Multi-WIM Paradigm: Comparing Urban Planning Variants with Worlds-in-Miniature

Karoline Brehm, Ephraim Schott, Manuel Hartmann, and Bernd Froehlich

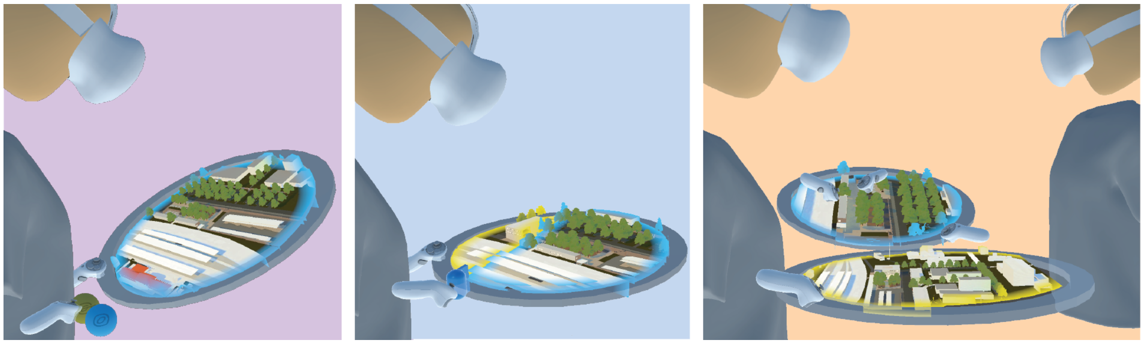

Fig. 1: Users comparing 3D urban planning variants in immersive virtual reality using, from left to right: Toggle, Swipe and Juxtaposed visualizations of WIM multiples.

Abstract—World-in-Miniatures (WIMs) are an established technique in virtual reality (VR) that provide an overview of the surrounding virtual environment. While traditionally used to support navigation over larger distances and distant manipulation in large-scale spaces, their potential for comparing alternative versions of a virtual environment has not been exploited. In this paper, we introduce a Multi-WIM paradigm that facilitates the detailed exploration and comparison of urban planning alternatives. Each alternative is visualized in a separate WIM. The WIMs can be arranged side by side (juxtaposed) or stacked. Stacking allows toggling between alternatives or using a swipe tool where on each side one alternative is shown. These three modes (juxtaposed, toggle, swipe) provide flexible ways of exploring differences and similarities between alternatives, with highlighting functionalities that reveal differences of functional units across designs. We evaluated the Multi-WIM paradigm in a single-factor VR user study (N=24), combining a single-user experiment with a multi-user exploration. Participants had to identify differences and estimate distances, areas and heights across urban planning alternatives. Results show that the stacked toggle mode was significantly faster for finding differences while the juxtaposed view better supported a landmark recall task. Furthermore, toggle was the overall preferred technique in the single-user mode. In contrast, in multi-user settings participants favored juxtaposed, as it fostered collaborative discussion and shared sensemaking. The identified design trade-offs between efficiency and collaboration indicate the need to incorporate all three WIM modes, together with seamless transitions among them, in urban planning applications so as to accommodate both loosely and tightly coupled phases of collaboration.

Index Terms—Virtual Reality, 3D Comparisons, Worlds-In-Miniature, Citizen Participation, Social XR

1 Introduction

Worlds-in-Miniature (WIMs) are a well-established technique in virtual reality (VR), primarily employed to facilitate navigation across large distances, to provide an exocentric overview and to enable distant manipulation in large-scale environments. Despite their widespread use for these purposes, the potential of WIMs for supporting systematic comparison of alternative virtual environment designs remains largely unexplored. In urban planning, where the evaluation of multiple design variants is central, WIMs likewise represent a promising tool. Yet, it remains unclear how such variants should be presented to users and how different presentation strategies influence the effectiveness and efficiency of comparison tasks when evaluating the design features of competing variants.

- *Karoline Brehm is with Bauhaus-Universität Weimar. E-mail: karoline.brehm@uni-weimar.de.*
- *Ephraim Schott is with Bauhaus-Universität Weimar. E-mail: ephraim.schott@uni-weimar.de.*
- *Manuel Hartmann is with Bauhaus-Universität Weimar. E-mail: manuel.hartmann@uni-weimar.de.*
- *Bernd Froehlich is with Bauhaus-Universität Weimar. E-mail: bernd.froehlich@uni-weimar.de.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

In this paper, we address this gap by investigating how multiple WIM comparison techniques and their transitions can be employed to support the detailed exploration and comparative evaluation of urban design variants. In our Multi-WIM paradigm, each alternative is visualized in a separate, navigable WIM, which can be arranged side by side (juxtaposed) or stacked (see Fig. 1). Stacking supports toggling between alternatives or employing a swipe tool in which each side of a separating plane reveals a different design variant. These three comparison techniques (juxtaposed, toggle, and swipe) provide flexible ways of exploring differences and similarities between alternatives. In addition, highlighting functional units across variants facilitates their localization.

We hypothesized that the three WIM comparison modes of our approach (toggle, swipe, juxtaposed) provide complementary strengths for different forms of spatial exploration and comparison. This leads to two research questions: (1) Which WIM modes best support specific spatial exploration and comparison tasks? and (2) How does the setting—single-user versus multi-user—affect their use? To address these questions, we evaluated the Multi-WIM paradigm in a single-factor VR user study (N=24), combining a single-user experiment with a multi-user exploration. In the single-user stage, we recorded participants performance in tasks such as locating key objects, identifying differences between variants, counting specified features, and estimating distances, areas, and sizes across designs. In the multi-user stage, participants collaboratively discussed various design aspects.

We administered questionnaires to collect subjective data about the participants' experience and preferences at each stage.

This work contributes a Multi-WIM paradigm that extends the established concept of WIMs beyond navigation to support systematic comparison of urban design alternatives. By introducing three complementary comparison modes (toggle, swipe, juxtaposed), we provide a flexible framework for spatial exploration and evaluation in both single-user and collaborative settings. Our empirical study demonstrates that the modes afford some distinct advantages: toggle and swipe are particularly effective for focused pairwise comparison, while juxtaposed appears beneficial for landmark recall. Moreover, the multi-user exploration revealed that the availability of multiple modes fosters productive collaboration, enabling participants to shift between loosely and tightly coupled analysis. Together, these findings demonstrate how WIMs can be employed for comparative evaluation tasks in VR-supported urban planning.

2 Related Work

Our WIM variant visualizations build on prior research in citizen participation, spatial knowledge acquisition, world-in-miniatures, and comparative visualization. To contextualize our approach, we review relevant work from these four domains, emphasizing the key insights and limitations that informed our design.

2.1 Urban Planning

In urban planning contexts, immersive stereoscopic displays have shown to support planning and evaluation by improving 3D perception and eliciting social presence (the "feeling of being there" [45]), which in turn fosters a better understanding of spatial relations [34, 35, 43, 47]. Building on this, a growing body of work explores augmented and mixed reality [36, 41] as well as virtual reality [7, 25, 33, 51] to enhance citizen participation. Our work contributes to a higher level of participation by explicitly enabling the effective comparison of multiple variants of the same environment, thereby supporting collaboration and exchange around different design alternatives.

