Investigating How Physical Surfaces Can Serve as Common-Region Cues for Perceptual Grouping of Virtual Elements in Augmented Reality

Xuanhui Yang
The Hong Kong University of Science and Technology
Hong Kong, China
xyangcx@connect.ust.hk

Xuning Hu
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
xuninghu@hkust-gz.edu.cn

Hai-Ning Liang
Computational Media and Arts Thrust
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, Guangdong, China
hainingliang@hkust-gz.edu.cn

Xiaojuan Ma
The Hong Kong University of Science and Technology
Hong Kong, China
mxj@cse.ust.hk

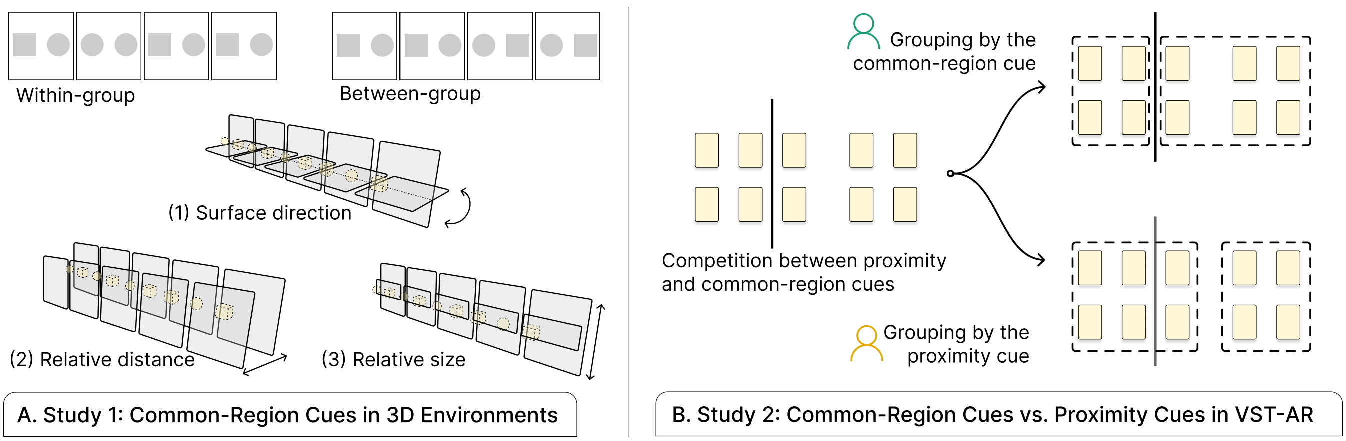


Figure 1: We conduct two user studies to investigate: (A) How can the surfaces serve as common-region cues that influence the perceptual grouping of the objects in 3D? (B) How do different grouping cues, such as common region and proximity, interact and affect users' perceptual grouping of virtual content in AR? Differences in users' physical environments and preferences for grouping principles can result in inconsistent perceptions and potential misinterpretations of the grouping information. For example, one user can group virtual elements based on the physical layout, using separated physical regions as grouping cues, while another user may not consider the use of physical context, leading to a different interpretation.

Abstract

Perceptual grouping enables people to organize elements into units according to intrinsic (e.g., proximity) and extrinsic (e.g., common region) principles. However, the role of physical surfaces as extrinsic grouping cues for virtual elements in Augmented Reality (AR) remains unclear. To provide a deeper understanding, we conducted two within-subject studies. The first study (N = 24) using repetition discrimination tasks revealed that surfaces can be common-region cues in 3D, with their influence depending on their distance to target objects along the viewing direction. Building on these findings, the second study (N = 24) employed both objective and subjective measures to capture the interaction between proximity and common-region cues in AR. Results indicate that competing cues reduce group clarity. They also enable us to distill people's strategies for improving the clarity by leveraging their physical and virtual environments. Finally, we propose design recommendations for future AR systems in assisted grouping tasks.

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3790315

CCS Concepts

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

Keywords

Perceptual Grouping, Common Region, Physical Context, Augmented Reality, User Study

ACM Reference Format:
Xuanhui Yang, Xuning Hu, Hai-Ning Liang, and Xiaojuan Ma. 2026. Investigating How Physical Surfaces Can Serve as Common-Region Cues for Perceptual Grouping of Virtual Elements in Augmented Reality. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3772318.3790315

1 Introduction

People often engage in tasks involving forming and recognizing groups among digital artifacts. For instance, during brainstorming sessions, similar ideas can be clustered by positioning the e-notes they are written on in close proximity [43, 77]. Likewise, when planning a conference, presentations are organized into sessions by arranging their associated blocks within a common grid [49]. During such tasks, people undergo a cognitive process known as perceptual grouping [2, 80, 82], which allows them to determine which elements in the visual scene belong together as parts of higher-level perceptual units based on various cues like visual similarity, distance, and common region among individual items.

Studies show that competition exists between intrinsic and extrinsic grouping cues when they co-occur. For example, elements located within a shared area indicated by an external frame or shading tend to be perceived as belonging together, which is induced by the extrinsic cue of a common region [68]. It is powerful enough to overwrite other intrinsic grouping principles, which are based on built-in properties of the elements, such as proximity and shape similarity [53, 60, 68]. Hence, it is a common practice to avoid the presence of potentially conflicting cues during grouping tasks [21, 52, 53].

When people bring work into the augmented reality (AR) environment to enjoy situated anchors and natural spatial interactions [54, 55], the situation gets complicated. In AR environments, users often use their physical surroundings to position virtual elements. Previous research indicated that users typically rely on the surfaces and/or the boundaries of physical objects as spatial affordance when placing virtual content [8, 55]. This suggests that features in the physical environment can serve as a common region cue for grouping, helping users perceive virtual elements positioned within them as part of the same group. Palmer [68] postulated that grouping by common region is influenced by perceived depth relations between the elements and the enclosing regions. Prior studies also suggested that perceptual grouping has effects on users' focus and memory of objects in 3D space [21, 22, 38]. However, it remains unclear to what extent the common region principle can be applied in 3D environments, particularly in AR contexts where physical and virtual elements coexist. There is an essential need to understand how physical contexts, such as surfaces, may be taken as external cues for the perceptual grouping of virtual artifacts in 3D environments. This knowledge can help in designing AR systems that better align with users' natural grouping perceptions, making effective use of environmental features to enhance grouping clarity or avoiding unintended confusions caused by the physical world.

To fill this gap, we conducted two studies to explore the effect of common region cues imposed by physical surfaces on the perceptual grouping of virtual objects in 3D spaces and whether spatial factors have an impact on this process. In particular, our first study (N = 24) adopted a type of repetition discrimination task (RDT) [61, 70, 79], a common experimental procedure to investigate perceptual grouping. As demonstrated in Figure 1 A, we manipulate various spatial factors of the surface. The task required participants to identify the repeated pair of elements when they were enclosed within the same or different surfaces. By using well-controlled stimuli in a virtual reality (VR) environment, we can access the specific contributions of different factors on perceptual grouping, minimizing confounds present in complex real-world scenarios. Our results showed that participants' response time (RT) is affected by the spatial relationship between the repeated pair of elements and the surfaces, suggesting that the common region principle was in force within certain 3D environments. This finding leads to a further question about its interaction with other coexisting grouping cues.

Prior research on perceptual grouping demonstrated that conflicting cues can weaken the grouping, leading to unstable perception or variations in grouping outcomes across individuals [52, 53, 59]. In an AR environment, extrinsic grouping cues provided by physical context may not align with intrinsic cues, such as proximity, especially when virtual content is not optimally positioned according to the physical context. As shown in Figure 1 B, such conflicts may occur when users observe grouping results made by others without knowing the intended grouping principle. These scenarios motivate us to explore how people perceive grouping when different perceptual cues interact in AR, either cooperatively or competitively.

Specifically, we conducted a second study (N = 24) to investigate how the co-presence of extrinsic cues from the physical context and intrinsic cues embedded within virtual elements influences users' perception and interpretation of grouping in AR. We compared different combinations of the intrinsic (proximity) and the extrinsic (common region) principles in a video see-through AR (VST-AR) environment, measuring participants' RT when one principle acted as the grouping signal and the other as a distractor. We also invited the participants to rate how difficult they found identifying the group and then re-arrange the visual items in the physical environment for better grouping clarity. We analyzed both quantitative and qualitative results to understand how cooperation, interference, and competition affect the two grouping mechanisms. Our findings indicate that competition between proximity and common region cues leads to challenging grouping situations. Participants adopted various strategies to enhance grouping clarity by leveraging both physical and virtual features.

To sum up, this paper makes the following contributions:

- We conducted a user study to investigate how the direction, relative distance, and size of physical surfaces influence perceptual grouping in 3D using a repetition discrimination task in a VR simulated environment.
- After confirming the grouping effects of 3D surfaces, we conducted a second user study to examine the impact and

user strategies related to the interaction between common region and proximity cues in an AR environment.
- Drawing on these findings, we discuss design recommendations for AR systems to better leverage physical environments as grouping factors and to mitigate ambiguities in perceptual grouping resulting from multiple principles.

2 Related Work

2.1 Perceptual Grouping

Perceptual grouping [80] refers to the cognitive process by which people organize discrete visual elements into higher-level units. Wertheimer's early work [82] introduced several key principles, such as proximity and similarity, that describe how spatial and visual relationships shape grouping. Later, Palmer expanded this framework with the principle of common region, emphasizing that elements enclosed within the same regions are likely to be perceived as a group [68]. These principles are often classified into intrinsic and extrinsic categories [69]. Intrinsic principles stem from visual attributes of the elements themselves, such as proximity and shared visual characteristics like color or shape (similarity). Extrinsic principles depend on the configuration of external structures, such as enclosing regions (common region) [68] or connecting lines (connectedness) [67] that impose organization across separate elements. These foundational principles of perceptual grouping play a critical role in interface design [5], interaction design [39, 45], and data visualization [2, 72], where designers and researchers leverage grouping cues to structure information, guide user attention, and improve usability.

Some studies have demonstrated that people perceive these cues with varying strength. For instance, the common region has the tendency to override similarity or connectedness [53, 61, 68]. Prior studies offer objective methods for evaluating these effects. In RDT [3, 61, 70, 79], participants are asked to identify repeated shapes from a row of objects with alternating shapes, each presented with various grouping cues. Notably, participants tend to respond more quickly to items within the same perceptual group, while responses slow down when items are separated across group boundaries. These patterns provide objective evidence of implicit grouping influences. Other studies combine multiple cues in cooperative or competitive configurations. Participants are then asked to determine the group based on a specific cue, and differences in response time reveal which cue dominates perception [52, 53, 59, 60]. For example, Luna et al.'s studies [52] suggest that the competition of the extrinsic principle of common region and intrinsic principles, such as proximity or similarity, leads to ambiguous grouping results, and the priority of those principles is subject to individual differences. Taken together, these findings highlight how subtle visual features guide perceptual grouping in 2D and establish methodological foundations for our investigation of physical environments as extrinsic grouping cues in AR.

2.2 Grouping in 3D Interaction Spaces

Grouping plays an important role in sensemaking tasks, where people organize diverse information into meaningful clusters [15, 81]. Prior research has investigated user strategies for sensemaking tasks and highlights the advantages of utilizing both large displays [1, 33, 48, 49, 75] and 3D environments [12, 47, 48, 50, 54, 56] for these activities. Their results on document and image arrangement show that, beyond semantic meaning, users intuitively apply Gestalt principles, especially proximity and common region principles, to group related items. Proximity consistently emerges as a primary organizing pattern, with participants placing elements close together to form groups [10, 81]. Some systems support this process by automatically clustering items based on proximity [78]. The environment itself also plays a crucial role in organizing groups, which aligns with the principle of common region. For instance, documents with different themes are often placed on separate walls in a virtual room [47]. Users may align virtual notes with familiar locations in reconstructed environments [65], and use furniture as natural boundaries for grouping in AR [55, 56]. These behaviors underscore the important interplay between spatial context and group organization.

Despite these empirical observations, only a handful of studies have directly examined the influence of grouping cues in 3D interaction spaces. For example, users are more easily distracted by objects sharing a shape or close spatial region [38], suggesting a strong role for perceptual grouping in 3D. Moreover, memory performance is higher for items located on the same wall than for items separated by a corner, indicating that the common region cue in 3D environments can meaningfully shape learning processes [21, 22]. Nevertheless, it remains unclear how strongly 3D environments shape users' understanding and perception of the grouping of the virtual elements in AR.

2.3 The Role of Physical Environments for Immersive Interaction

In augmented and mixed reality (AR/MR), the physical environment enables the placement of virtual elements [6, 17, 26, 36, 37, 55, 73]. Consequently, designing environment-aware virtual content placement becomes a critical consideration. Previous research has investigated approaches such as accounting for physical space constraints [20, 63], anchoring digital elements to nearby surfaces [8], respecting semantic associations [6, 7, 51], and adhering to social norms by avoiding occlusion of people and key physical contexts [13, 44]. Collectively, these strategies improve usability and legibility across a range of spatial settings [6, 7, 16], emphasizing that the physical environment should be viewed not merely as a constraint but also as a resource for creating effective interactions. Moreover, leveraging the physical environment presents additional challenges in remote collaborative contexts, where shared content must be synchronized across heterogeneous spaces [85]. To address these challenges, systems have allowed remote users to make use of their own environments by spatially mapping similar objects within different rooms (e.g., whiteboards, tables, and chairs) [24, 25, 36, 37], generating mutual shared spaces [29, 32, 35, 65, 66], redirecting virtual elements according to each physical environment [18, 76], or establishing shared spatial anchors independent of local layouts [30, 31, 62].

