes. Our analyses revealed that head and hand synchrony followed a U-shaped curve over time, and a downward trend was observed for nonzero object edit synchrony. When we modeled whether dyads temporally aligning their object editing behaviors (i.e., nonzero vs. zero synchrony), or whether individuals were actively creating, editing, or deleting objects (i.e., nonzero vs. zero task engagement), we found a consistent inverted-U pattern in their likelihood over time. In particular, the likelihood of these engaged moments increased early on in the activity, peaked around the midpoint of the activity, and declined afterward. We found that prior XR experiences are predictive of collaborative behaviors (e.g., object creation synchrony, object deletion frequency), while group size negatively predicted the frequency of object creation. There was no evidence linking collaborative behaviors to an individual's perception of collaborative outcomes, while a higher frequency of object deletion is linked to a higher perception of group closeness. These findings highlight the possibilities of leveraging user behaviors in VR to understand the psychological and social processes underlying collaborations and demonstrate how VR can be used to investigate collaborative interactions at scale.

Acknowledgments

We thank the teams at ENGAGE and Meshy AI for their help and support. We would also like to thank Gaby Harari for her helpful feedback. This work was supported by the Stanford Graduate Fellowship.

Conflicts of interest

The authors have no conflicts of interest to declare.

Data availability statement

The data reported in this article cannot be shared publicly due to privacy concerns, but will be shared on reasonable request to the corresponding author.

Notes

1 XR experiences were measured, instead of VR experiences, as other experiments conducted through the same university course also utilized mixed reality technologies.
2 Group size reflects the number of the users in each group who are involved in the VR activities.
3 For this model, time was centered and scaled to facilitate convergence. Its time-related coefficients represent the effect of time per one unit of standard deviation.

References

Andreas, K., Tsiatsos, T., Terzidou, T., & Pomportsis, A. (2010). Fostering collaborative learning in Second Life: Metaphors and affordances. *Computers & Education*, *55*(2), 603–615. https://doi.org/10.1016/j.compedu.2010.02.021

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

Atkins, D. C., Baldwin, S. A., Zheng, C., Gallop, R. J., & Neighbors, C. (2013). A tutorial on count regression and zero-altered count models for longitudinal substance use data. *Psychology of Addictive Behaviors*, *27*(1), 166–177. https://doi.org/10.1037/a0029508

Bernieri, F. J., Reznick, J. S., & Rosenthal, R. (1988). Synchrony, pseudosynchrony, and dissynchrony: Measuring the entrainment process in mother−infant interactions. *Journal of Personality and Social Psychology*, *54*(2), 243–253. https://doi.org/10.1037/0022-3514.54.2.243

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*(2), 103–124. https://doi.org/10.1207/S15327965PLI1302_01

Chen, H. J., Yan-Ting, C., & Chia-Han, Y. (2022). Behaviors of novice and expert designers in the design process: From discovery to design. *International Journal of Design*, *16*(3), 59–76. https://doi.org/10.57698/v16i3.04

Duffy, S. E., Feist, M. I., & McCarthy, S. (2014). Moving through time: The role of personality in three real-life contexts. *Cognitive Science*, *38*(8), 1662–1674. https://doi.org/10.1111/cogs.12124

Duran, N. D., Paige, A., & D'Mello, S. K. (2024). Multi-level linguistic alignment in a dynamic collaborative problem-solving task. *Cognitive Science*, *48*(1), e13398. https://doi.org/10.1111/cogs.13398

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, *47*(1), 273–305. https://doi.org/10.1146/annurev.psych.47.1.273

Freeman, G., Acena, D., McNeese, N. J., & Schulenberg, K. (2022). Working together apart through embodiment: Engaging in everyday collaborative activities in social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction*, *6*, 1–25. https://doi.org/10.1145/3492836

Gächter, S., Starmer, C., & Tufano, F. (2015). Measuring the closeness of relationships: A comprehensive evaluation of the 'Inclusion of the Other in the Self' scale. *PLOS ONE*, *10*(6), e0129478. https://doi.org/10.1371/journal.pone.0129478

Griffin, A. S., Guillette, L. M., & Healy, S. D. (2015). Cognition and personality: An analysis of an emerging field. *Trends in Ecology & Evolution*, *30*(4), 207–214. https://doi.org/10.1016/j.tree.2015.01.012

Hadley, L. V., Whitmer, W. M., Brimijoin, W. O., & Naylor, G. (2021). Conversation in small groups: Speaking and listening strategies depend on the complexities of the environment and group. *Psychonomic Bulletin & Review*, *28*, 632–640. https://doi.org/10.3758/s13423-020-01821-9