2.2 Spatial Knowledge Acquisition

Urban development projects need to allow stakeholders, in particular citizens, to form a mental model of the environment, a process known as spatial knowledge acquisition or cognitive mapping [13]. This process leads to an understanding of the spatial layout as well as the relationships between entities within a space, which is categorized into knowledge about specific buildings (*landmark* knowledge), paths between locations (*route* knowledge), and the overall layout of areas (*survey* or *configurational* knowledge) [1, 38, 44]. These forms of knowledge can be acquired through different sources and perspectives [1, 6]: either through navigation within a real or virtual environment (navigation learning), through external sources such as maps, WIMs, or images (map learning), or through verbal instructions (description learning). Acquired spatial knowledge is typically measured through spatial memory tasks like recalling or placing landmarks or sketching space layouts and routes [1, 19, 25].

Our work addresses the acquisition of survey knowledge through WIMs. The tasks in our study are informed by the mentioned forms of knowledge acquisition required for building a mental model of a design variant and its particular characteristics.

2.3 Worlds-in-Miniature

The World-in-Miniature (WIM) metaphor provides users with an exocentric view of their surroundings, supporting spatial knowledge acquisition by allowing the miniature to be moved, rotated, and scaled, thereby offering perspectives from different angles and distances. Introduced nearly 30 years ago, WIMs were designed to seamlessly integrate overview and detail within virtual environments [48]. Since then, WIMs have been explored for a variety of purposes, most commonly to support navigation for single users or groups [3, 8, 37, 49], to provide spatial overviews [10, 27, 32, 51], or to enable distant manipulation [4, 20, 22, 49]. Application domains are diverse, ranging from biomedical data analysis [9] and storytelling [54] to the exploration of astrophysical environments [27] and architectural design [51]. Interaction techniques such as scaling, panning, and slicing have been continuously refined to address limitations and adapt WIMs to different contexts [10, 27, 52]. Beyond interaction, several studies have compared WIMs with 2D maps in the context of locomotion [3, 12]. Englmeier et al. [14] examined alternative WIM visualizations, contrasting a planar WIM with a spherical "little planet" WIM, and found that despite its unconventional form, the latter outperformed the original WIM in terms of navigation efficiency. In the domain of spatial knowledge acquisition, Hsu et al. [19] demonstrated that reduced WIMs can enhance landmark and route recall, and that both WIMs and maps are most effective when complemented by a first-person view enriched with building information. Danyluk et al. [11] provide a comprehensive analysis of the WIM design space and identify WIM multiples as an underrepresented dimension. Since WIMs support spatial understanding they were a natural choice for enabling the comparison and exploration of different urban design variants. However, little work has addressed the intersection of multiple WIMs and variant comparison. Nam et al. [32] introduced "WIMs in Wedges" to compare forestry data, yet their approach did not place WIMs at the center of comparison, nor did it investigate alternative WIM visualization techniques to support different comparison tasks. Other examples include VRGit by Zhang et al. [53], which employs multiple WIMs to visualize differences in versioned room arrangements, and TimeTables [55], which uses multiple WIM-like views to explore and analyze spatio-temporal energy consumption data. While both systems provide multiple views, they do not provide explicit mechanisms for comparison beyond visual juxtaposition.

2.4 Comparisons in 2D and 3D Visualizations

While WIMs have not previously been applied to the comparison of design variants, the cognitive tasks involved are closely related to well-studied comparison techniques in 2D data visualization and visual analytics. Prior works have shown that multiple coordinated views can facilitate effective comparison across alternatives both in desktop [39] and immersive scenarios [15, 28]. Brushing-and-linking techniques further support this process by highlighting related content across views [23, 39]. For identifying differences among competing 2D stimuli, studies have compared common paradigms such as juxtaposition, superimposition, sliders, and explicit encoding [17, 24]. Each approach has been shown to support different tasks; for example, slider metaphors are particularly effective for analyzing before-and-after images [16].

Similar techniques have also been investigated in 3D, for example superimposition to explore generative designs [30] or juxtaposition of small multiples to explore a 3D gesture dataset [26]. Kim et al. [24] emphasize that 3D data should ideally be spatially aligned, simultaneously visible, and information-preserving, but since this is often not feasible, balancing of design tradeoffs is necessary.

While the cognitive processes that affect comparison are well established in 2D visualization, comparable studies have not been conducted with WIMs. It is plausible that similar results extend to single-user WIM interactions, yet it remains unclear whether they also hold in collaborative scenarios where multiple users jointly compare variants. These concepts informed the design of our Multi-WIM paradigm.

3 System Implementation

Our work investigates the use of WIMs for different comparison tasks, which arise in diverse contexts, for example when contrasting before-and-after scans of buildings, comparing medical imaging results, or assessing variations of architectural designs.

Driven by a collaboration with local urban planners, our work focuses on recreating citizen information events and decision-making processes in immersive environments to improve citizen participation in urban planning. Typically, such events involve presenting the current situation of a location, introducing planned changes and alternative design options, and then discussing these alternatives with respect to planning criteria and citizen concerns.

To replicate these decision processes and enhance them using VR, we developed a multi-user application using Unity3D and Netcode.

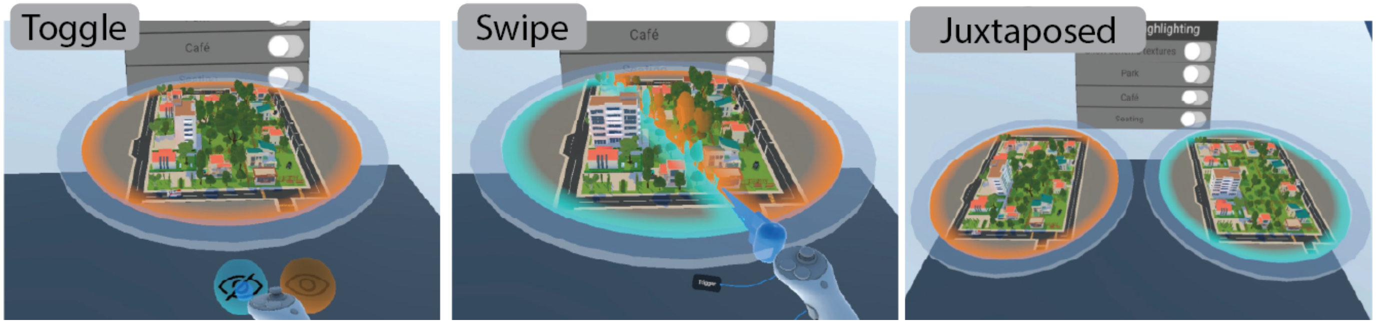

Fig. 2: The interactive comparative visualizations modes Toggle, Swipe and Juxtaposed were evaluated in the study.