Despite the recognized importance of environment-aware design for usability and collaboration, less attention has been paid to how physical environments serve as perceptual grouping cues that shape users' understanding of groups within AR/MR. When the grouping principles of virtual elements are not aligned with those implied by physical structures, users may face ambiguity or misinterpretation. Thus, our objective is to investigate how environmental features function as grouping cues and interact with intrinsic grouping cues, providing insights for future research on leveraging physical environments in AR/MR.

3 Study 1: Common Region Cues in 3D Environments

This first within-subjects study examined whether common region cues induced by external surfaces in 3D spaces affect people's perceptual grouping of virtual objects, and whether various spatial properties of the surfaces contribute to this process. We employed RDT, a widely used method in psychology that assesses the influence of different factors on grouping strength by measuring participants' RT in identifying a repeated object pair [70]. Although the task is abstract and does not directly ask participants to group items, it enables an objective comparison of how spatial factors affect basic group perception and reduces the impact of other confounding factors.

3.1 Apparatus

The experiment was conducted in a VR-simulated environment to enable better control of variables and prevent interference from other factors, in line with previous work [41, 74]. The scene was developed using Unity 2022.3.53f1, with the Meta Quest Pro as the head-mounted display (HMD). The experiment settings, procedures, and participants' responses are managed and logged using Unity Experiment Framework (UXF) [4]. Participants had to remain seated in a fixed position throughout the experiment, observe the stimulus in a virtual environment that contained only an empty table and a blank wall, and respond by pressing the designated buttons on the controllers using both thumbs. All experimental tasks were executed on a high-performance PC equipped with an Intel Core i9 processor and an NVIDIA GeForce RTX 4060 GPU to maintain stable frame rates and real-time stimulus presentation.

3.2 Stimuli and Experimental Setup

As shown in Figure 2, each trial displayed a row of nine objects along with an array of five surfaces in the simulated environment. The objects alternated between cubes and spheres, except for a single adjacent pair of the same shape (the target pair). Among the nine objects, the target pair could appear at four random positions in the center (objects 3–4, 4–5, 5–6, or 6–7, counted from the left), while ensuring the periphery remained populated. The task for participants is to determine the shape of the target pair as quickly as possible while maintaining high accuracy.

The position and size of the spheres and cubes stay constant throughout the entire experiment. They were placed 65 cm from the participant and 15 cm below eye level, simulating interaction range within arm's reach and keeping targets within 15° below the line of sight to reduce neck strain [71]. At this viewing distance, each sphere or cube was 3 cm wide and spaced 5 cm apart. The spacing ensured visible surface boundaries between objects and the total width is limited to the natural range of eye movement in

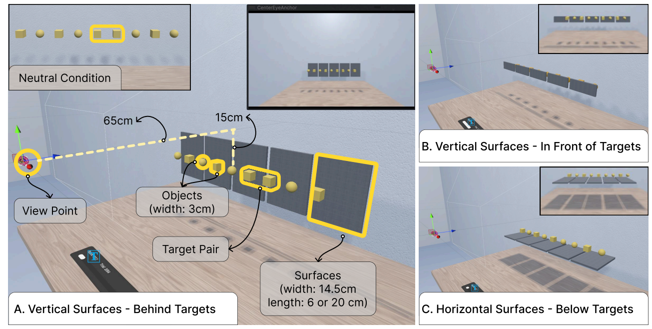


Figure 2: Experimental setup and stimuli for Study 1. Participants viewed a row of nine objects (alternating spheres and cubes, with one adjacent repeated pair as the target) presented along with the surfaces of varying direction, placement, and size. (A) Vertical surfaces positioned behind the target objects. The viewpoint (65 cm viewing distance, 15 cm below eye level), object dimensions (3 cm width), and surface width (14.5 cm) were controlled across conditions. (B) Vertical surfaces positioned in front of the targets, with the targets remaining visible to participants. (C) Horizontal surfaces positioned below the targets.

the horizontal direction (40°–45°) [19]. To appropriately cover two objects, the width of the surface (along the x-axis) was 14.5 cm, and the separation between the surfaces was 1.5 cm. The length of the surfaces is one of our manipulation factors, as specified in subsection 3.3. To simulate AR scenarios in which virtual objects stay visible despite being occluded by physical objects [40, 58], the shaders of the foreground objects were altered to render these elements in front of other geometries, thereby maintaining their visibility even when their actual position was behind the surface (Figure 2 B).

3.3 Study Design

We employed a within-subjects design with four independent variables (IVs), focusing on the potential impact of surface properties related to 3D spaces (see Figure 3 and Figure 4). The first IV is the grouping types of the target pair, used to test whether surface cues affect perceptual grouping. This factor had two levels: *within-group* (target pair projected on the same surface), or *between-group* (target pair on different surfaces). We also included a *neutral* condition without the group-indicating surfaces. The second IV is the surface directions (*vertical* or *horizontal*). Direction is one of the key properties in 3D space [23], and previous studies show that users tend to attach objects to both vertical and horizontal surfaces as displayable or placeable affordances [37]. The third and forth IVs are the relative distance and size of the surfaces compared to the target pair, because prior research indicates that the depth and size of regions can influence the strength of grouping caused by common region cues [68]. The levels of surface distance and size were determined from a pilot study (N=7) to ensure enough variety without too many trials. Distance indicates where objects were attached to the surfaces (0 cm), near (±5 cm), or far (±10 cm) from surfaces on both sides. More specifically, distance refers to changes in the depth of vertical surfaces and the height of horizontal surfaces, and negative values indicate surfaces before or above the target. Size reflects surface length, specified as tight (6 cm) or loose (20 cm) regions. Aside from the factors above, we controlled other surface features, such as texture, color, and illumination [23].

In the pilot study, participants reported that when two adjacent stimuli changed significantly, the resulting visual impact would delay their reaction. Therefore, we only display conditions with the same surface direction within each test block. Under each surface direction condition, the 21 different types of stimuli (2 grouping types × 5 distance levels × 2 size levels + 1 neutral condition) are combined with two types of target pair shapes, four possible target pair positions, and three repetitions. Thus, each participant observed a total of 1008 test trials, which were divided into 12 blocks of 84 trials each.

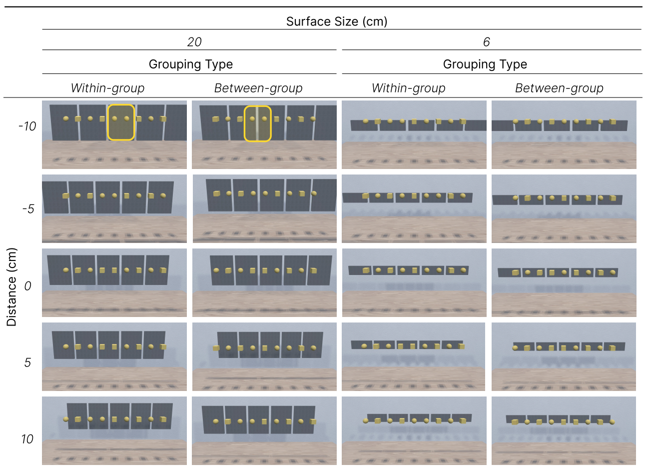


Figure 3: Experimental conditions (2 grouping types × 5 distance levels × 2 size levels) when the surfaces are vertical. The two images in the upper left corner highlight the target pair. Negative distances represent surfaces in front of the objects.

3.4 Participants and Procedure

We recruited 24 participants (15 females and 9 males) through university-wide advertisements from local universities. The average age was 21.5 years (SD = 1.8). Most of them are undergraduate or postgraduate students with diverse academic backgrounds, including engineering, mathematics, computer science, and design, and all have normal or corrected-to-normal vision. Most participants had prior VR experiences using an HMD: one participant reported using VR daily, three used it weekly, six used it monthly, 11 used it less than once a month, and only three participants reported that they had never used VR before.

After signing the consent form, participants completed the built-in nine-point eye-tracking calibration, and verified tracking quality using the headset's test scene. The main session comprised **16 blocks** in total: three preparatory blocks followed by one task-practice block, and then 12 test blocks.

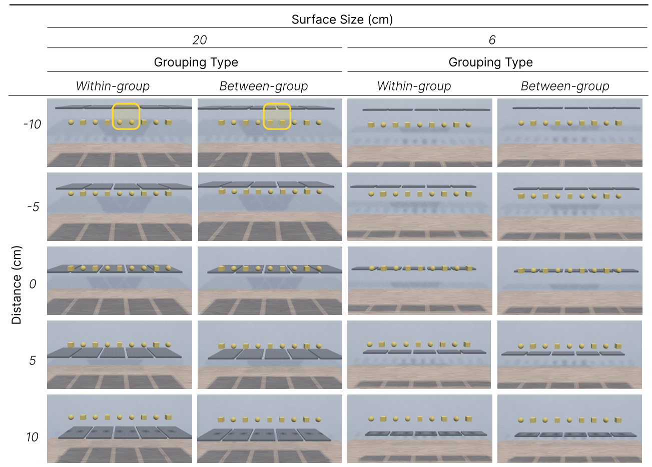


Figure 4: Experimental conditions (2 grouping types × 5 distance levels × 2 size levels) when the surfaces are horizontal. The two images in the upper left corner highlight the target pair. Negative distances represent surfaces above the objects.

Preparatory stage (Blocks 1–3). We gradually increase the difficulty of the tasks to make them easier for participants to learn. This stage was for participants to practice using the controllers and learn the response mapping between shapes and left and right controllers, e.g., the left controller button for a sphere and the right one for a cube. In the first two blocks, an object of the same shape appeared at a random time, while an object of random shapes appeared in Block 3. Participants were instructed to respond as quickly as possible and press the left or right controller button accurately corresponding to the object shape. The response mapping remained fixed for each participant throughout all the following trials and was balanced across participants.

Main stage (Block 4: practice, Blocks 5–16: test). Participants were required to specify the shape of the target pair among nine objects as quickly and accurately as possible. After each response, the stimulus disappeared, regardless of accuracy, and the next stimulus appeared after a one-second delay. The 12 test blocks implemented the factorial combinations specified in subsection 3.3. The order of surface directions was counterbalanced across participants, and trial order within each block was randomized to minimize order effects.

Participants could remove the headset, rest between blocks, and were reminded to maintain concentration. After completing all test blocks, they filled out a questionnaire rating the perceived impact of the experimental variables, and the NASA Task Load Index (NASA-TLX) [27]. The session lasted approximately **50 minutes**, and each participant received a 7 USD *gift voucher* as a token of appreciation.

3.5 Measurements

We employed a set of dependent variables encompassing both objective and subjective measures.

- **RT and dwell time:** RT was defined as the time (in seconds) from stimulus onset to the first button press on each trial. We also measured dwell time on the ground-truth targets until the response, to distinguish whether differences in participants' responses arose from the perceptual grouping of the target pair or from the delay in finding the target. The eye-tracking data was obtained via the integrated sensors of the Meta Quest Pro headset and Meta's Unity eye-tracking API[1].
- **Subjective feedback:** After completing all tasks, participants provided 5-point Likert-scale ratings regarding the perceived influence of identifying the target pair in different experimental conditions (1 = not at all, 2 = a little, 3 = moderately, 4 = very much, 5 = extremely). The detailed questionnaire items are reported in Appendix A. In addition, subjective workload of the experiment was assessed using the raw NASA-TLX questionnaire [27] after task completion.

4 Study 1: Results

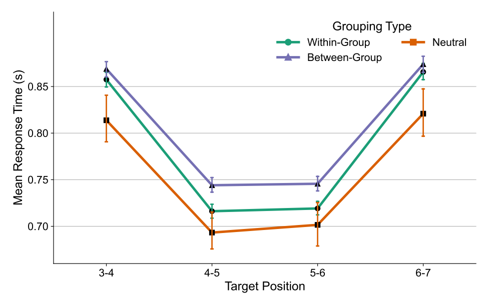
From the experiments, we collected a total of 24,192 trials (24 participants × 1008 trials, see subsection 3.3). For the analysis of RT and eye-tracking data, we removed three non-overlapping types of outliers: (1) the first trial in each block (288 trials, 1.19%); (2) error trials (536 trials, 2.22%); and (3) correct trials whose RTs fell outside a fixed range of 0.2–1.5 s (1007 trials, 4.16%). These exclusions follow common practice [53, 60] because such trials are unlikely to reflect typical behavior. Examples include unusually slow responses due to distraction, and overly fast or incorrect responses caused by accidental inputs). All subsequent analyses were conducted on the cleaned dataset (n = 22,361).

We first assessed the normality of all the data using the Shapiro-Wilk test and Q-Q plots. Since the results showed that the data violated the assumption of normality, we employed a Repeated Measures Aligned Rank Transformed ANOVA (RM ART ANOVA) using ARTool in R [84]. Following the suggestion of the previous study [42], we report effect sizes as partial eta-squared (η_p^2) for RM-ANOVA. The post-hoc analyses were carried out with Wilcoxon's signed-rank tests, and rank-biserial correlation values (r_b) were reported as effect sizes. Based on the rules-of-thumb [11], η_p^2 = 0.01/0.06/0.14 and r_b = 0.125/0.304/0.465 correspond to small, medium, and large effects, respectively.

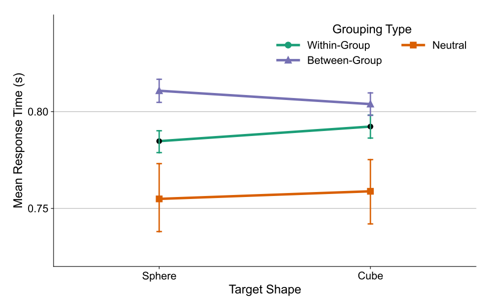
4.1 Response Time and Dwell Time

4.1.1 Grouping Type. First, we performed statistical analyses of the RT and dwell time under different grouping types (*within-group*, *between-group*, *neutral*) while ignoring other IVs. The results indicated a main effect of grouping type on both measures (RT: $F(2, 46) = 18.772$, $p < .001$, $\eta_p^2 = .417$, dwell time: $F(2, 46) = 7.371$, $p = .002$, $\eta_p^2 = .225$). Table 1-*Overall* shows the pairwise comparisons and the results indicate the similar pattern for RT and dwell time. Collapsing across surface properties, target positions, and shapes, *within-group* conditions yield significantly faster RT ($M = 0.747$, $SD = 0.071$) and shorter dwell time ($M = .418$, $SD = .051$) than *between-group* conditions (RT: $M = 0.772$, $SD = 0.081$, dwell time: $M = .433$, $SD = 0.047$). The *neutral* condition showed the fastest RT ($M = .721$, $SD = .076$) and the shortest dwell time ($M = .412$, $SD = .045$). Pairwise comparisons showed that *neutral* differed significantly from *between-group* for both measures, while the difference between *neutral* and *within-group* was significant for RT but not for dwell time.