Hamilton, D. L., Sherman, S. J., & Castelli, L. (2002). A group by any other name—The role of entitativity in group perception. *European Review of Social Psychology*, *12*(1), 139–166. https://doi.org/10.1080/14792772143000049

Han, E., Miller, M. R., DeVeaux, C., Jun, H., Nowak, K. L., Hancock, J. T., Ram, N., & Bailenson, J. N. (2023). People, places, and time: A large-scale, longitudinal study of transformed avatars and environmental context in group interaction in the metaverse. *Journal of Computer-Mediated Communication*, *28*(2), zmac031. https://doi.org/10.1093/jcmc/zmac031

Han, E., Wang, P., DeVeaux, C., Harari, G. M., & Bailenson, J. N. (2025). Understanding the role of virtual mobility on how and what people create in virtual reality. *Thinking Skills and Creativity*, *56*, 101699. https://doi.org/10.1016/j.tsc.2024.101699

Harmon-Jones, E., & Allen, J. J. (2001). The role of affect in the mere exposure effect: Evidence from psychophysiological and individual differences approaches. *Personality and Social Psychology Bulletin*, *27*(7), 889–898. https://doi.org/10.1177/0146167201277011

Heerey, E. A. (2015). Decoding the dyad: Challenges in the study of individual differences in social behavior. *Current Directions in Psychological Science*, *24*(4), 285–291. https://doi.org/10.1177/0963721415570731

Hong, S. W., Jeong, Y., Kalay, Y. E., Jung, S., & Lee, J. (2016). Enablers and barriers of the multi-user virtual environment for exploratory creativity in architectural design collaboration. *CoDesign*, *12*(3), 151–170. https://doi.org/10.1080/15710882.2015.1081239

Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, *27*(6), 949–960. https://doi.org/10.1521/soco.2009.27.6.949

Howard, C., Di Eugenio, B., Jordan, P., & Katz, S. (2017). Exploring initiative as a signal of knowledge co-construction during collaborative problem solving. *Cognitive Science*, *41*(6), 1422–1449. https://doi.org/10.1111/cogs.12415

Hutchins, E. (2000). Distributed cognition. *International Encyclopedia of the Social and Behavioral Sciences*, *138*(1), 1–10.

Jiang, H., & Yen, C. (2009). Protocol analysis in design research: A review. *Proceeding of IADSR*.

Knoblich, G., & Sebanz, N. (2006). The social nature of perception and action. *Current Directions in Psychological Science*, *15*(3), 99–104. https://doi.org/10.1111/j.0963-7214.2006.00415.x

Lehikko, A., Nykänen, M., & Ruokamo, H. (2025). Previous VR experience increases embodiment in immersive VR safety training: An exploratory mixed-methods study. *Virtual Reality*, *29*(3), 126. https://doi.org/10.1007/s10055-025-01196-z

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, *36*(8), 1404–1426. https://doi.org/10.1111/j.1551-6709.2012.01269.x

Maloney, D., Freeman, G., & Robb, A. (2021). Stay connected in an immersive world: Why teenagers engage in social virtual reality. Proceedings of the 20th Annual ACM Interaction Design and Children Conference, 69–79. https://doi.org/10.1145/3459990.3460703

Maloney, D., Freeman, G., & Wohn, D. Y. (2020). "Talking without a voice" understanding non-verbal communication in social virtual reality. *Proceedings of the ACM on Human-Computer Interaction*, *4*, 1–25. https://doi.org/10.1145/3415246

Markowitz, D., Harari, G., Han, E., Miller, M., DeVeaux, C., & Bailenson, J. (2025). Individual differences and contextual influences on group behavior in virtual reality: An exploratory person−environment interactionist perspective. *PRESENCE: Virtual and Augmented Reality*, *34*, 395–414. https://doi.org/10.1162/PRES.a.11

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*, 1–27. https://doi.org/10.1145/3479544

Mogan, R., Fischer, R., & Bulbulia, J. A. (2017). To be in synchrony or not? A meta-analysis of synchrony's effects on behavior, perception, cognition and affect. *Journal of Experimental Social Psychology*, *72*, 13–20. https://doi.org/10.1016/j.jesp.2017.03.009

Montoya, M. M., Massey, A. P., & Lockwood, N. S. (2011). 3D collaborative virtual environments: Exploring the link between collaborative behaviors and team performance. *Decision Sciences*, *42*(2), 451–476. https://doi.org/10.1111/j.1540-5915.2011.00318.x