Our system allows participants to explore various architectural and environmental models either by navigating them at full scale or by inspecting them through WIMs. In the following sections, we detail our approach, including the Multi-WIM paradigm, its comparison modes, and additional highlighting features.

3.1 Multi-WIM Paradigm

To foster both individual and collaborative exploration of environment variants, we propose the Multi-WIM paradigm. The idea builds on the observation that evaluating and comparing multiple alternatives involves different tasks of varying complexity. Prior research has shown that spatial comparison tasks, such as orienting oneself, locating features, identifying differences, estimating area and size, or comparing functionalities, cannot be fully supported by a single visualization technique alone. Following the taxonomy by Kim et al. [24], we therefore designed a hybrid interaction concept for WIMs that integrates three established comparison strategies (juxtaposition, interchangeable, and swipe) into a flexible framework. This allows us to compensate for the trade-offs of each technique and support a broad range of comparison tasks within one unified paradigm. We also considered using semi-transparent superimposition or explicit encoding approaches, but initial tests showed that these approaches were not suitable for comparing our urban models.

The core principle of the Multi-WIM paradigm is that each urban design variant is represented by its own dedicated WIM. Thus, when comparing multiple variants, a distinct WIM is available for each urban design. While WIMs can be moved, rotated, and scaled independently or in lockstep, they can also be combined into a *WIM Stack*. A WIM stack spatially aligns multiple WIMs, providing a shared viewport and enabling access to the toggle and swipe comparison modes (see Fig. 2).

In our implementation, each WIM is equipped with a surrounding circular handle for direct hand manipulation. This handle allows users to grab, rotate, place, and scale the WIM within their workspace. The extent and shape of the WIM are defined by the radius of this circle, which also determines the viewport of the model. Consequently, the circle radius defines both the geometry and the size of the WIM in the virtual environment. By grabbing the displayed "world" with one hand, users can pan the WIM content whereas two-handed interaction enables rotation and zooming. For WIM Stacks, the same zoom and pan level is consistently applied across all variants.

The stackable WIM interaction concept permits users to freely arrange WIMs and variants, realized in our implementation through drag-and-drop manipulation via the circular handle. This allows users to seamlessly transition between comparison modes: juxtaposed variants can be simply overlaid which makes them snap into a stacked WIM or separated again by pulling them apart. The stacked WIM allows switching between Toggle and Swipe modes. In the following, these Toggle, Swipe, and Juxtaposed modes are described in detail.

3.1.1 Toggle Comparison Technique

The Toggle mode is an interactive comparative visualization categorized as an interchangeable approach by Kim et al. [24]. The terms blitting and temporal multiplexing are also sometimes used to describe the technique [29,46]. Variants contained in one WIM are spatially aligned, with one variant visible at a time. Variant selector widgets act as radio buttons, which allows the user to toggle between variants.

3.1.2 Swipe Comparison Technique

The Swipe mode is an interactive superimposition approach, in which variants contained in a WIM are spatially aligned. It is an established technique for image and map comparison [29]. In contrast to Toggle, two variants can be visible at the same time, as they are shown on opposite sides of a clip plane that is orthogonal to the WIM surface. This allows the user to swipe between variants through direct hand interaction with a slider handle, revealing larger or smaller sections of the variants. To support more than two stacked variants, selector widgets allow the users to configure the visible pair.

3.1.3 Juxtaposed Comparison Technique

The Juxtaposed mode implements a widely used comparative design by displaying variants in separate WIMs that are placed side by side. Zoom and pan can be either performed in lockstep between WIMs or independently.

3.2 Feature Highlighting Technique

A feature highlighting technique was implemented to assist users in locating specific functional entities, structures or areas in the variants and to support comparison of their areas and sizes.

Feature highlights allow the user to visually single out components, such as functional entities in an urban design that are identified by name in the control menu. In our urban planning context, users might view a life-sized variant from an egocentric perspective, using the WIM as an overview. In this case, feature highlights provide a link between world and miniature, providing controls for accessing additional information and providing orientation in larger environments such as a city block. Feature highlights are similar to brushing and linking as they link functional entities across variants that might appear at different locations and in different form. So far, these entities were preselected by the urban planners to focus the attention on the most relevant aspects. However, they could also be interactively selected in one variant and then highlighted in the others, following the classic brushing-and-linking approach.

4 Evaluation

Our evaluation focused on assessing the different comparison modes of our proposed Multi-WIM paradigm in the context of variant comparison and spatial exploration. Two central research questions guided this evaluation:

- **RQ1** What are the different qualities of the WIM modes in the contexts of spatial exploration and comparison?
- **RQ2** How do single-user versus multi-user settings influence the suitability of these modes?

The following subsections describe the methodology we used to evaluate the WIM modes and to investigate our research questions.

Fig. 3: Three fictional worlds, each represented by two variants created to support different types of comparison tasks

4.1 Experimental Setup

The study was conducted in a quiet computer lab with an open area of approximately 10 m × 5 m, providing sufficient space for participants to move freely without the need for additional virtual navigation. During the single-user stage, a mobile partition wall was placed to separated users. The experimental application was implemented in Unity 6 and executed in PC-VR mode ("link mode") on a Meta Quest 3 headset at 72 Hz, displayed at the device's native resolution (2.064 × 2,208 pixels per eye). This setup ensured a stable and high-fidelity VR experience.

4.1.1 WIM Modes

The stacking interaction was disabled for the study to reduce the number of interactions users had to learn. Instead, for the single-user stage of the experiment, one or two WIMs were configured according to the condition (Toggle, Swipe, Juxtaposed).

For the multi-user stage of the experiment, the virtual environment was configured with four WIMs: A WIM for Toggle and Swipe respectively, each containing the variant pair, as well as two WIMs showing the variants in juxtaposition (see Fig. 4).

4.1.2 Materials

Four fictional city blocks (Tutorial World, World 1, World 2, World 3) were created for the single-user part of the experiment using a commercially available 3D model pack (see Fig. 3). For each city block, two variants **A** and **B** were created by introducing differences for each study task.

During the multi-user session, participants explored two design variants showing potential layouts for the future development of the Weimar train station, a site likely familiar to students at our institution. This real-world example was chosen for its relevance and recognizability, and it reflects the citizen participation use case that originally motivated the Multi-WIM paradigm.

4.1.3 Comparison and Spatial Knowledge Acquisition Tasks

In the single-user stage, participants completed the following seven tasks in the given order:

- FIND: Mark the location of a given target object.
- LANDMARK: Memorize the same four landmarks (structures or buildings) that were placed at different locations in the variants, and mark their locations in a reduced model.
- DIFFERENCES: In a highlighted area, mark six locations where differences occurred between the two variants.
- COUNT: Count the occurrences of a given object and indicate which variant has a higher count.
- AREA, SIZE and DISTANCE (ESTIMATION): Indicate which variant had the larger area or size of a given structure, or greater walking distance between two given locations.