4.1.2 Target Position and Target Shape. Target position had a large effect on both RT and dwell time. As shown in Figure 5a, RT varied significantly according to target position ($F(3, 69) = 63.179$, $p < .001$, $\eta_p^2 = .729$), with repeated elements near the center of the visual field (positions *4–5* and *5–6*) detected more quickly than those in peripheral locations (positions *3–4* and *6–7*). Dwell time on the correct target likewise varied significantly by target position ($F(3, 69) = 33.467$, $p < .001$, $\eta_p^2 = .592$). Although the interaction between grouping type and target position reached statistical significance for RT ($F(6, 138) = 2.200$, $p = .046$, $\eta_p^2 = .084$), its effect size was moderate. In addition, no significant interaction with grouping type was observed for dwell time ($F(6, 138) = .144$, $p = .989$, $\eta_p^2 = .007$). Figure 5b shows the interaction between grouping type and target shape. There was no significant interaction between these two variables for RT ($F(2, 46) = 1.249$, $p = .296$, $\eta_p^2 = 0.056$). The absence of a significant interaction effect suggests that the influence of grouping type on RT and dwell time follows similar patterns across different target positions and shapes. Therefore, subsequent analyses were performed without splitting the trials by target position or shape.



(a) Grouping Type × Target Position



(b) Grouping Type × Target Shape

Figure 5: Mean Response Time (RT) as a function of Grouping Type, Target Position, and Target Shape.

[1] https://developers.meta.com/horizon/documentation/native/android/move-eye-tracking [Accessed: 2025-07-13]

Table 1: The Post Hoc Results of Response Time and Dwell Time by Grouping Type

	Response Time					Dwell Time				
Comparison	M_{diff}	SD_{diff}	z	p	r_b	M_{diff}	SD_{diff}	z	p	r_b
Overall										
W-B	-0.026	0.029	-4.228	< **.001**	-.987	-0.014	0.023	-2.580	**.015**	-.690
N-W	-0.026	0.038	-2.742	**.006**	-.640	-0.007	0.030	-1.516	.129	-.376
N-B	-0.052	0.059	-3.600	< **.001**	-.840	-0.021	0.031	-2.891	**.011**	-.718
Vertical surfaces [Distance(cm)]										
W-B [Overall]	-0.034	0.037	-4.228	< **.001**	-.986	-0.017	0.034	-2.491	**.012**	-.633
W-B [-10]	-0.045	0.051	-3.285	**.001**	-.767	-0.022	0.071	-1.689	.095	-.415
W-B [-5]	-0.567	0.054	-4.028	< **.001**	-.940	-0.017	0.053	-1.829	.067	-.427
W-B [0]	-0.045	0.051	-3.571	< **.001**	-.833	-0.029	0.060	-2.615	**.009**	-.636
W-B [5]	-0.029	0.037	-2.086	**.037**	-.487	-0.009	0.065	-0.243	.807	-.058
W-B [10]	-0.010	0.030	-1.314	.189	-.306	-0.005	0.062	-0.400	.689	-.099
Horizontal surfaces [Distance(cm)]										
W-B [Overall]	-0.012	0.026	-2.385	**.017**	-.556	-0.013	0.028	-2.379	**.017**	-.604
W-B [-10]	-0.001	0.049	0.057	.954	-.013	-0.008	0.049	-0.852	.394	-.202
W-B [-5]	-0.022	0.048	-2.200	**.027**	-.513	-0.010	0.053	-0.654	.512	-.156
W-B [0]	-0.018	0.064	-1.142	.253	-.267	-0.022	0.049	-2.191	**.028**	-.521
W-B [5]	-0.014	0.053	-1.514	.129	-.353	-0.011	0.055	-0.471	.637	-.114
W-B [10]	-0.015	0.054	-1.314	.189	-.307	-0.015	0.064	-0.989	.323	-.236

W: Within-group, B: Between-group, N: Neutral.

4.1.3 Surface Direction. We conducted further analyses to explore how surface properties influence the strength of grouping caused by common region cues in 3D space, thus excluding *neutral* trials where no surfaces were present (21,272 trials remaining). The interaction between grouping type and surface direction was significant regarding both RT ($F(1, 23) = 9.316$, $p = .005$, $\eta_p^2 = .396$) and dwell time ($F(1, 23) = 6.684$, $p = .016$, $\eta_p^2 = .225$). As shown in Figure 6, within-group trials elicited significantly shorter RT than between-group trials when the surface direction was *vertical*. A similar, though smaller, effect was also observed for *horizontal* surfaces. Therefore, the results suggest that the surfaces influence perceptual grouping under both directions, albeit stronger in the *vertical* case.

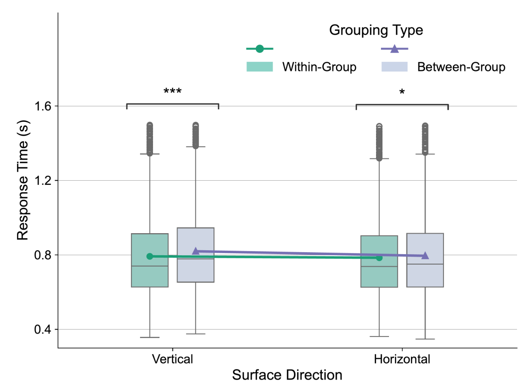


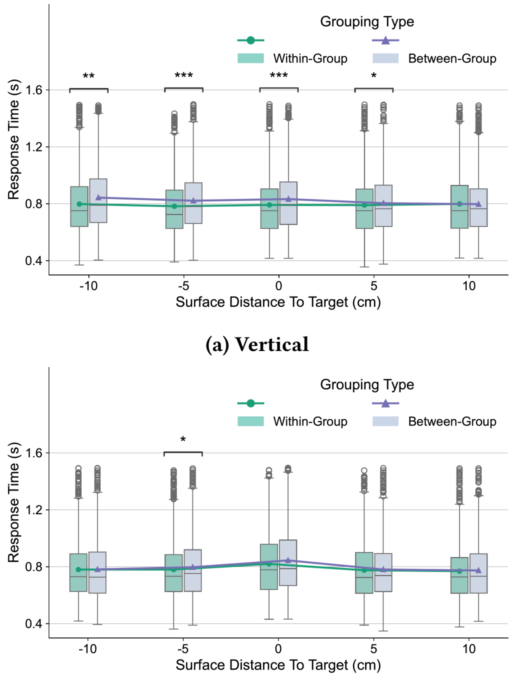
Figure 6: Response Time (RT) by Grouping Type and Surface Direction. $p < .05^{*}$; $p < .01^{**}$; $p < .001^{***}$

4.1.4 Surface Distance and Size. We tested the effects of surface distance and size separately for each surface direction. In the vertical direction, the interaction between grouping type and distance had a significant effect on RT ($F(4, 92) = 6.349$, $p < .001$, $\eta_p^2 = .211$) (Figure 7a), but it did not significantly affect dwell time ($F(4, 92) = 1.057$, $p = .382$, $\eta_p^2 = .043$). As shown in Table 1-*Vertical surfaces*, surfaces placed in front of the target (negative distance) or at the same distance enhanced their effects on grouping perception. These result in significantly larger RT differences between *within-* and *between-group* conditions than surfaces placed behind the virtual items (positive distance). In contrast, as shown in Table 1-*Horizontal surfaces* and Figure 7b, the grouping effects in the horizontal direction showed only small differences across distance levels for RT ($F(4, 92) = 1.112$, $p = .355$, $\eta_p^2 = .061$), as well as for dwell time ($F(4, 92) = .973$, $p = .426$, $\eta_p^2 = .041$). Surface size did not interact significantly with grouping type to affect RT in either direction (all $p > .05$).

4.2 Subjective Feedback

The subjective questionnaire primarily examined how the variables we controlled influenced participants' subjective perceptions. The scores are on a 5-point scale; the smaller the score, the less influenced it is.

Participants' subjective ratings indicated that *target position* significantly affected their responses ($F(2, 46) = 31.81$, $p < .001$, $\eta_p^2 = .547$). As shown in Table 2, the *center* position ($M = 1.42$,



(a) Vertical

(b) Horizontal

Figure 7: Response Time (RT) by Grouping Type and Surface Distance. $p < .05^{*}$; $p < .01^{**}$; $p < .001^{***}$

SD = 0.88) was rated significantly less influenced than both the *left* (*M* = 2.63, *SD* = 1.06) and *right* positions (*M* = 2.83, *SD* = 0.96). No significant difference was observed between the left and right positions.

For grouping type, a significant effect was observed under both viewing directions (Table 2), with *within-group* arrangements (V: *M* = 1.42, *SD* = 0.93, H: *M* = 1.54, *SD* = 0.78) were rated significantly smaller than *between-group* (V: *M* = 3.00, *SD* = 1.14; H: *M* = 2.54, *SD* = 1.02). By contrast, neither surface distance nor surface size had significant effects on participants' ratings under either surface direction (distance: vertical $p = .343$, horizontal $p = .361$; size: vertical $p = .419$, horizontal $p = .080$).

Additionally, the NASA-TLX scores reflected a moderate level of task load (*M* = 31.25, *SD* = 12.57) [28], supporting that the task demand was sufficient to maintain participant engagement and data reliability throughout the experiment.

4.3 Summary

In summary, both objective and subjective measures show that surfaces can serve as common region cues in 3D space within our experimental settings. Notably, this grouping effect sometimes occurs without users' conscious awareness. The grouping cue is strong enough when surfaces interact with or are positioned in front of targets, particularly when the targets and surfaces are aligned along the viewing direction. In these cases, participants' responses are influenced even if they do not subjectively notice the distance difference. In our task, since participants viewed the object from the side, vertical surfaces – aligned with the viewing axis – had a slightly stronger effect on grouping than horizontal surfaces. The results regarding surface direction may be reversed under a top-down view. We further discuss the results in section 7.

Table 2: The Post Hoc Results of Subjective Rating by Target Position and Grouping Type

Comparison	M_{diff}	SD_{diff}	z	p	r_b
Target position					
Center - Left	-1.208	1.141	-3.580	< **.001**	-.895
Center - Right	-1.416	0.881	-4.102	< **.001**	-1.00
Left - Right	-0.208	0.977	-1.040	.298	-.378
Grouping type					
W-B (V)	-1.583	1.176	-3.911	< **.001**	-.904
W-B (H)	-1.000	0.722	-3.873	< **.001**	-1.00

W: Within-group, B: Between-group.
V: Vertical surface, H: Horizontal surface.

5 Study 2: Common Region Cues vs. Proximity Cues in VST-AR

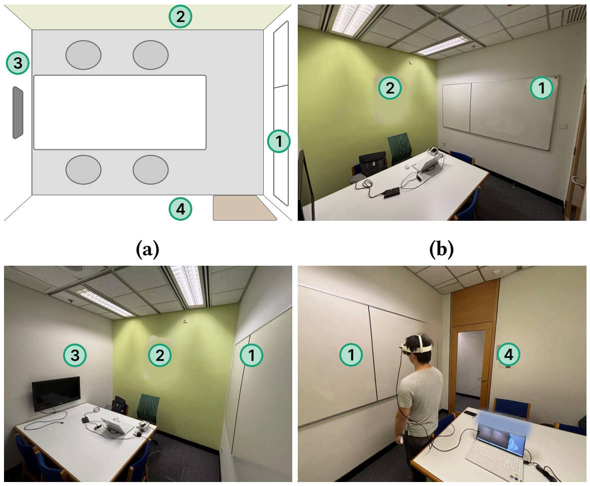
Study 1 demonstrated that surfaces can serve as grouping cues in certain 3D environments. Building on these findings, we conducted a subsequent controlled within-subjects study to examine how users perceive and react to grouped virtual cards when intrinsic (proximity) and extrinsic (common region) grouping cues are independent, cooperative, interfering, or competitive in the VST-AR environment. We devised three objective or subjective grouping-related tasks, simulating the process of users judging and adjusting group layouts in real-world tasks [82].

5.1 Apparatus

The experiment was conducted in a study room (2.5 × 3.5 *m*) with a whiteboard split into two uneven parts by a one-cm-wide black magnetic border, mimicking a typical work environment (see Figure 8). A high-performance laptop ran the Unity project (version 2022.3.53f1), streaming content in real time to a Meta Quest 3 HMD. Virtual content was seamlessly integrated with the real-world scene by enabling Passthrough on the HMD, resulting in an immersive VST-AR experience. A spatial anchor was predefined in the room along the partition border on the whiteboard to ensure precise and consistent alignment of virtual elements with the physical environment throughout the experiment.

5.2 Stimuli and Experimental Setup

Informed by the results of Study 1, we focused only on cases with a vertical background surface and had participants view from the side, ensuring virtual objects were projected naturally onto the physical surfaces along their viewing direction. We also ignored the impact of target position on perceptual grouping and created a simple layout using a small number of cards, as illustrated in Figure 9.