Nokes-Malach, T. J., Meade, M. L., & Morrow, D. G. (2012). The effect of expertise on collaborative problem solving. *Thinking & Reasoning*, *18*(1), 32–58. https://doi.org/10.1080/13546783.2011.642206

Olson, G. M., & Olson, J. S. (2000). Distance matters. *Human–Computer Interaction*, *15*(2–3), 139–178. https://doi.org/10.1207/S15327051HCI1523_4

Pierce, J. L., & Jussila, I. (2010). Collective psychological ownership within the work and organizational context: Construct introduction and elaboration. *Journal of Organizational Behavior*, *31*(6), 810–834. https://doi.org/10.1002/job.628

Rannastu, M., Siiman, L. A., Mäeots, M., Pedaste, M., & Leijen, Ä. (2019). Does group size affect students' inquiry and collaboration in using computer-based asymmetric collaborative simulations?. *Advances in Web-Based Learning–ICWL 2019: 18th International Conference, Magdeburg, Germany, September 23–25, 2019, Proceedings*. 143–154. Springer, Cham. https://doi.org/10.1007/978-3-030-35758-0_14

Roberts, B. W., & Yoon, H. J. (2022). Personality psychology. *Annual Review of Psychology*, *73*(1), 489–516. https://doi.org/10.1146/annurev-psych-020821-114927

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

Sagnier, C., Loup-Escande, E., & Valléry, G. (2019). Effects of gender and prior experience in immersive user experience with virtual reality. *Advances in Intelligent Systems and Computing*, 972, 305–314. Springer, Cham. https://doi.org/10.1007/978-3-030-19135-1_30

Santoso, M., Wang, P., Han, E., & Bailenson, J. N. (2025). Conversational dynamics in social virtual reality: A large-scale, longitudinal study of speech acts and nonverbal behavior. *Computers in Human Behavior*, *170*, 1–13. https://doi.org/10.1016/j.chb.2025.108691

Saqr, M., Nouri, J., & Jormanainen, I. (2019). A learning analytics study of the effect of group size on social dynamics and performance in online collaborative learning. In *European Conference on Technology Enhanced Learning*, 466–479. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-29736-7_35

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, *1*(2), 353–367. https://doi.org/10.1111/j.1756-8765.2009.01024.x

Smith, H. J., & Neff, M. (2018). Communication behavior in embodied virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3173574.3173863

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

Tarr, B., Slater, M., & Cohen, E. (2018). Synchrony and social connection in immersive virtual reality. *Scientific Reports*, *8*(1), 3693. https://doi.org/10.1038/s41598-018-21765-4

Thom-Santelli, J., Cosley, D., & Gay, G. (2010). What do you know? Experts, novices and territoriality in collaborative systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1685–1694. https://doi.org/10.1145/1753326.1753578

Van Dyne, L., & Pierce, J. L. (2004). Psychological ownership and feelings of possession: Three field studies predicting employee attitudes and organizational citizenship behavior. *Journal of Organizational Behavior*, *25*(4), 439–459. https://doi.org/10.1002/job.249

Wallot, S., Mitkidis, P., McGraw, J. J., & Roepstorff, A. (2016). Beyond synchrony: Joint action in a complex production task reveals beneficial effects of decreased interpersonal synchrony. *PLoS One*, *11*(12), e0168306. https://doi.org/10.1371/journal.pone.0168306

Wang, P., Santoso, M., Han, E., Srirangarajan, T., & Bailenson, J. (2026). Virtual Placemaking: Self-built environments and revisiting shared memories in virtual reality increase group cohesion. *Journal of Computer-Mediated Communication*, *31*(1), zmaf023. https://doi.org/10.1093/jcmc/zmaf023

Wang, P., Han, E., Queiroz, A., DeVeaux, C., & Bailenson, J. (2025). Predicting and understanding turn-taking behavior in open-ended group activities in virtual reality. *Proceedings of the ACM on Human-Computer Interaction*, *9*, 1–40. https://doi.org/10.1145/3757498

Wang, P., Miller, M. R., Han, E., DeVeaux, C., & Bailenson, J. N. (2024). Understanding virtual design behaviors: A large-scale analysis of the design process in Virtual Reality. *Design Studies*, *90*, 101237. https://doi.org/10.1016/j.destud.2023.101237

Zheng, C., & Tversky, B. (2024). Putting it together, together. *Cognitive Science*, *48*(2), e13405. https://doi.org/10.1111/cogs.13405

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information