The tasks required the user to either indicate locations in the WIM by placing a marker, or indicating the color-coded variant **A** or **B** on a task panel as the answer to a comparison question.

4.1.4 Feature Highlighting

In the single-user stage of the experiment, the feature highlight menu contained entries for the items or areas that were featured in the tasks. The menu was shown only during COUNT and estimation tasks (AREA, SIZE, DISTANCE) to encourage its use only for tasks that related to the features. In the multi-user stage, the menu was always visible and the selected features were highlighted in all four WIMs.

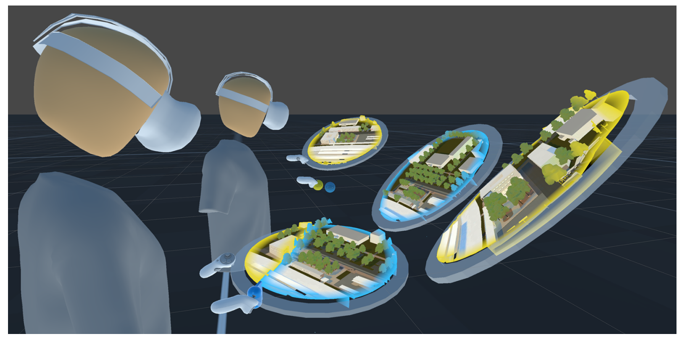

Fig. 4: Colocated users compare two design variants (orange, blue). The front user swipes stacked WIMs and uses juxtaposed WIMs, while the back user uses toggle WIM.

4.2 Study Design

We employed a single-factor within-subjects study design. The independent variable was the *WIM mode* with three levels: *Toggle*, *Swipe*, and *Juxtaposed*. The experiment encompassed a single-user and a multi-user stage.

In the single-user stage, participants completed the task set once under each WIM mode condition, with World 1, World 2 and World 3 appearing in fixed order. The order of conditions was fully counterbalanced across participants.

In the multi-user stage, participants were asked to collaborate on answering discussion prompts about four aspects of the train station designs: design of the bus terminal, identify which features are present in each of the designs, design of car parking, preferences from a cyclist perspective.

4.2.1 Measures

To address RQ1 we assessed effectiveness, usability and participant preferences of each WIM mode. For RQ2, we evaluated the participants' opinions with respect to the suitability of the different modes for the multi-user setting.

As a measure of effectiveness, the task completion time (TCT) was recorded in seconds. Depending on the task, we also determined the error rate (COUNT, AREA, SIZE, DISTANCE, DIFFERENCES) or distance of the marker from the target location (LANDMARK), measured in distance units in the reference frame of the 3D model.

After experiencing each condition, participants filled out a questionnaire about the WIM mode. In this questionnaire, the usability of the WIM mode was measured using the System Usability Scale (SUS) [5] and custom 7-point Likert scales for rating frustration and agreement with the statement "With <WIM mode> I always knew which variant I am looking at." Participants also gave the WIM mode a 1 to 10 star rating and described what they did and did not like about it in free text. The suitability for each task was evaluated by rating agreement with the statement "The <WIM Mode> was very useful for the <TASK>." on a 7-point scale.

After experiencing all three WIM modes, participants completed a final single user questionnaire. They indicated their overall preference and provided reasoning for their most and least liked condition in free-text responses. Suitability of each WIM mode across *all* tasks was rated on a 7-point scale. Feature highlighting was evaluated using custom 7-point Likert items rating usefulness, ease of use, clarity, and whether it complemented the Multi-WIM, making it easier to understand. A demographic questionnaire was also completed at this stage.

After the multi-user session, participants filled out a final questionnaire. In this questionnaire the suitability of conditions in a multi-user settings was rated and participants explaining their reasoning for the WIM mode they found most and least suitable. Aspects relevant for collaboration were measured using two custom 7-point Likert items: "I could easily point out content to others when using <WIM mode>" (Point Out) and "I could follow what content was being discussed when using <WIM Mode>" (Follow Discussion). Co-presence was evaluated using a subscale from the Networked Minds Social Presence Inventory (NMSPI) [18].

4.3 Procedure

Participants were welcomed to the lab and informed about the study and the experimental procedure. Written consent was obtained prior to conducting the single-user (parallel) stage, followed by the multi-user stage of the experiment. Interactions were practiced by participants following information presented on a tutorial panel in the virtual environment, each under the guidance of a study conductor. Before the first trial, each WIM interaction technique was practiced (zoom+pan, marker placement, activating feature highlights). Interaction with the WIM modes were practiced at the beginning of each condition (widgets for toggling variants, using slider to swipe, viewing juxtaposed WIMs).

4.4 Participants

The study was conducted with 24 participants (13 male, 11 female) aged between 23 and 34 years (M = 26.5, $\sigma = 3.1$). Participants were recruited via university mailing lists and received a compensation of 15 Euro. No participant ended the study early. Most participants were students in the Computer Science or Human Computer Interaction field (n = 21), only few from other fields (n = 2), and one participant did not provide this information (n = 1). In regard to familiarity with VR, 10 participants reported being very familiar and three participants reported having never used VR before (M = 3.7, $\sigma = 1.4$). Five participants played video games rather often (M = 2.6, $\sigma = 1.4$). In regard to the multi-user stage of the experiment, 10 participants already knew the other person.

4.5 Hypotheses

To investigate our research questions, we perform an exploratory statistical analysis on the recorded measures to compare the three WIM modes. To structure the analysis and reporting of results, we formulate hypotheses that are evaluated using non-parametric statistical tests. Specifically, we apply Friedman tests to identify differences between the WIM mode conditions, followed by Bonferroni-corrected pairwise post-hoc tests where applicable. In cases where no significant differences are observed, additional Bayesian analysis was used to assess the strength of evidence for the absence of effects.

With regard to RQ1, we expected that specific modes would provide advantages for particular task types, and that participants would value and rate the modes differently:

- Usability (**H1**) and overall ratings (**H2**) will be different between WIM modes.
- Performance will be different between WIM modes in the FIND (**H3**), LANDMARK (**H4**), COUNT (**H5**) and DIFFERENCES (**H6**) task, as well as the ESTIMATION tasks (**H7**).

We expect that Toggle would yield the best TCT in the FIND task, as it supports fast switching for comparison technique. We also expect that the stacked WIM modes (Toggle, Swipe), which offer spatially aligned

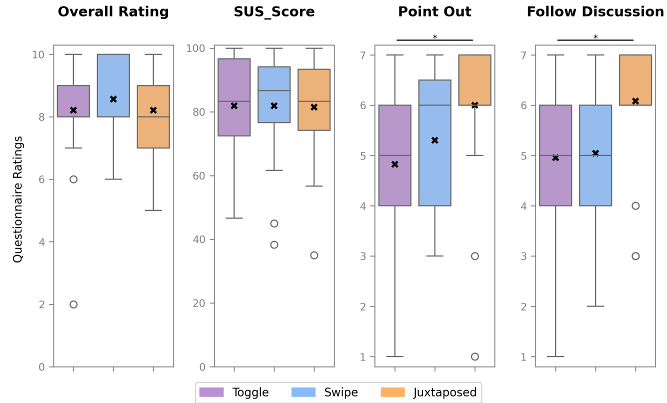

Fig. 5: Results per WIM mode for 1) overall rating, 2) SUS score, as well as 3) ease of pointing out features and 4) ease of following discussions in multi-user setting.

model variants in a shared viewport, will benefit the DIFFERENCES and ESTIMATION tasks, leading to lower TCTs. In contrast, we assume that the LANDMARK and COUNT tasks can be completed sequentially. In this case, the Juxtaposed technique, which does not require interaction to reveal both variants, might yield the lowest TCT.