(a) (b)

(c) (d)

Figure 8: Experimental space of Study 2. (a) Top-down view of the study room. The room features a table in the middle, with each wall hosting a different element: (1) a whiteboard, (2) a colorful wall, (3) a screen, and (4) a door. (b) and (c) Different views of the study room. The whiteboard was split by a black magnetic border. (d) Participant standing in front of the whiteboard.

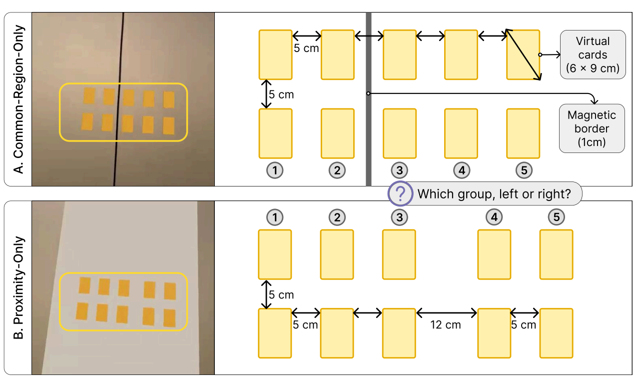
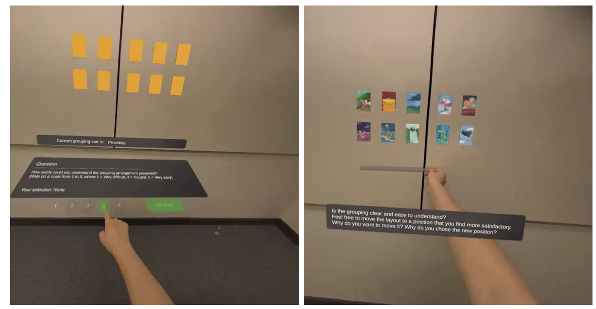


Figure 9: Stimuli and setup for Study 2. Participants viewed ten virtual cards arranged in a 2×5 grid within the VST-AR environment. Card placement and spacing varied by condition: (A) In the common-region-based condition, all cards were evenly spaced (5 cm apart) with a physical border 1 cm wide. (B) In the proximity-based condition, inter-group spacing was increased (12 cm), and a virtual background covered the physical border. Participants were asked to determine which group the central column belonged to.

The experimental stimuli consist of ten rectangular cards of a common size (6 cm wide × 9 cm high), arranged in a 2×5 grid. The five columns were divided into two groups from columns 2–3 or 3–4 (counting from the left). The primary task was to determine which group the central column belonged to based on a given grouping cue, similar to previous studies [52, 53, 60]. It was carried out in three stages as if participants had just joined a grouping task and needed to first understand the existing groups. **Stage 1**: participants determined the grouping as quickly as possible, focusing on a specific cue. **Stage 2** (Figure 10a): they rated the perceived difficulty of grouping for each stimulus. **Stage 3** (Figure 10b): they could freely reposition the entire set of virtual cards whenever grouping appeared ambiguous or unclear.



(a) Stage 2 (b) Stage 3

Figure 10: Experimental setup for (a) Stage 2: The participant selects a score for the presented condition, and (b) Stage 3: The participant rearranges the cards as needed. Images capture the participant's right-eye perspective in the AR environment.

The default horizontal and vertical spacing between all cards was 5 cm. Grouping was indicated by regions separated by a black border on the physical whiteboard or by increasing the gap between groups to 12 cm. In a pilot study (N=3), we confirmed that this setup made groupings distinguishable under single-cue conditions and created ambiguity when competing cues co-occurred, as expected based on a prior study [52]. The combinations of the two grouping conditions was explained in subsection 5.3. In the pilot study, we initially tried cards with random pictures to measure RT for group identification in Stage 1. However, one participant misunderstood the task, spending time carefully observing the card content instead of proximity and common region cues. To eliminate visual confounds when measuring RT, all cards in Stage 1 and the first half of Stage 2 were painted in an identical color. To make the task more engaging and reduce abstraction, we covered the cards with colorful images selected from the board game Dixit[2] during the second half of Stage 2 and Stage 3.

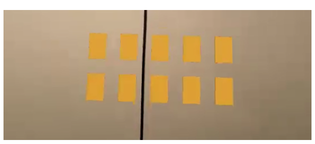
Virtual cards were placed at predetermined locations in the physical environment using a spatial anchor. The bottom and top of the display area were about 140 and 163 cm above the floor. This arrangement ensured that all virtual content could be viewed as on the whiteboard, and that the black border of the whiteboard consistently appeared at the intended spacing. At the start of the experiment, participants were guided to stand about 80–90 cm away from the whiteboard. They could slightly adjust their position to optimize their view of the virtual overlays and the physical whiteboard. They remained stationary in Stages 1 and 2 and were permitted to move around the room in Stage 3. Since all grouping manipulations

[2] https://www.libellud.com/en/our-games/dixit/

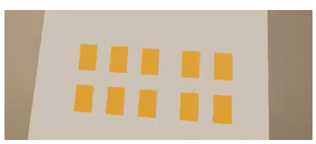
took place along the horizontal axis, small differences in viewing angles caused by participant height were not expected to influence their perceptual grouping.

5.3 Study Design

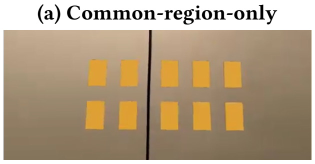
Stage 1 and **Stage 2** aimed for direct comparison of the strength of different grouping cues by analyzing RT and subjective ratings, targeting three independent IVs as within-subject factors. The first IV is attention, the designated cue used to determine the group (*proximity-based* or *common-region-based*). The second IV is stimulus types that represents single or combined grouping cues with different relationships (see Figure 11). In common-region-based blocks, three stimulus types were included: *common-region-only*, *cooperation* – two grouping cues indicate the same grouping, and *competition* – the proximity cue suggesting opposite grouping of the central column. This structure was mirrored in proximity-based blocks, with *proximity-only* as the single grouping cue and the physical regions serving as either *cooperation* or *competition* indicators. Additionally, a new condition – *interference* – was included in proximity-based blocks to reflect situations where virtual contents are placed on the border of the whiteboard as if the system did not consider physical context features. The last IV is distance between the virtual cards and the physical whiteboard (0 cm or 10 cm). Distance levels were established based on findings from Study 1, which showed that grouping by common region cues is strong enough when targets are at the same depth as the surfaces and weakens when objects are positioned 10 cm in front of the surface.



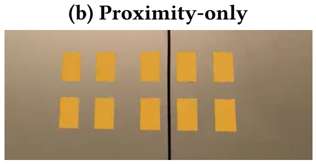
(a) Common-region-only



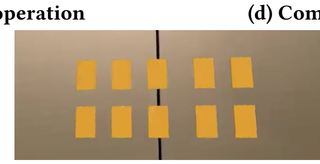
(b) Proximity-only



(c) Cooperation



(d) Competition



(e) Interference

Figure 11: Five stimulus types in study 2 from the participant's first-person perspective (views in the right eye).

Stage 3 is intended to encourage participants to use their subjective judgment to optimize the perceptual clarity of the groupings. It only includes two within-subject factors: five stimulus types (Figure 11) and two distance levels. In this stage, participants were not given a specific cue to attend to. Instead, they were asked to interpret and adjust the display based on their own understanding of the stimulus.

In summary, this study design resulted in a total of 84 trials ((3 + 4) stimulus types × 2 distance levels × 3 within-block repetition × 2 attention repetition) in **Stage 1**, 28 trials ((3 + 4) stimulus types × 2 distance levels × 2 card content conditions) in **Stage 2**, and 10 trials (5 stimulus types × 2 distance levels) in **Stage 3** for each participant.

5.4 Participants and Procedure

We recruited 24 additional participants (13 females and 11 males) for this second study, with no overlap with Study 1 and an average age of 24.7 years (SD = 4.0). All participants have normal or corrected-to-normal vision and were recruited from a local university through online advertisements, with diverse backgrounds in fields such as computer science, chemistry, ocean science, and art. Most participants had prior experience with immersive technologies (VR and AR HMDs). For VR, 18 reported using it less than once a month, three used it weekly, and only three had never used VR before. For AR, 16 participants used it no more than once a month, two used it weekly, and six had no experience with AR.

After signing the consent form, participants were guided to the designated location after they could see the real environment through the HMD. Each experimental session consisted of a preparatory stage, three task stages, and a post-experiment semi-structured interview stage. Participants were allowed to rest between stages. The session lasted about **50 minutes** (30 minutes in AR and 20 minutes for the interview), and each participant received a 7 USD *gift voucher* as a token of appreciation.

Preparatory stage (Blocks 1–2). This stage allowed participants to get familiar with using the controllers. In each trial, a single card appeared randomly on the left or right side of the whiteboard for participants to learn the mapping between selections and controller buttons. Since the selections correspond naturally to the left and right controllers, fewer preparation blocks were used in this study compared to the beginning of Study 1.

Stage 1 (Blocks 3–4: practice, Blocks 5–8: test). Prior to each block, participants were informed of the cue to attend to (fixed throughout the block) and instructed to quickly identify which side the central column belonged to according to the attention. Similar to Study 1, each trial disappeared after a response, with the next trial appearing one second later. The order of two types of attention was balanced across participants and trials were randomized in each block to reduce order effects. Based on experience from Study 1, a pre-trial was added at the beginning of each test block, as prolonged RT was typically observed in the first trial.

Stage 2 (Block 9: practice, Blocks 10–13: test). Participants rated the difficulty they experienced in determining groupings based on the assigned cue across various stimulus and distance conditions. In this stage, participants used their free hands to press virtual rating buttons in midair for a natural interaction [57]. They were instructed to respond at their own pace and maintain a consistent scoring standard. A stimulus disappeared only after people pressed the 'confirm' button. Cards in blocks 10–11 were identical in color, while featured random colorful images in blocks 12–13 as preparation for the next stage.

Stage 3 (Block 14: practice, Block 15: test). At this stage, participants were instructed to improve the visibility of image-card groupings based on their own understanding without being told the underlying grouping principles. They could relocate the virtual cards as a whole anywhere in the experiment space by pinching a gray bar. During these adjustments, participants were asked to explain the rationale behind their actions.

Interview Stage. After completing the tasks, participants removed the headset and took part in a semi-structured interview. Participants were asked about their impressions of the competing cues and their personal preference for using *proximity* and *common region* principles in determining groups. Participants were then asked to envision how they might approach grouping tasks in AR environments under various scenarios. These included working alone (in single or multiple workspaces), and collaborating with others (co-located or remote). Insights from this stage will be briefly summarized in subsection 6.3 and discussed in subsection 7.2.

5.5 Measurements

The objective and subjective measures recorded in this study are as follows:

- **Response time (RT) – Stage 1:** RT was defined similarly in Study 1.
- **Subjective rating – Stage 2:** Participants rate *"How easily could you understand the grouping arrangement presented?"* in each trial, on a 5-point scale (1 = very difficult, 3= neutral, 5 = very easy).
- **User behavior – Stage 3:** Participants' first-person views were streamed from the HMD to the laptop and recorded via Zoom.

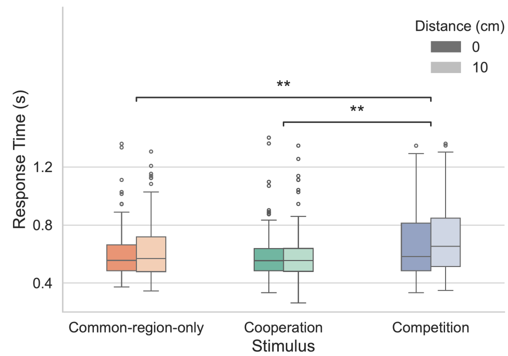
6 Study 2: Results

For both RT and subjective ratings, we applied statistical analyses. In **Stage 1**, we collected a total of 2016 trials (24 participants × 84 trials), where the first trial of each block is not counted as described in subsection 5.4. Consistent with Study 1, we further removed error trials (53 trials, 2.63%) and trials with RTs outside the range of 0.2–1.5 seconds (59 trials, 2.93%). After these exclusions, the final dataset comprised 1904 trials (94.44%) for subsequent analysis. In **Stage 2**, trials with cards of identical color (n=336) were analyzed separately from those with colorful pictures (n=336). As in Study 1, normality was violated, so we used the same analysis framework: RM ART ANOVA for main effects, followed by Wilcoxon signed-rank tests for post-hoc comparisons.

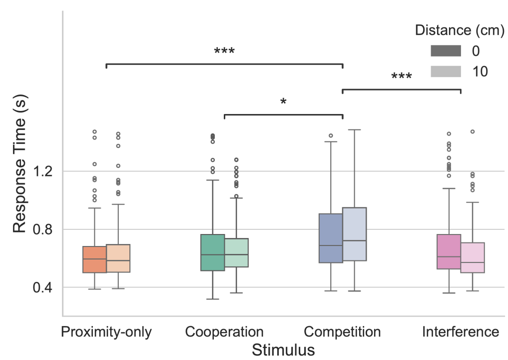
For the behavioral and audio data collected in **Stage 3** and during the interview, we conducted a thematic analysis. Two researchers independently performed open coding to identify patterns related to users' adjustment behaviors, their intentions, and descriptions of various conditions. The initial codes were iteratively refined into a shared codebook, ensuring consistency and capturing emerging themes. After the coding process, all authors discussed the findings together to reach a consensus on the results.

6.1 Response Time

6.1.1 Stimulus Type. First, RM ART ANOVA tests reveal that RT differed across different stimuli under both two types of attention: *common-region-based* ($F(2,46) = 17.173$, $p < .001$, $\eta_p^2 = .427$) and *proximity-based* ($F(3,69) = 12.939$, $p < .001$, $\eta_p^2 = .360$). The pairwise comparisons are listed in Table 3. When proximity and common region cues compete, RT is significantly slower than when a single cue is present or when both cues cooperate (see Figure 12). However, in the additional *interference* condition, RT was similar with the *proximity-only* and *cooperation* cases. Additionally, we tested the two single-cue conditions (*common-region-only* vs. *proximity-only*) and found no significant difference on RT ($M_{diff} = -0.042$, $SD_{diff} = 0.139$, $z = -1.257$, $p = .208$, $r_b = -.293$).