Addressing RQ2, we considered that the social dynamics of collaboration as well as non-verbal communication such as deictic pointing would affect the perceived suitability and preference of the WIM modes:

- In the multi-user setting, the perceived ease of pointing out content (H8) and the ease of following a discussion about content (H9) will be different between WIM modes.

We expected Juxtaposed to best support communication in the multi-user setting, because multiple users can view and point to details in both variants without potentially hiding areas or variants that the other user is interested when swiping or toggling.

5 Results

In this section, we report on the results of the study. We conducted statistical analyses of the measured variables to examine whether there were significant differences that support our hypotheses.

All participants ($N = 24$) completed the task set described in Sec. 4.1.3 using each of the three WIM modes. After each condition, as well as after the single- and multi-user parts, questionnaires were administered, resulting in five per participant.

We began the analysis by testing for normality using the Shapiro–Wilk test. As the data did not meet the assumption of normality, we applied the non-parametric Friedman test. When significant effects were found ($p < .05$), Bonferroni-corrected Wilcoxon signed-rank tests were conducted as post-hoc analyses, resulting in following significance levels: $*p < .017$, $**p < .003$, $***p < .0003$. Where no significant effects were observed with respect to our hypothesis, we conducted post-hoc Bayesian repeated-measures ANOVAs in JASP using software defaults (uninformative prior) to quantify evidence in favor of the null hypothesis [21, 31, 40]. We report the Bayes Factor BF_{01} which indicates how likely the observed data is to occur under the null hypothesis compared to the alternative hypothesis, and interpret the value according to the JASP guidelines [50].

5.1 Usability and Overall Rating

Overall rating and SUS scores are depicted in Fig. 5.

The mean SUS scores for Toggle (81.81), Swipe (78.89), and Juxtaposed (81.18) all fall within the "good" usability range [2], indicating comparable usability across modes. A Friedman test revealed no statistically significant effect of WIM mode on usability ($\chi^2(2) = 0.15$, $p = .926$, $W = .001$). Overall, participants rated Swipe highest on a 10 star scale (M = 8.57, $\sigma = 1.20$), followed by Juxtaposed (M = 8.22, $\sigma = 1.41$) and Toggle (M = 8.22, $\sigma = 1.76$). A Friedman test showed

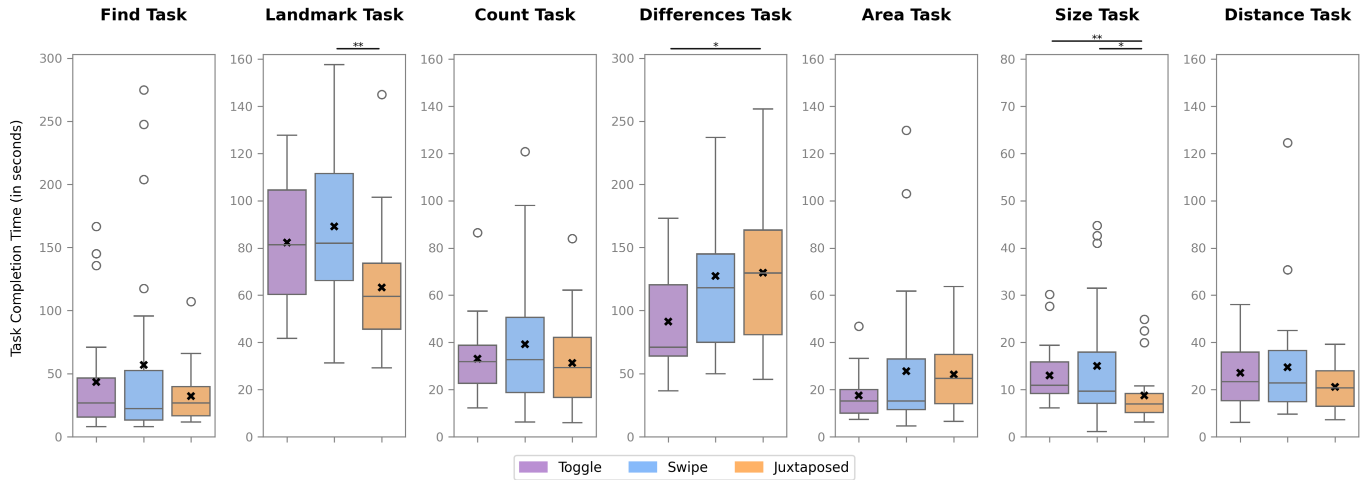

Fig. 6: Task Completion Times for all tasks with significance levels in post-hoc tests (*$p < .017$, **$p < .003$, ***$p < .0003$)

	Toggle		Swipe		Juxtaposed	
Task	*M*	σ	*M*	σ	*M*	σ
FIND	43.4	44.3	56.9	77.2	**32.1**	22.0
LANDMARK	82.4	25.3	89.1	34.3	**63.3**	25.9
COUNT	33.1	15.2	39.2	27.2	**31.3**	19.8
DIFFERENCES	**91.1**	42.2	127.3	80.4	130.0	61.4
AREA	**17.5**	10.1	27.8	30.8	26.5	16.0
DISTANCE	27.1	14.4	29.4	24.6	**21.1**	8.5
SIZE	13.0	6.2	15.0	12.5	**8.7**	5.7

Table 1: Task Completion Time (in seconds). Mean (M) and standard deviation (σ) per task and condition.

no significant effect of WIM mode on overall ratings, ($\chi^2(2) = 0.11$, $p = .947$, $W = .002$).

The post-hoc Bayesian analysis provides moderate evidence in favor of the null hypothesis $\mathbf{H1}_0$ for the SUS score ($BF_{01} = 8.318$) and $\mathbf{H2}_0$ for the overall ratings ($BF_{01} = 5.557$), which indicates that these measures are similar for the different WIM modes.

The feature highlighting received high mean ratings in regards to usefulness (M = 6.42, $\sigma = 1.14$), ease of use (M = 6.46, $\sigma = 1.35$), clarity (M = 6.46, $\sigma = 1.44$) and tended to be perceived by participants as complementing the WIM modes well in the right context (M = 6.33, $\sigma = 0.87$).

5.2 Task Results

As described in Sec. 4.2.1, the Task Completion Time (TCT, in seconds) as well as task-specific measures were recorded for each task. Outcomes are reported in Tab. 1 and Fig. 6.

Task complexity varied greatly, for instance, the LANDMARK task required placing eight markers, whereas the FIND task required only two. Consequently, we emphasize that TCT can only be compared between conditions within the same task. Overall, 72 TCT values were collected for each task.

5.2.1 Find Task

All targets were found by all participants. A Friedman test revealed no statistically significant effect of WIM mode on TCT in the FINDING task ($\chi^2(2) = 0.083$, $p = .9592$, $W = .002$) and a Bayesian post-hoc test showed weak evidence ($BF_{01} = 2.780$) for the null hypothesis $\mathbf{H3}_0$.

5.2.2 Landmark Task

In terms of accuracy, participants placed landmarks closest to their original positions in the Juxtaposed condition (M = 9.17, $\sigma = 7.482$), followed by Toggle (M = 11.703, $\sigma = 8.175$) and Swipe (M = 15.287, $\sigma = 11.854$). A Friedman test showed no significant effect of WIM mode on distance to target ($\chi^2(2) = 1.083$, $p = .5818$, $W = .023$).

However, the WIM mode did have a statistically significant effect on TCT ($\chi^2(2) = 10.58$, $p = .005$, $W = .220$), which supports **H4**. Post-hoc tests showed that Juxtaposed (M = 63.3) was significantly faster than Swipe (M = 89.1) ($W = 35.0$, $p = .0005$).

5.2.3 Count Task

Participants made no counting errors with any WIM mode. A Friedman test revealed no statistically significant effect of WIM mode on TCT, ($\chi^2(2) = 0.33$, $p = .847$, $W = .007$) and the Bayesian post-hoc test revealed rather weak evidence ($BF_{01} = 3.432$) for the null hypothesis $\mathbf{H5}_0$.

5.2.4 Differences Task

Participants achieved high accuracy across all conditions, with Toggle yielding the lowest error rate (7.64%), followed by Swipe (8.33%) and Side-by-Side (9.03%).

A Friedman test revealed a significant effect of WIM mode on TCT, ($\chi^2(2) = 6.75$, $p = .034$, $W = .141$), supporting **H6**. Post-hoc analysis showed that Toggle (M = 91.15) was significantly faster than Juxtaposed (M = 129.96) ($W = 65.0$, $p = .0137$).

5.2.5 Estimation Tasks

Estimation tasks required participants to compare functional entities across variants, in terms of SIZE, AREA, and DISTANCE.

In terms of accuracy, Toggle consistently produced the lowest error rates. For AREA, error rates were 20.8% for Toggle, 25.0% for Swipe, and 45.8% for Juxtaposed. In the DISTANCE task, participants performed best with Toggle (0% errors), followed by Juxtaposed (8.3%) and Swipe (16.7%). Accuracy in the SIZE task was high across all conditions, with only a single error occurring in Toggle (4.2%), while Juxtaposed and Swipe showed no errors.

For TCT, a Friedman test revealed a significant effect of WIM mode on TCT in the SIZE task, ($\chi^2(2) = 12.25$, $p = .0022$, $W = .255$). However, no significant effects were found for AREA ($\chi^2(2) = 2.250$, $p = 0.3247$, $W = .047$) or DISTANCE ($\chi^2(2) = 4.083$, $p = 0.1298$, $W = 0.085$). Post-hoc comparisons for SIZE showed that Juxtaposed (M = 8.7) was significantly faster than both Swipe (M = 15.0, $W = 68.0$, $p = .017$) and Toggle (M = 13.0, $W = 46.0$, $p = .002$) .

Bayesian post-hoc tests for AREA ($BF_{01} = 1.464$) proved inconclusive and evidence in favor of the null hypothesis regarding DISTANCE ($BF_{01} = 2.259$) is weak. We therefore find no evidence regarding **H7** and conclude that the ESTIMATION tasks require further, separate investigation in future studies.

	Toggle		Swipe		Juxtaposed	
Question	*M*	*σ*	*M*	*σ*	*M*	*σ*
Frustration	2.52	1.62	2.26	1.29	**2.57**	2.02
Variant Clarity	5.91	1.38	6.04	1.26	**6.52**	1.12
Task Suitability						
Across Tasks	5.29	1.77	**5.54**	1.59	5.08	1.56
FIND	5.74	1.05	6.04	1.30	**6.17**	0.98
LANDMARK	5.17	1.85	5.57	1.34	**6.09**	1.28
COUNT	6.00	1.31	**6.39**	0.84	6.22	1.09
DIFFERENCES	5.91	1.65	**6.17**	1.03	4.65	1.87
ESTIMATION	**5.91**	1.24	5.65	1.50	5.22	1.48

Table 2: Mean (M) and standard deviation (σ) of frustration, perceived clarity and task suitability by condition on a 7-point scale.

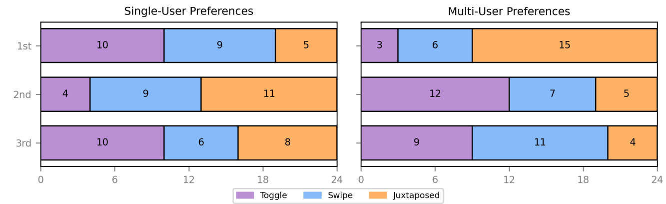

Fig. 7: Stacked bar charts visualizing participants' preferences.

5.3 Frustration, Clarity and Task Suitability

As stated in Sec. 4.2.1, the questionnaires gathered data on overall ratings and system usability (reported in Sec. 5.1), as well as frustration, clarity, and the perceived suitability of each WIM mode for the different tasks. Results are summarized in Tab. 2.

Overall, participants reported low frustration (M = 2.4, $\sigma = 1.65$) and found it clear which variant they were looking at (M = 6.2, $\sigma = 1.27$). The WIM modes were rated as suitable across tasks at an average rating of M = 5.3.

A Friedman test revealed no significant effects of condition on task-specific suitability ratings (all $p > .1207$), with the exception of DIFFERENCES ($\chi^2(2) = 9.123$, $p = .0104$, $W = 0.198$), where post-hoc analysis shows that Juxtaposed (M = 4.65) was rated significantly lower than both Swipe (M = 6.17, $W = 17.0$, $p = .0026$) and Toggle (M = 5.91, $W = 20.0$, $p = .007$).

5.4 Multi-User Results and Preferences

Overall, participants assessed co-presence with an average score of 5.86 ($\sigma = 1.01$) in the multi-user part, indicating that they perceived each other and could effectively attract the other's attention; requirements for successful collaboration in the given tasks.

For "Point Out," the Juxtaposed condition received the highest ratings (M = 6.00, $\sigma = 1.45$), followed by Swipe (M = 5.30, $\sigma = 1.43$) and Toggle (M = 4.83, $\sigma = 1.78$). A Friedman test confirmed a small but significant effect, ($\chi^2(2) = 6.00$, $p = .050$, $W = .130$) which supports **H8**. Post-hoc analysis indicated that Juxtaposed was rated significantly higher than Toggle ($W = 28.5$, $p = .007$).