(a) Attention: Common-region-based



(b) Attention: Proximity-based

Figure 12: Response Time (RT) by Stimulus and Distance. $P < .05^{*}$; $P < .01^{**}$; $p < .001^{***}$

6.1.2 Distance. Response time was not significantly affected by distance ($F(1,23) = 1.130$, $p = .298$, $\eta_p^2 = .046$). However, its interaction with stimulus type was significant ($F(6,138) = 2.983$, $p = .009$, $\eta_p^2 = .114$), indicating that distance influenced RT only under certain stimulus conditions. Specifically, a significant difference was observed in the interference condition: when participants grouped the cards by proximity, RT at distance 10 cm was faster than at distance 0 cm ($M_{diff} = -0.063$, $SD_{diff} = 0.122$, $z = -2.428$, $p = .015$, $r_b = -.567$).

6.2 Subjective Ratings

We evaluated the two settings with different card content (*identical-color* and *random-figure*) separately. As shown in Figure 13, subjective difficulty ratings differed significantly across stimulus conditions in both attention conditions (*proximity-based*: $F(3,69) = 16.987$, $p < .001$, $\eta_p^2 = .428$; *common-region–based*: $F(2,46) =$

Table 3: The Post Hoc Results of Response Time by Stimulus Types

Comparison	M_{diff}	SD_{diff}	z	p	r_b
Attention: CR-based					
CR-Coop	−0.013	0.044	−1.199	.230	−.298
CR-Comp	−0.077	0.093	−3.423	**.001**	−.852
Coop-Comp	−0.090	0.109	−3.001	**.003**	−.748
Attention: Prox-based					
Prox-Coop	−0.319	0.893	−1.771	.115	−.413
Prox-Comp	−0.113	0.102	−3.971	< **.001**	−.927
Prox-Intf	0.001	0.082	−.141	.886	−.033
Coop-Comp	−0.081	0.141	−2.685	**.014**	−.627
Coop-Intf	0.033	0.108	1.428	.184	.333
Comp-Intf	0.115	0.090	4.114	< **.001**	.960

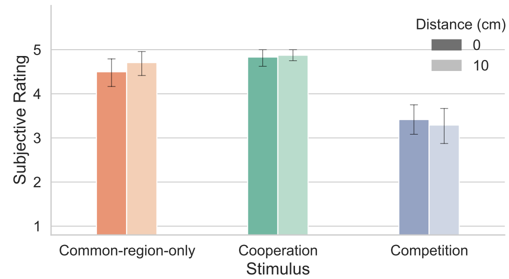
CR: Common-region-only, Coop: Cooperation,
Comp: Competition, Prox: Proximity-only, Intf: Interference.

59.416, $p < .001$, $\eta_p^2 = .721$). Pairwise comparisons, detailed in Table 4, show that both the *cooperation* and single cue conditions were rated as significantly easier than the *competition* condition under both attention types. Additionally, within the *proximity-based* block, the *cooperation* and *interference* conditions received similar ratings, both rated as more difficult than the *proximity-only* condition but easier than the *competition* condition. Furthermore, participants' ratings remained consistent even when the cards were filled with random figures (*proximity-based*: $F(3, 69) = 19.481, p < .001, \eta_p^2 = .459$; *common-region-based*: $F(2, 46) = 15.571$, $p < .001$, $\eta_p^2 = .403$), indicating the robustness of these subjective difficulty effects.

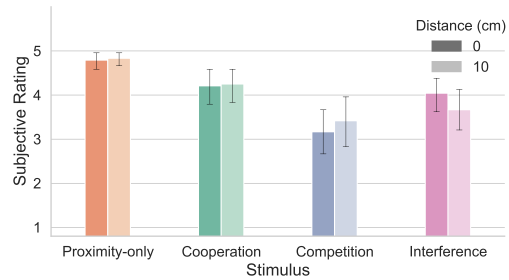
6.3 User Behaviors

We found that users employed various adjustment strategies to improve grouping clarity under different conditions of the grouping cues. Below, we summarize these strategies into five key themes that illustrate how participants actively leverage both virtual and physical elements of the environment to resolve uncertainty in our experimental setup.

6.3.1 No Adjustment Needed. When the grouping results were sufficiently clear, participants generally left the cards unmoved, which was most common in the *proximity-only*, *common-region-only*, and *cooperation* conditions. For example, some participants noted that "the virtual background covers the border on the whiteboard" (P13) and some explained "the cards are clearly separated by the black border" (P5). Even in conditions involving *interference* or *competition*, some participants reported no need to adjust the cards, as they chose to disregard one of the conflicting cues in favor of a single, dominant cue to guide their decisions. Several participants relied on the proximity cue. For example, P8 and P24 explained that the five columns are grouped by "the large spacing" and they can "ignore the physical context". In the *competition* condition, P2 observed that parallax shifts produced an "unstable relationship between the border and the virtual cards," and thus concluded that "the background doesn't matter" when the cards were not coplanar



(a) Attention: Common-region-based



(b) Attention: Proximity-based

Figure 13: Subjective ratings by stimulus type and card distance when all cards are the same color. Error bars indicate the 95% confidence interval (CI).

with the whiteboard. In contrast, other participants in the competition condition prioritized the common region cue. For instance, P13 stated that grouping by the background boundary is "the first impression." Moreover, P10 ignored the proximity factor "because the gap is not particularly large."

In some cases, under the *competition* condition, participants resisted making adjustments because multiple interpretations seemed equally valid, preferring not to make changes without additional context. As P2 explained, s/he would not "touch it rashly" without further contextual information. P8 explained the reason for no adjustment is that "either one of the principles is very clear and not difficult to utilize." P12 described a possible multi-level grouping: "First, . . . two groups according to the background. Then, the large group . . . is divided into two sub-groups according to the proximity."

6.3.2 Utilize Physical Environments. Participants frequently integrated physical features of the space to support grouping clarity. This included using surface boundaries, different colored surfaces, or areas at varying distances to segment regions.

Some participants assumed that grouping should follow proximity, and they often found background features to be distracting. As P10 and P22 put it, "The line in the middle of a group is distracting." To address this, some participants repurposed the border from an interference to a helpful cue. One reason for aligning two grouping principles was "to unify the results" (P3) and "dual confirmation . . . eliminates any ambiguity" (P13). Another explanation was to find more explicit divisions than proximity-only, as participants remarked, "find a more obvious boundary" (P14) and "physical divisions are stronger and more intuitive" (P6). Even subtle discomfort

Table 4: The Post Hoc Results of Subjective Rating by Stimulus Types

	Rating in *identical-color* blocks					Rating in *random-figure* blocks				
Comparison	M_{diff}	SD_{diff}	z	p	r_b	M_{diff}	SD_{diff}	z	p	r_b
Attention: CR-based										
CR-Coop	−0.250	0.552	0	**.039**	−.857	−0.083	0.319	0	.206	−.500
CR-Comp	1.250	0.737	4.078	**< .001**	1.00	0.708	0.896	3.179	**.001**	.897
Coop-Comp	1.500	0.847	4.156	**< .001**	1.00	0.792	0.919	3.427	**.002**	.963
Attention: Prox-based										
Prox-Coop	0.583	0.985	2.975	**.004**	.885	0.458	0.943	2.153	**.037**	.603
Prox-Comp	1.521	1.229	3.938	**< .001**	1.00	1.521	1.047	4.131	**< .001**	1.00
Prox-Intf	0.958	0.907	3.655	**< .001**	1.00	0.812	0.719	3.659	**< .001**	.971
Coop-Comp	0.937	1.209	3.061	**.004**	.776	1.062	1.288	3.269	**.002**	.772
Coop-Intf	0.375	1.209	1.339	.181	.347	0.354	1.108	0	.150	.404
Comp-Intf	−0.562	0.838	−2.695	**.008**	−.694	−0.708	0.833	−3.174	**.002**	−.869

CR: Common-region-only, Coop: Cooperation, Comp: Competition, Prox: Proximity-only, Intf: Interference.

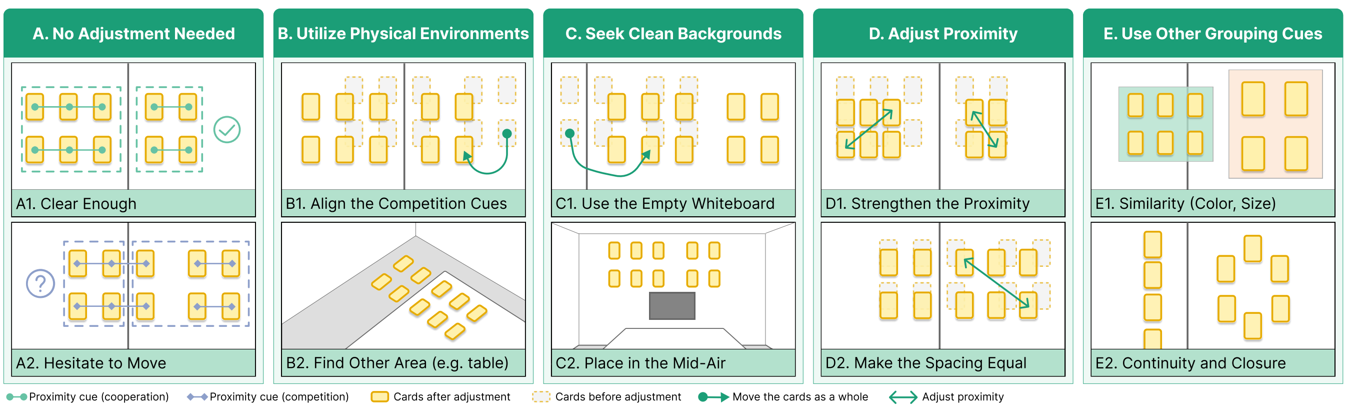


Figure 14: Summary of adjustment strategies observed in Stage 3. (A) No adjustment needed—either grouping was clear (A1) or ambiguous (A2). (B) Leveraging physical features to clarify groupings, such as (B1) using borders on the whiteboard that previously introduced competing cues, or (B2) referencing the table edge contrasted against the floor. (C) Moving cards to a clean background—for example, (C1) the empty whiteboard or (C2) open mid-air space. (D) Adjusting proximity cues in various ways, including (D1) decreasing intra-group spacing and increasing inter-group spacing, or (D2) eliminating proximity-based grouping by equalizing all spaces. (E) Integrating other grouping cues, such as (E1) color and shape similarity, (E2) continuity, and closure (e.g., forming a circle).

sometimes led to adjustments. P5 admitted to avoiding placing cards on the border, even though it did not affect her perception of grouping.

Beyond the border on the whiteboard, participants explored other environmental anchors. When spacing felt ambiguous, some participants relied on the boundary of the virtual background because s/he "need an external line to help with quick judgment" (P14), as illustrated in Figure 15 A. Others experimented with features in the room such as the table, TV screen, or door frame, citing color contrast or depth differences between regions as additional aids for grouping (P16, see Figure 15 B, C, and D).

6.3.3 Seek Clean Backgrounds. Another prominent strategy was to relocate cards to visually uncluttered areas. When cues conflicted or caused interference, some participants moved cards onto the blank whiteboard (Figure 15 E) or into empty mid-air (Figure 15 F) to "make it easier to see" (P18) and "eliminate the distraction" (P2). Some participants preferred arranging cards in mid-air against a neutral white wall because "it looks cleaner and more comfortable" (P14).

Interestingly, even in the *common-region-only* condition, which was designed as a simple grouping, several participants still moved the cards to uncluttered backgrounds. This is because some participants believe that equal spacing among the cards did not convey any intended grouping. For example, P18 assumed "a significant distance between groups, and the border may only serve to emphasize, rather than determine, that separation." In some trials, participants

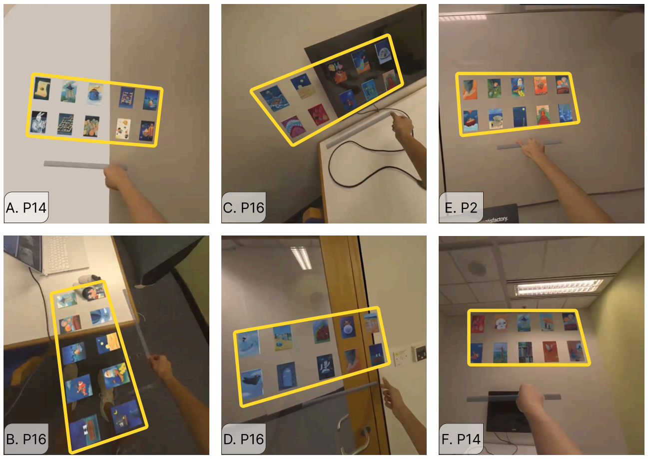


Figure 15: Participants' adjustment results. (A) Use the virtual background and the physical whiteboard. (B) Use the table edge. (C) Use the TV screen. (D) Use the door frame. (E) Place at the empty place of the whiteboard. (F) Place in the mid-air.

observed that the cards were slightly forward of the board, which they interpreted as a cue to rely less on physical context. As P2 remarked, "they don't seem to fit in with the background . . . but the boundary behind them may interfere with later tasks."

6.3.4 Adjust Proximity. Although the task instructed participants to manipulate the ten cards as a whole, several participants expressed a desire to fine-tune the arrangement by adjusting the spacing within and between groups. For example, in the *common-region-only* condition, participants described reducing spacing within the group and increasing between-group spacing (P7, P3). Alternatively, some participants preferred to adjust the cards to minimize proximity cues and emphasize grouping by the background instead. As P10 explained in a *competition* trial, "I might move [some cards] so that all the spacings are equal. Then, I can distinguish the groups based on the border."