Similarly, for "Follow Discussion," Juxtaposed again scored highest (M = 6.09, $\sigma = 1.20$), compared to Swipe (M = 5.04, $\sigma = 1.58$) and Toggle (M = 4.96, $\sigma = 1.58$). A Friedman test revealed a significant effect of WIM mode ($\chi^2(2) = 6.86$, $p = .032$, $W = .149$) in support of **H9**. Post-hoc analysis showed that Juxtaposed was rated significantly higher than Toggle ($W = 31.5$, $p = .005$).

Figure 7 illustrates that in the single-user part, stacked WIM modes were preferred by 75% of participants, with Toggle being the most favored technique (41.7%), followed by Swipe (33.3%), while the remaining 25% favored the Juxtaposed mode. In the multi-user part, however, preferences clearly shifted: Juxtaposed became the most favored technique (62.5%), followed by Swipe (25.0%) and Toggle (12.5%).

5.5 Qualitative Feedback

To assess the qualitative feedback, we analyzed collected answers from open-ended questions using axial coding.

Toggle: Participants appreciated the Toggle mode for its simple switching interaction (Comment Occurrence, CO = 11), its support for comparisons (CO = 7), ease of use (CO = 3), and, in some cases, the clear distinguishability of variants (CO = 3). As P15 noted in the toggle questionnaire: "[The Toggle mode] made the process of switching from one variant to another very simple."

In the final single-user questionnaire, ten participants favored Toggle, citing ease of use (CO = 9), support for comparison (CO = 4), and quick interaction (CO = 3) as reasons. Participants generally used this technique to rapidly alternate between variants, which, as P20 remarked, "made differences easy to find out." In the multi-user context, however, only three participants preferred this mode, mainly for its comparison benefits (CO = 2).

18 participants also mentioned drawbacks. Several noted that it was unclear which variant was currently displayed (CO = 5), which made comparisons difficult (CO = 3) and, in some cases, caused eye strain (CO = 2) due to the need to "pay attention to color" (P17). Consequently, ten participants ranked Toggle as their least favorite mode in the final single-user questionnaire, mainly due to interaction issues (CO = 5) such as "having to locate the toggle buttons" (P15) and finding it "hard to compare and to remember the features of each place" (P3). In the multi-user setting, nine participants also disliked this mode, describing it as confusing.

Swipe: Participants valued the Swipe mode primarily for its ease of comparison (CO = 9), especially when identifying subtle differences between variants. They also emphasized its overall ease of use (CO = 7), followed by the benefits of clear color coding (CO = 4) which supported the identification of variants and allowed to switch quickly between variants (CO = 4). In the final single-user questionnaire, nine participants favored Swipe, describing it as interactive and enjoyable (CO = 4), and highlighting its benefits for comparison (CO = 4). In the multi-user setting, six participants rated it as the most suitable, emphasizing support for comparison (CO = 5) for instance, P15 explained: "In multi user setting using the swipe one was the easiest because I didn't have to move around a lot and also most questions asked were based on estimation and difference and for both I will prefer swipe." A few also noted its collaborative aspect, such as P6: "We could together use one map and move the slider together taking turns."

At the same time, Swipe was associated with several drawbacks. Eighteen negative comments were recorded, including issues with the grabbing gesture (CO = 7), overlap of views (CO = 3), and annoyance with repeated swiping actions (CO = 3). P13 described it as "not really comfortable having to keep the button pressed." Others criticized the small inspection area (P9) or inconsistencies when zooming. In the combined questionnaire, six participants named Swipe as their least favorite, describing it as confusing (CO = 4) or effortful to use (CO = 3). In the multi-user context, eleven participants considered it the least suitable, citing interaction difficulties (CO = 9) and challenges for collaboration (CO = 2). As P20 summarized: "I think it is hard to switch between versions meanwhile discussing something."

Juxtaposed: Participants valued the Juxtaposed mode for clear variant differentiation (CO = 9), ease of use (CO = 10) without requiring "extra action" (P11), and support for direct comparison (CO = 9). In the final single-user questionnaire, only five participants favored this technique, highlighting its straightforward use (CO = 3) and the advantage of simultaneous visibility (CO = 4). In contrast, in the multi-user questionnaire 15 participants preferred Juxtaposed, particularly for its benefits in comparison (CO = 10) and cooperation (CO = 5). P13 noted that "it was easier to mark the differences and to discuss and observe at[sic] both variants at the same time" and P21 mentioned that multiple users "can check both tables at the same time" indicating that Juxtaposed supported parallel interactions. P1 further reflected that viewing both variants provided "time to think on my own and not being controlled by the other person's decisions," suggesting that the mode also supported individual reflection within collaborative settings.

The most frequent complaint about the Juxtaposed mode concerned the need to look back and forth between the two WIMs (CO = 7). Several participants also mentioned memorization and comparison difficulties (CO = 6), with P1 noting that they often had to remember details from one view before checking the other. Additional issues included interaction problems such as linked zooming (CO = 2). Five participants reported no dislikes at all (CO = 5).

In the single-user questionnaire, eight participants reported Juxtaposed as their least favorite technique, mainly due to the increased mental load (CO = 4) and the effort of frequent head movement (CO = 4). In the multi-user questionnaire, only four participants considered Juxtaposed least suitable. Reasons included interaction issues (CO = 2), such as difficulties handling the display when multiple users were involved (P14, P15), and space requirements (CO = 3), with participants noting that the side-by-side setup occupied too much room (P4, P14). One participant (P6) also criticized that both maps zoomed simultaneously.

6 Discussion

In this section, we first discuss the implications of the quantitative and qualitative results of our study in regard to the suitability of Toggle, Swipe and Juxtaposed modes for variant comparison in single-user versus collaborative multi-user settings. Secondly, we discuss the implications of our work for VR in citizen participation and for application in other domains.

6.1 Variant Comparison with Multi-WIM Paradigm

Overall, our findings suggest that our interaction paradigm for multiple WIMs is suitable for variant comparison and spatial knowledge acquisition: On average all WIM modes achieved good SUS scores (M = 80,6) and high ratings (M = 8,3). We found moderate evidence in favor of the null hypothesis $\mathbf{H1}_0$ ($BF_{01} = 8.318$) and $\mathbf{H2}_0$ ($BF_{01} = 5.557$) that these measures do not differ by WIM mode. The condition had no significant effect on the accuracy in the landmark placement task ($\chi^2(2) = 1.083$, $p = .5818$, $W = .023$). FIND and COUNT tasks were completed with zero errors, SIZE with one error. Participants on average missed 0.5 out of six DIFFERENCES and 8.3% of DISTANCE estimation responses were incorrect. Only the AREA task was often answered wrong at 38.86% error rate, making it an outlier among the ESTIMATION tasks. This points to a possible flaw in framing of the question or the material. Nonetheless, participants were able to complete the tasks in all conditions.