6.3.5 Substitute and Compensate for Other Grouping Cues. Finally, participants sometimes referenced Gestalt principles beyond proximity and common region, turning to alternative cues to guide grouping decisions. (1) Color similarity: Several participants indicated that both content and background color influenced their grouping choices. For example, P8's attention was drawn to the color of the cards, "this side is darker . . . those are all light colors." P7 wanted that "each group . . . background colors would differ." P16 also deliberately selected alternative areas with strong color contrast. (2) Shape similarity: P7 suggested enlarging cards within an important group to increase perceptual contrast and make the grouping clearer. (3) Continuity and Closure: P3 described reorganizing the cards into enclosed or continuous arrangements, such as vertical lines or circular layouts, to support perceptual grouping.

6.4 Summary

In summary, when two grouping cues coexist and compete, participants find it more challenging to interpret how items are grouped, both objectively and subjectively. Although individuals differ in their preferences and understanding of proximity and common region cues, most participants aim for either a single unambiguous cue or multiple cues that work together. We further discuss the implications of the results in section 7.

7 Discussion

7.1 The Role of Physical Surfaces in Providing Perceptual Grouping of Virtual Elements

In Study 1, participants often responded more quickly when target pairs were within-group than between-group, with the neutral condition producing the fastest responses. This finding is consistent with previous 2D perceptual grouping research [3, 61, 70], suggesting that the experimental setup shaped users' perceptual grouping. Similar to previous studies [70], the particularly rapid responses in the neutral condition may result from less cluttered stimuli that required less attentional effort, thereby facilitating faster processing. However, surfaces are not always effective grouping cues in 3D environments. For instance, grouping effects weakened when vertical surfaces were positioned farther behind the targets or when horizontal surfaces were viewed from the side. One possible explanation is that participants in Study 1 were not explicitly instructed to attend to surfaces. Since the objects were observed from a side view, they could be naturally projected onto nearby vertical surfaces along participants' gaze. Consequently, participants could more easily ignore the surfaces that were distant or viewed obliquely because relating objects to those surfaces required extra effort [64], which weakened the common region cue. We speculate that the effect of surface direction on perceptual grouping may vary if the viewing direction shifts, such as switching to a top-down view.

In contrast, Study 2 revealed that the distance between the cards and the background had little impact on grouping by common region. This difference may be attributed to the fact that we explicitly instructed participants to use the background as a grouping cue in Study 2. Although a few participants note that parallax or large distances made the physical surfaces less useful, most of them efficiently relied on the background boundary to segment the cards and were unaware of the distance differences. Additionally, differences in depth perception between VR and VST-AR environments may have contributed, since previous research [83] has found greater depth estimation errors and distance underestimation in VST-AR. Participants' behavior also supports our guess that the effect of surface on grouping is related to the viewing direction and the object-surface relationships. When participants used horizontal surfaces (e.g., a table) to reinforce grouping, they aligned the cards with the table and viewed them from a top-down perspective (e.g., P16). Others also mentioned using the table's horizontal boundaries to classify items placed on it (e.g., P5).

Furthermore, we found no significant effect of surface area on grouping in Study 1. Even when a larger whiteboard was used in Study 2, the background remained a salient cue for grouping. Participants often referred to the whiteboard border when describing factors that aided or interfered with grouping. However, in terms of surface size, it is sufficient as long as the surface can accommodate the set of cards. Together, these results suggest that boundaries, rather than surface size, may be the critical features [14, 21] users rely on to distinguish groups.

In summary, physical surfaces facilitate the grouping of virtual objects most effectively when virtual content is spatially close to the physical plane and when boundaries between regions are clearly visible. Importantly, the impact of these cues also depends on individual differences, such as whether users recognize the background as relevant in the context of the task.

7.2 The Interaction of Multiple Grouping Cues

In Study 2, both the RT and subjective ratings indicate that participants found it difficult to determine grouping results when the two grouping cues were in competition. Since each cue was sufficiently strong on its own, participants hesitated when the cues conflicted and relied on different cues according to their preferences, which could lead to ambiguity in understanding. Besides, complex backgrounds are more likely to resemble the interference condition in the real-world environments. The interference mode did not affect participants' perceptions as much as we expected. This may be because, under this condition, the physical background neither provided a clear grouping cue nor obscured the large spacing between the two groups by proximity. However, we still need to be mindful of this situation, as the interference may cause uncomfortable experience for some users.

Furthermore, our interview results indicate that users adopt a variety of strategies for interacting with the environment when performing grouping tasks under different work modes, and the rationale behind these strategies can help us reveal improvements to future AR systems. Participants reported that when working alone, their grouping criteria could be relatively vague, such as "just separate them a bit" (P16) or "imagine a boundary" (P8). However, in co-located collaboration, users preferred to actively use "clearer boundaries" to indicate groupings, because multiple users may "perceive distances differently" and thus "require extra verbal negotiation" (P3, P4). For single user having multiple workspaces, it may "require additional effort to rearrange virtual content" if the system does not consider the impact of the changing environments on grouping (P9). Maintaining a consistent grouping effect of the environment on virtual content or avoiding new interference may potentially help improve the efficiency of completing higher-level tasks in multiple workspaces, (e.g., memory retention and learning processes [21, 22]). For remote collaboration scenarios, most participants preferred to minimize interference and confusion from their physical surroundings by seeking a clean background. For example, P10 expressed concern that grouping defined by boundaries within his/her own environment might be difficult for remote collaborators to understand. Participants speculated that leveraging physical space features would be more efficient only if collaborators could mutually recognize grouping boundaries (e.g., P13, P21). Therefore, systems that present matched grouping cues to all participants may support users to utilize their physical environment in remote grouping tasks, providing a better immersive experience.

7.3 Generalizability to Real-World Tasks

Perceiving individual elements as groups is one of the fundamental processes in perception [82]. We extracted several key spatial properties of surfaces in Study 1 and simulated the process of understanding the grouping results in Study 2. Although real-world scenarios are more complex, the simple spatial and grouping factors studied in this paper likely form a basic building block of low-level perceptual organization in AR applications. Building on our findings, future AR systems could incorporate these properties to better account for the environment's influence on virtual object grouping. For example, an AR brainstorming system could provide suggested usage areas for users by automatically extracting surface features and identifying regions with clear boundaries. Based on identifying areas that align with basic grouping principles, the system can further customize the selection of interactive areas according to task characteristics, user preferences, and other factors.

Although we only investigated the interaction between two principles, our findings could be extended to combinations with other intrinsic or extrinsic principles, such as environment as connectedness cues [67] combining with color similarity cues. Our results show that the physical environment can compete with intrinsic grouping cues in AR, indicating that we cannot ignore the role of the environment in grouping. Identifying relevant cues and ensuring their cooperation may provide a fundamental mechanism for AR systems supporting grouping tasks. In real-world tasks, users may only focus on semantic grouping criteria without explicitly paying attention to the underlying grouping principles. For example, one participant focused on how the cards were grouped based on their visual content rather than grouping cues in our Study 2 pilot stage. Previous research has shown that extrinsic grouping principles are more easily weakened by top-down influences [3]. When a user unfamiliar with the task context enters a collaboration where others have already focused on semantic content, s/he may be more influenced by environmental factors that others have overlooked, impairing task comprehension. By aligning multiple grouping cues, such systems may help to reduce users' cognitive burden during complex, semantic-focused tasks that are prone to ambiguity (e.g., tasks in [86]). Nonetheless, these findings pertain to basic perceptual grouping effects, and their generalizability to real-world settings remains to be tested; further discussion appears in subsection 7.5.

7.4 Design Recommendations

Based on the effective grouping influence of physical surfaces and these varying user strategies, we propose several design recommendations for future AR systems regarding the use of physical environments. **(1) Leverage physical boundaries as grouping cues.** AR systems can benefit from recognizing and utilizing physical structures in the environment, such as walls, desks, or natural regions, as extrinsic cues for grouping virtual content. These cues complement intrinsic virtual features, helping users efficiently organize digital information across diverse environments. Automatic placement strategies (e.g., avoiding occlusion or placing elements away from immediate boundaries [44]) can further reduce friction in interaction. **(2) Provide adaptive filtering of environmental distractions.** Because cluttered or salient real-world features can interfere with the perception of groups, AR systems should allow users to suppress or visually transform such background elements.

Adaptive filtering techniques, such as dynamically reducing contrast or altering textures of physical surroundings (e.g., [9, 46]), can improve the clarity of groupings. **(3) Support cross-environment grouping awareness in collaboration.** In multi-user or remote settings, systems should clearly indicate how physical contexts are used for grouping (e.g., the boundaries of the physical objects being utilized). Where appropriate, they can provide alternative representations or constraints to enable remote participants to understand grouping intentions without direct access to the local environment (e.g., [35, 66]). **(4) Detect and surface grouping ambiguities.** Since grouping cues are inherently subjective, systems should help users identify potential ambiguities(e.g., [34]), when a cluster of objects could be interpreted in multiple ways. Gentle alerts or visual indicators can draw attention to such cases, supporting negotiation and clarification within both individual and collaborative tasks.

7.5 Limitations and Future Work

A primary limitation of our study is that we did not investigate scenarios in which the distance between physical backgrounds and virtual objects varied widely. Because participants' responses were measured at a fixed location, we were unable to assess the effects of larger distance changes, which could confound user attention. Consequently, we could not determine a threshold at which users cease to rely on the physical background. Future research could identify effective perceptual thresholds or ranges of physical environments as grouping cues for both understanding users and system design.

Additionally, although we manipulated some spatial properties of the surfaces, we did not fully examine other important factors, such as card spacing and its interaction with card-background distance. In realistic settings, boundaries may not always be visible due to stereoscopic vision or limited spacing, whereas our experiments controlled for these issues by using sufficiently large gaps. The influence of diverse physical surfaces and 3D objects, such as different types of furniture with varying color and depth, also remains to be explored.

Finally, our tasks were intentionally abstract to isolate grouping principles, but this approach may restrict the generalizability of our findings. For example, our experimental setup matched the number of common regions to the number of groups. It remains unclear whether this effect persists when grouping dynamically changes and the group number exceeds what the environment affords. Incorporating task-specific scenarios in future research could clarify how variables such as group size, number of groups, and semantic relationship between the task and the environmental context affect user grouping behavior. Extending studies to cover more complex scenarios, including multiple environments and remote collaboration, would further illuminate the broader applicability of these results.

8 Conclusion

This work is motivated by a desire to understand how physical environments shape the perceptual grouping of virtual artifacts in AR. To explore this, we conducted two user studies. The first study used a repetition discrimination task in VR to systematically examine how physical surfaces influence grouping in 3D space. Our results demonstrated that spatial factors significantly impacted participants' grouping performance, confirming that the common region principle extends to 3D environments and is influenced by spatial features such as surface direction and distance. Building on these insights, the second study investigated how extrinsic grouping cues from the physical context interact with intrinsic cues embedded in virtual items within an AR scenario. The results revealed that conflicting cues created perceptual challenges, leading participants to adopt a variety of strategies to clarify group structures.

Together, these findings deepen our understanding of how physical and virtual cues jointly influence perceptual grouping in AR. Drawing on this integrated perspective, we propose design recommendations for future AR systems to better leverage environmental affordances and minimize unintended conflicts between grouping principles, thereby aligning more closely with users' natural perceptions.

Acknowledgments

We are deeply grateful to the participants of our user studies for their time and effort. This work is partially supported by the Research Grants Council of the Hong Kong Special Administrative Region under General Research Fund (GRF) with Grant No. 16207923.

References

[1] Christopher Andrews, Alex Endert, and Chris North. 2010. Space to Think: Large High-Resolution Displays for Sensemaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 55–64. doi:10.1145/1753326.1753336

[2] Juhee Bae and Benjamin Watson. 2014. Reinforcing Visual Grouping Cues to Communicate Complex Informational Structure. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (Dec. 2014), 1973–1982. doi:10.1109/TVCG.2014.2346998

[3] Diane M. Beck and Stephen E. Palmer. 2002. Top-down Influences on Perceptual Grouping. *Journal of Experimental Psychology: Human Perception and Performance* 28, 5 (2002), 1071–1084. doi:10.1037/0096-1523.28.5.1071

[4] Jack Brookes, Matthew Warburton, Mshari Alghadier, Mark Mon-Williams, and Faisal Mushtaq. 2020. Studying Human Behavior with Virtual Reality: The Unity Experiment Framework. *Behavior Research Methods* 52, 2 (April 2020), 455–463. doi:10.3758/s13428-019-01242-0

[5] Duncan P. Brumby and Susan Zhuang. 2015. Visual Grouping in Menu Interfaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 4203–4206. doi:10.1145/2702123.2702177

[6] Han Joo Chae, Youli Chang, Minji Kim, Gwanmo Park, and Jinwook Seo. 2020. ARphy: Managing Photo Collections Using Physical Objects in AR. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3334480.3382885

[7] Yifei Cheng, Yukang Yan, Xin Yi, Yuanchun Shi, and David Lindlbauer. 2021. SemanticAdapt: Optimization-based Adaptation of Mixed Reality Layouts Leveraging Virtual-Physical Semantic Connections. In *The 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21)*. Association for Computing Machinery, New York, NY, USA, 282–297. doi:10.1145/3472749.3474750

[8] Yi Fei Cheng, Christoph Gebhardt, and Christian Holz. 2023. InteractionAdapt: Interaction-driven Workspace Adaptation for Situated Virtual Reality Environments. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3586183.3606717

[9] Yi Fei Cheng, Hang Yin, Yukang Yan, Jan Gugenheimer, and David Lindlbauer. 2022. Towards Understanding Diminished Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3491102.3517452

[10] Hyunsung Cho, Yukang Yan, Kashyap Todi, Mark Parent, Missie Smith, Tanya R. Jonker, Hrvoje Benko, and David Lindlbauer. 2024. MineXR: Mining Personalized Extended Reality Interfaces. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3613904.3642394

[11] Jacob Cohen. 1988. *Statistical power analysis for the behavioral sciences*. Routledge. doi:10.4324/9780203771587

[12] Kylie Davidson, Lee Lisle, Kirsten Whitley, Doug A. Bowman, and Chris North. 2023. Exploring the Evolution of Sensemaking Strategies in Immersive Space to Think. *IEEE Transactions on Visualization and Computer Graphics* 29, 12 (Dec. 2023), 5294–5307. doi:10.1109/TVCG.2022.3207357

[13] Diego Drago, Jacob Bhattacharyya, Fahim Kawsar, and Stephen Anthony Brewster. 2025. The Effects of Physical Context on Collaborator Placement in Augmented Reality Meetings. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 517, 12 pages. doi:10.1145/3706599.3720195

[14] James H Elder and Steven W Zucker. 1998. Evidence for Boundary-Specific Grouping. *Vision Research* 38, 1 (1998), 143–152. doi:10.1016/S0042-6989(97)00138-7

[15] Alex Endert, Seth Fox, Dipayan Maiti, Scotland Leman, and Chris North. 2012. The Semantics of Clustering: Analysis of User-Generated Spatializations of Text Documents. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, Capri Island Italy, 555–562. doi:10.1145/2254556.2254660

[16] Barrett Ens, Eyal Ofek, Neil Bruce, and Pourang Irani. 2015. Spatial Constancy of Surface-Embedded Layouts across Multiple Environments. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction (SUI '15)*. Association for Computing Machinery, New York, NY, USA, 65–68. doi:10.1145/2788940.2788954

[17] Andreas Fender, Philipp Herholz, Marc Alexa, and Jörg Müller. 2018. OptiSpace: Automated Placement of Interactive 3D Projection Mapping Content. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3173574.3173843

[18] Daniel Immanuel Fink, Johannes Zagermann, Harald Reiterer, and Hans-Christian Jetter. 2022. Re-Locations: Augmenting Personal and Shared Workspaces to Support Remote Collaboration in Incongruent Spaces. *Proc. ACM Hum.-Comput. Interact.* 6, ISS (Nov. 2022), 556:1–556:30. doi:10.1145/3567709

[19] Edward G. Freedman. 2008. Coordination of the Eyes and Head during Visual Orienting. *Experimental Brain Research* 190, 4 (Oct. 2008), 369–387. doi:10.1007/s00221-008-1504-8

[20] Ran Gal, Lior Shapira, Eyal Ofek, and Pushmeet Kohli. 2014. FLARE: Fast Layout for Augmented Reality Applications. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 207–212. doi:10.1109/ISMAR.2014.6948429

[21] B. Garsoffky, M. Benkert, and S. Schwan. 2025. BEING SURROUNDED BY LEARNING CONTENT IN AN IMMERSIVE VIRTUAL ROOM: SPATIAL GROUPING INFLUENCES THE CORNER EFFECT ON MEMORY. In *EDULEARN25 Proceedings* (Palma, Spain) *(17th International Conference on Education and New Learning Technologies)*. IATED, 4452–4455. doi:10.21125/edulearn.2025.1147

[22] Bärbel Garsoffky and Stephan Schwan. 2024. Room Corners and How They Influence the Memory of Visual Information Arranged on Walls. *Scientific Reports* 14, 1 (May 2024), 12022. doi:10.1038/s41598-024-62648-1

[23] James J Gibson. 1950. The perception of visual surfaces. *The American journal of psychology* 63, 3 (1950), 367–384. doi:10.2307/1418003

[24] Jens Emil Sloth Grønbæk, Ken Pfeuffer, Eduardo Velloso, Morten Astrup, Melanie Isabel Sønderkær Pedersen, Martin Kjær, Germán Leiva, and Hans Gellersen. 2023. Partially Blended Realities: Aligning Dissimilar Spaces for Distributed Mixed Reality Meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3544548.3581515

[25] Jens Emil Sloth Grønbæk, Juan Sánchez Esquivel, Germán Leiva, Eduardo Velloso, Hans Gellersen, and Ken Pfeuffer. 2024. Blended Whiteboard: Physicality and Reconfigurability in Remote Mixed Reality Collaboration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3613904.3642293

[26] Violet Yinuo Han, Hyunsung Cho, Kiyosu Maeda, Alexandra Ion, and David Lindlbauer. 2023. BlendMR: A Computational Method to Create Ambient Mixed Reality Interfaces. *Proc. ACM Hum.-Comput. Interact.* 7, ISS (Nov. 2023), 436:217–436:241. doi:10.1145/3626472

[27] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. doi:10.1177/154193120605000909 arXiv:https://doi.org/10.1177/154193120605000909

[28] Keith C. Hendy, Katherine M. Hamilton, and Louise N. Landry. 1993. Measuring subjective workload: When is one scale better than many? *Human Factors* 35, 4 (1993), 579–601. doi:10.1177/001872089303500404

[29] Jaylin Herskovitz, Yi Fei Cheng, Anhong Guo, Alanson P. Sample, and Michael Nebeling. 2022. XSpace: An Augmented Reality Toolkit for Enabling Spatially-Aware Distributed Collaboration. *Proc. ACM Hum.-Comput. Interact.* 6, ISS (Nov. 2022), 568:277–568:302. doi:10.1145/3567721

[30] Xincheng Huang and Robert Xiao. 2024. SurfShare: Lightweight Spatially Consistent Physical Surface and Virtual Replica Sharing with Head-mounted Mixed-Reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 4 (Jan. 2024), 162:1–162:24. doi:10.1145/3631418

[31] Xincheng Huang, Michael Yin, Ziyi Xia, and Robert Xiao. 2024. VirtualNexus: Enhancing 360-Degree Video AR/VR Collaboration with Environment Cutouts and Virtual Replicas. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3654777.3676377

[32] Andrew Irlitti, Mesut Latifoglu, Thuong Hoang, Brandon Victor Syiem, and Frank Vetere. 2024. Volumetric Hybrid Workspaces: Interactions with Objects in Remote and Co-located Telepresence. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3613904.3642814

[33] Raphaël James, Anastasia Bezerianos, and Olivier Chapuis. 2023. Evaluating the Extension of Wall Displays with AR for Collaborative Work. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3544548.3580752

[34] Hyeon Jeon, Ghulam Jilani Quadri, Hyunwook Lee, Paul Rosen, Danielle Albers Szafir, and Jinwook Seo. 2024. CLAMS: A Cluster Ambiguity Measure for Estimating Perceptual Variability in Visual Clustering. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (Jan. 2024), 770–780. doi:10.1109/TVCG.2023.3327201

[35] Mohammad Keshavarzi, Allen Y. Yang, Woojin Ko, and Luisa Caldas. 2020. Optimization and Manipulation of Contextual Mutual Spaces for Multi-User Virtual and Augmented Reality Interaction. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 353–362. doi:10.1109/VR46266.2020.00055

[36] Dooyoung Kim, Seonji Kim, Selin Choi, and Woontack Woo. 2024. Spatial Affordance-aware Interactable Subspace Allocation for Mixed Reality Telepresence. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 1256–1265. doi:10.1109/ISMAR62088.2024.00142

[37] Seonji Kim, Dooyoung Kim, Jae-Eun Shin, and Woontack Woo. 2024. Object Cluster Registration of Dissimilar Rooms Using Geometric Spatial Affordance Graph to Generate Shared Virtual Spaces. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 796–805. doi:10.1109/VR58804.2024.00099

[38] Ruud Koolen. 2019. On Visually-Grounded Reference Production: Testing the Effects of Perceptual Grouping and 2D/3D Presentation Mode. *Frontiers in Psychology* 10 (Oct. 2019), 2247. doi:10.3389/fpsyg.2019.02247

[39] J. Kratt, T. Niese, R. Hu, H. Huang, S. Pirk, A. Sharf, D. Cohen-Or, and O. Deussen. 2018. Sketching in Gestalt Space: Interactive Shape Abstraction through Perceptual Reasoning. *Computer Graphics Forum* 37, 6 (2018), 188–204. doi:10.1111/cgf.13321

[40] Ernst Kruijff, J. Edward Swan, and Steven Feiner. 2010. Perceptual Issues in Augmented Reality Revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*. 3–12. doi:10.1109/ISMAR.2010.5643530

[41] Jérémy Lacoche, Eric Villain, and Anthony Foulonneau. 2022. Evaluating Usability and User Experience of AR Applications in VR Simulation. *Frontiers in Virtual Reality* 3 (July 2022). doi:10.3389/frvir.2022.881318

[42] Daniel Lakens. 2013. Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for t-Tests and ANOVAs. *Frontiers in Psychology* 4 (Nov. 2013). doi:10.3389/fpsyg.2013.00863

[43] Joon Hyub Lee, Donghyeok Ma, Haena Cho, and Seok-Hyung Bae. 2021. Post-Post-it: A Spatial Ideation System in VR for Overcoming Limitations of Physical Post-it Notes. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3411763.3451786

[44] Zhipeng Li, Christoph Gebhardt, Yves Inglin, Nicolas Steck, Paul Streli, and Christian Holz. 2024. SituationAdapt: Contextual UI Optimization in Mixed Reality with Situation Awareness via LLM Reasoning. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3654777.3676470

[45] David Lindlbauer, Michael Haller, Mark Hancock, Stacey D. Scott, and Wolfgang Stuerzlinger. 2013. Perceptual Grouping: Selection Assistance for Digital Sketching. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces (ITS '13)*. Association for Computing Machinery, New York, NY, USA, 51–60. doi:10.1145/2512349.2512801

[46] David Lindlbauer, Jörg Mueller, and Marc Alexa. 2017. Changing the Appearance of Real-World Objects By Modifying Their Surroundings. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3954–3965. doi:10.1145/3025453.3025795

[47] Lee Lisle, Kylie Davidson, Edward J.K. Gitre, Chris North, and Doug A. Bowman. 2021. Sensemaking Strategies with Immersive Space to Think. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 529–537. doi:10.1109/VR50410.2021.00077

[48] Lee Lisle, Kylie Davidson, Leonardo Pavanatto, Ibrahim A. Tahmid, Chris North, and Doug A. Bowman. 2023. Spaces to Think: A Comparison of Small, Large, and Immersive Displays for the Sensemaking Process. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 1084–1093. doi:10.1109/ISMAR59233.2023.00125

[49] Can Liu, Olivier Chapuis, Michel Beaudouin-Lafon, Eric Lecolinet, and Wendy E. Mackay. 2014. Effects of display size and navigation type on a classification task.

In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 4147–4156. doi:10.1145/2556288.2557020

[50] Jiazhou Liu, Barrett Ens, Arnaud Prouzeau, Jim Smiley, Isobel Kara Nixon, Sarah Goodwin, and Tim Dwyer. 2023. DataDancing: An Exploration of the Design Space For Visualisation View Management for 3D Surfaces and Spaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3544548.3580827

[51] Feiyu Lu, Mengyu Chen, Hsiang Hsu, Pranav Deshpande, Cheng Yao Wang, and Blair MacIntyre. 2024. Adaptive 3D UI Placement in Mixed Reality Using Deep Reinforcement Learning. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3613905.3651059

[52] Dolores Luna and Pedro R Montoro. 2011. Interactions between Intrinsic Principles of Similarity and Proximity and Extrinsic Principle of Common Region in Visual Perception. *Perception* 40, 12 (Dec. 2011), 1467–1477. doi:10.1068/p7086

[53] Dolores Luna, Cristina Villalba-García, Pedro R. Montoro, and José A. Hinojosa. 2016. Dominance Dynamics of Competition between Intrinsic and Extrinsic Grouping Cues. *Acta Psychologica* 170 (Oct. 2016), 146–154. doi:10.1016/j.actpsy.2016.07.001

[54] Weizhou Luo, Mats Ole Ellenberg, Marc Satkowski, and Raimund Dachselt. 2025. Documents in Your Hands: Exploring Interaction Techniques for Spatial Arrangement of Augmented Reality Documents. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–22. doi:10.1145/3706598.3713518

[55] Weizhou Luo, Anke Lehmann, Hjalmar Widengren, and Raimund Dachselt. 2022. Where Should We Put It? Layout and Placement Strategies of Documents in Augmented Reality for Collaborative Sensemaking. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3491102.3501946

[56] Weizhou Luo, Anke Lehmann, Yushan Yang, and Raimund Dachselt. 2021. Investigating Document Layout and Placement Strategies for Collaborative Sensemaking in Augmented Reality. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3411763.3451588

[57] Tiffany Luong, Yi Fei Cheng, Max Möbus, Andreas Fender, and Christian Holz. 2023. Controllers or Bare Hands? A Controlled Evaluation of Input Techniques on Interaction Performance and Exertion in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 29, 11 (Nov. 2023), 4633–4643. doi:10.1109/TVCG.2023.3320211

[58] Márcio C. F. Macedo and Antônio L. Apolinário. 2023. Occlusion Handling in Augmented Reality: Past, Present and Future. *IEEE Transactions on Visualization and Computer Graphics* 29, 2 (Feb. 2023), 1590–1609. doi:10.1109/TVCG.2021.3117866

[59] Pedro R. Montoro and Dolores Luna. 2015. Does the Relative Strength of Grouping Principles Modulate the Interactions between Them? *The Spanish Journal of Psychology* 18 (Jan. 2015), E33. doi:10.1017/sjp.2015.33

[60] Pedro R. Montoro, Cristina Villalba-García, Jacobo Albert, Antonio Prieto, and José A. Hinojosa. 2025. Temporal Neural Dynamics of the Competition between Extrinsic and Intrinsic Grouping Principles in Vision: An ERP Study. *Neuropsychologia* 208 (Feb. 2025), 109081. doi:10.1016/j.neuropsychologia.2025.109081

[61] Pedro R. Montoro, Cristina Villalba-García, Dolores Luna, and José A. Hinojosa. 2017. Common region wins the competition between extrinsic grouping cues: Evidence from a task without explicit attention to grouping. *Psychonomic Bulletin & Review* 24, 6 (Dec. 2017), 1856–1861. doi:10.3758/s13423-017-1254-3

[62] Jens Müller, Simon Butscher, Stefan P. Feyer, and Harald Reiterer. 2017. Studying Collaborative Object Positioning in Distributed Augmented Realities. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia (MUM '17)*. Association for Computing Machinery, New York, NY, USA, 123–132. doi:10.1145/3152832.3152856

[63] Aziz Niyazov, Barrett Ens, Kadek Ananta Satriadi, Nicolas Mellado, Loic Barthe, Tim Dwyer, and Marcos Serrano. 2023. User-Driven Constraints for Layout Optimisation in Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3544548.3580873

[64] Christine Nothelfer and Steven Franconeri. 2018. Visual Processing of Spatial Relations Within and Between Objects. *Journal of Vision* 18, 10 (Sept. 2018), 321–321. doi:10.1167/18.10.321

[65] Nels Numan, Shwetha Rajaram, Balasaravanan Thoravi Kumaravel, Nicolai Marquardt, and Andrew D Wilson. 2024. SpaceBlender: Creating Context-Rich Collaborative Spaces Through Generative 3D Scene Blending. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–25. doi:10.1145/3654777.3676361

[66] Michael Pabst, Linda Rudolph, Nikolas Brasch, Verena Biener, Chloe Eghtebas, Ulrich Eck, Dieter Schmalstieg, and Gudrun Klinker. 2025. MRUnion: Asymmetric Task-Aware 3D Mutual Scene Generation of Dissimilar Spaces for Mixed Reality Telepresence. *IEEE Transactions on Visualization and Computer Graphics* 31, 5 (May 2025), 3354–3364. doi:10.1109/TVCG.2025.3549878

[67] Stephen Palmer and Irvin Rock. 1994. Rethinking Perceptual Organization: The Role of Uniform Connectedness. *Psychonomic Bulletin & Review* 1, 1 (March 1994), 29–55. doi:10.3758/BF03200760

[68] Stephen E Palmer. 1992. Common region: A new principle of perceptual grouping. *Cognitive Psychology* 24, 3 (July 1992), 436–447. doi:10.1016/0010-0285(92)90014-S

[69] Stephen E Palmer. 1999. *Vision science: Photons to phenomenology.* MIT press.

[70] Stephen E. Palmer and Diane M. Beck. 2007. The repetition discrimination task: An objective method for studying perceptual grouping. *Perception & Psychophysics* 69, 1 (Jan. 2007), 68–78. doi:10.3758/BF03194454

[71] Sai Akhil Penumudi, Veera Aneesh Kuppam, Jeong Ho Kim, and Jaejin Hwang. 2020. The effects of target location on musculoskeletal load, task performance, and subjective discomfort during virtual reality interactions. *Applied Ergonomics* 84 (April 2020), 103010. doi:10.1016/j.apergo.2019.103010

[72] Nicholas F. Polys, Doug A. Bowman, and Chris North. 2011. The Role of Depth and Gestalt Cues in Information-Rich Virtual Environments. *International Journal of Human-Computer Studies* 69, 1 (Jan. 2011), 30–51. doi:10.1016/j.ijhcs.2010.05.007

[73] Xun Qian, Fengming He, Xiyun Hu, Tianyi Wang, Ananya Ipsita, and Karthik Ramani. 2022. ScalAR: Authoring Semantically Adaptive Augmented Reality Experiences in Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3491102.3517665

[74] Eric Ragan, Curtis Wilkes, Doug A. Bowman, and Tobias Hollerer. 2009. Simulation of Augmented Reality Systems in Purely Virtual Environments. In *2009 IEEE Virtual Reality Conference*. 287–288. doi:10.1109/VR.2009.4811058

[75] Eric D. Ragan, Alex Endert, Doug A. Bowman, and Francis Quek. 2012. How Spatial Layout, Interactivity, and Persistent Visibility Affect Learning with Large Displays. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, Capri Island Italy, 91–98. doi:10.1145/2254556.2254576

[76] Ludwig Sidenmark, Tianyu Zhang, Leen Al Lababidi, Jiannan Li, and Tovi Grossman. 2024. Desk2Desk: Optimization-based Mixed Reality Workspace Integration for Remote Side-by-side Collaboration. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3654777.3676339

[77] David Straker. 2009. *Rapid problem solving with post-it notes.* Balance.

[78] Ibrahim A. Tahmid, Lee Lisle, Kylie Davidson, Chris North, and Doug A. Bowman. 2022. Evaluating the Benefits of Explicit and Semi-Automated Clusters for Immersive Sensemaking. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 479–488. doi:10.1109/ISMAR55827.2022.00064

[79] Cristina Villalba-García, Mikel Jimenez, Dolores Luna, José A Hinojosa, and Pedro R Montoro. 2021. Competition between Perceptual Grouping Cues in an Indirect Objective Task. *Quarterly Journal of Experimental Psychology* 74, 10 (Oct. 2021), 1724–1736. doi:10.1177/17470218211010486

[80] Johan Wagemans, James H. Elder, Michael Kubovy, Stephen E. Palmer, Mary A. Peterson, Manish Singh, and Rüdiger von der Heydt. 2012. A Century of Gestalt Psychology in Visual Perception: I. Perceptual Grouping and Figure–Ground Organization. *Psychological Bulletin* 138, 6 (2012), 1172–1217. doi:10.1037/a0029333

[81] John Wenskovitch and Chris North. 2021. An Examination of Grouping and Spatial Organization Tasks for High-Dimensional Data Exploration. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (Feb. 2021), 1742–1752. doi:10.1109/TVCG.2020.3028890

[82] Max Wertheimer. 1938. Laws of organization in perceptual forms. In *A source book of Gestalt psychology.*, Willis D. Ellis (Ed.). Kegan Paul, Trench, Trubner & Company, London, 71–88. doi:10.1037/11496-005

[83] Franziska Westermeier, Larissa Brubach, Carolin Wienrich, and Marc Erich Latoschik. 2024. Assessing Depth Perception in VR and Video See-Through AR: A Comparison on Distance Judgment, Performance, and Preference . *IEEE Transactions on Visualization & Computer Graphics* 30, 05 (May 2024), 2140–2150. doi:10.1109/TVCG.2024.3372061

[84] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. doi:10.1145/1978942.1978963

[85] Emily Wong, Adélaïde Genay, Jens Emil Sloth Grønbæk, and Eduardo Velloso. 2025. Spatial Heterogeneity in Distributed Mixed Reality Collaboration. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 980, 19 pages. doi:10.1145/3706598.3714033

[86] Shengchen Zhang, Weiwei Guo, and Xiaohua Sun. 2025. Align with Me, Not TO Me: How People Perceive Concept Alignment with LLM-Powered Conversational Agents. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3706599.3720126

A Subjective Feedback Questionnaire in Study 1

The questionnaire used a 5-point Likert scale to capture the user's subjective feelings about the influence of experimental conditions on their response time for identifying the target (1 = *not at all*, 2 = *a little*, 3 = *moderately*, 4 = *very much*, and 5 = *extremely*).

(1) **Target position.** To what extent did the *position of the target objects* affect your response time? Please provide separate ratings for the positions illustrated in Fig. 16: On the left side, at the center, and on the right side.

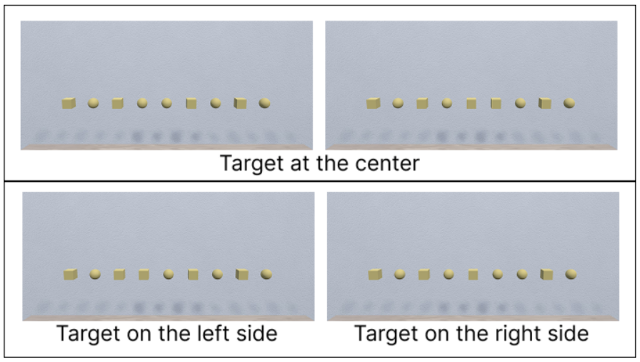


Figure 16: Illustrations for Question 1: target-position conditions.

(2) **Grouping type (vertical).** To what extent did the *relationship between the surface and the target objects* affect your response time when the surface direction was **vertical**? Please provide separate ratings for the scenarios illustrated in Fig. 17: targets separated by two surfaces (between-group) and targets in the same surface (within-group).

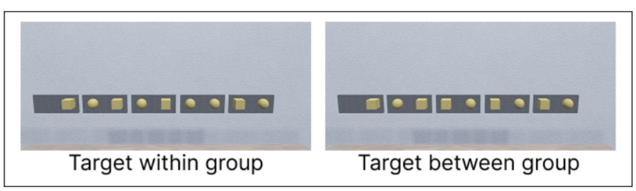


Figure 17: Illustrations for Question 2: grouping type (vertical surface).

(3) **Surface distance (vertical).** To what extent did the *distance of the surface* affect your response time when the surface direction was **vertical**? Please provide separate ratings for the scenarios illustrated in Fig. 18: behind and close from the target, behind and far to the target, in front and close to the target, and in front and far from the target.

(4) **Surface size (vertical).** To what extent did the *scale of the surface* affect your response time when the surface direction was **vertical**? Please provide separate ratings for the scenarios illustrated in Fig. 19: when the surface is large and when the surface is small.

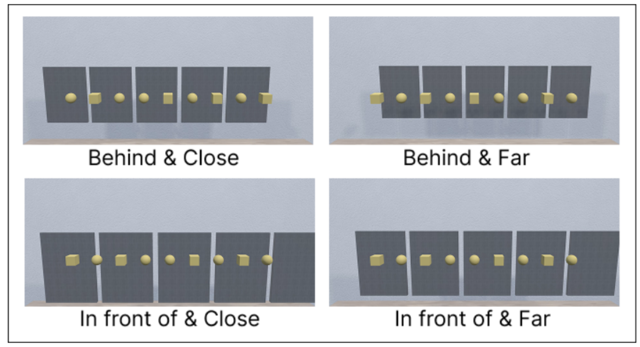


Figure 18: Illustrations for Question 3: surface distance levels (vertical surface).

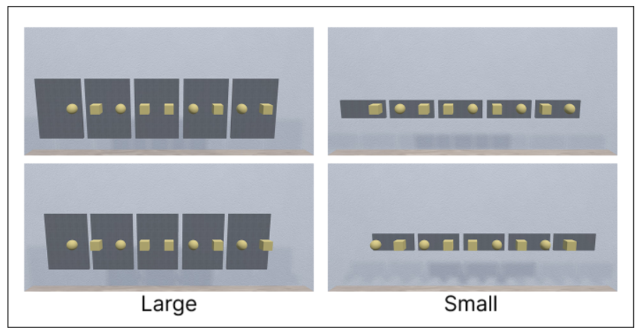


Figure 19: Illustrations for Question 4: surface scale levels (vertical surface).

(5) **Grouping type (horizontal).** To what extent did the *relationship between the surface and the target objects* affect your response time when the surface direction was **horizontal**? Please provide separate ratings for the scenarios illustrated in Fig. 20: targets separated by two surfaces (between-group) and targets in the same surface (within-group).

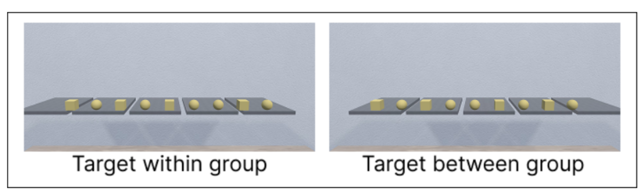


Figure 20: Illustrations for Question 5: grouping type (horizontal surface).

(6) **Surface distance (horizontal).** To what extent did the *distance of the surface* affect your response time when the surface direction was **horizontal**? Please provide separate ratings for the scenarios illustrated in Fig. 21: above and close from the target, above and far to the target, under and close to the target, and under and far from the target.

(7) **Surface size (horizontal).** To what extent did the *scale of the surface* affect your response time when the surface

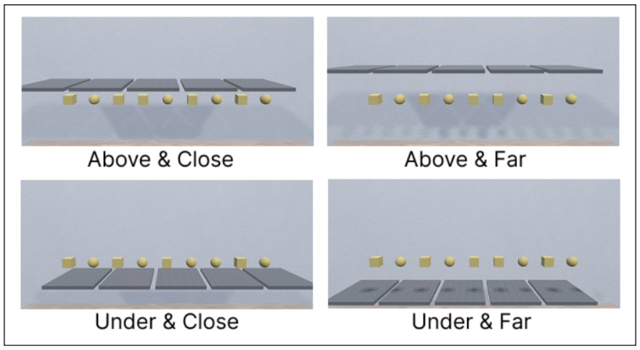


Figure 21: Illustrations for Question 6: surface distance levels (horizontal surface).

direction was **horizontal**? Please provide separate ratings for the scenarios illustrated in Fig. 22: when the surface is large and when the surface is small.

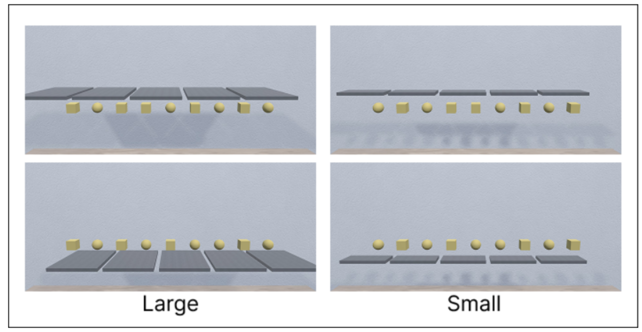


Figure 22: Illustrations for Question 7: surface scale levels (horizontal surface).