With respect to **RQ1**, we hypothesized that performance in the tasks would differ by WIM mode and found partial support for this in three tasks: LANDMARK, DIFFERENCES and SIZE. In addition, subjective ratings from the questionnaires showed that participants indeed favored different techniques for specific tasks. Interestingly, the same technique might be liked or disliked for the same characteristic. For example, users commented both very positively and negatively on the swipe interaction. It required a greater range of motion than other techniques, which one participant enjoyed for the "sense of more control" and quick "back and forth" (P13), while another disliked that "more effort" (P16) was required.

While there was no significant effect of condition on the accuracy of LANDMARK placement, participants completed the task significantly faster using Juxtaposed compared to Swipe ($W = 35.0$, $p = .0005$), which supports **H4** . This effect could be explained by users memorizing landmarks more easily, leading to more confidence in giving answers. However, compared to Swipe and Toggle, no interaction with a slider or widget was required to reveal the other variant, since both were simultaneously visible, possibly speeding up the completion of marker placement. This warrants further investigation of spatial knowledge acquisition in virtual environments composed of different variants.

In the COUNT task, Participants achieved a 0% error rate across conditions and a Bayesian post-hoc analysis found weak evidence for the null hypothesis $\mathbf{H5}_0$ ($BF_{01} = 3.432$).

On average, both Toggle and Swipe yielded faster TCTs than Juxtaposed in the DIFFERENCES task, with a significant difference between Toggle and Juxtaposed ($W = 65.0$, $p = .041$), which provides support for **H6**. The better performance also seems to be reflected in user preferences, with Toggle and Swipe receiving significantly higher suitability ratings than Juxtaposed (Toggle $W = 17.0$, $p = .0027$; Swipe $W = 20.0$, $p = .007$). In the qualitative feedback, Swipe and Toggle were both valued for finding differences. For example, one participant commented that Toggle "made differences easy to find out" (P20) and with Swipe P13 was able to "quickly see the difference between both variants." This supports our assumption that there is a benefit in providing swipe and toggle techniques to support comparison tasks, particularly those that require detection of subtle differences between variants.

For the ESTIMATION tasks, results were inconclusive regarding **H7**. Juxtaposed achieved significantly faster TCT compared to Toggle and Swipe in the SIZE task ($W = 46.0$, $p = .002$). In contrast, for the DISTANCE task a Bayesian post-hoc analysis found weak evidence pointing in favor of the WIM mode having no effect on TCT. With regard to the AREA task, results were inconclusive (). In the qualitative feedback, easy comparison was mentioned as a benefit in each case with seven comment occurrences for Toggle and 9 occurrences for Swipe and Juxtaposed mode.

6.2 Single User vs. Multi User

Regarding **RQ2** we found that the multi-user setting affected user preference for the conditions: At 62.5%, 2.5 times as many participants indicated Juxtaposed as the most suitable option in a multi-user setting, compared to 25% who preferred it in a single-user setting. Qualitative feedback gives us more insight into why the Juxtaposed technique tended to be rated as most suitable over stacked WIM modes: Nine participants perceived Toggle as confusing in the multi-user setting and other nine reported difficulties with the Swipe interaction (CO=9). In contrast, Juxtaposed was valued for supporting collaboration (CO=5), with drawbacks primarily related to a crowded workspace (CO=3); potentially an artifact of providing all modes simultaneously in the multi-user stage (see Fig. 4).

In regard to factors related to collaboration, participants rated the ease of pointing out things to the other user and the ability to follow the discussion significantly higher in Juxtaposed compared to Swipe and Toggle respectively (Swipe M = 5.04, $\sigma = 1.58$; Toggle M = 4.96, $\sigma = 1.58$). Thus, we found support for **H8** and **H9**.

6.3 Implications

The results of our study have implications for the design of WIM multiples for both single-user and multi-user settings. Different visualization modes should be available and it must be possible to fluidly switch between them. When transferring from the context of urban planning to other application domains, it might be necessary to incorporate other types of comparative visualization. For example, no explicit encoding techniques were incorporated because there is no obviously suitable encoding operation to apply to two urban planning variants, but they exist for other kinds of 3D models. Therefore, when adopting Multi-WIM paradigm to other application context, we should incorporate visualizations that were designed for the given type of data if possible.

6.4 Limitations and Future Work

Several limitations apply to this work. Notably, we recruited a sample of 24 participants from the student body at our institution, most of whom have a technical background and some prior VR experience, which is not representative of the general population. Furthermore, it is not ideal that participants completed the single-user stage in the same room, separated by a partition wall, which allowed them to overhear conversations between the other participant and study conductor. It is possible that this slightly affected performance or participants' opinions.

In each stage of the study, the number of variants displayed as WIMs was limited to two alternative designs. The stackable WIM interaction described in Sec. 3.1 can deal with a larger number of variants. However, a good usability of the stackable WIM interaction concept would be an important factor in the overall usability of the Multi-WIM paradigm and warrants further evaluation, as we conclude that multiple visualizations should be available to the user and allow for fluid switching.

The selected visualizations have some limitations. In particular, the swipe technique by design only allows comparison of two variants and it is not obvious if and how it should be extended to more than two variants. Selecting two variants from a stack of WIMs might be the best option. Additionally, the circular shape of the viewport might not be equally suitable for toggle, swipe and juxtaposed views. For other datasets and tasks, other shapes might be more efficient. Another open question relates to the spatial arrangement of multiple WIMs in the workspace, as well as other possible layout as have been explored in prior work on immersive comparison of maps [42, 46].

During the multi-user stage we present all WIM modes simultaneously, requiring users to walk to each mode to make use of it. For loosely coupled collaboration phases, it seems appropriate that each user has a separate WIM that can be used in all three modes for all variants under discussion. During tighter collaboration phases, users might focus on either one or transfer their findings and settings to a larger, shared WIM.

In our study, we focused only on interactions and tasks related to the Multi-WIM concept. We did not include egocentric views of variants and how they relate to the exocentric view onto a WIM. However, intuitive multi-user interaction for the selection, manipulation, and navigation of variants while maintaining a coherent relationship between the egocentric view and the WIM view remains a considerable challenge for future work and would also require closer investigation of spatial knowledge acquisition.

7 Conclusion

Our Multi-WIM paradigm extends prior research by integrating established comparison techniques (such as juxtaposition and swipe interactions) into WIMs. This approach enables users to seamlessly switch between different comparison modes, allowing them to select suitable or preferred visualizations depending on the task at hand. Our controlled study with 24 participants examined single- and multi-user conditions through a set of tasks including spatial knowledge acquisition (landmark recall) and comparisons such as finding entities, identifying differences, and comparing distances and sizes. The results reveal that for identifying differences, toggling or swiping was beneficial and preferred by users, yielding significantly faster TCTs than juxtaposition. For other tasks, the modes appeared to be equally suitable, with participants expressing diverging preferences. In the collaborative setting, however, this pattern shifted, with 62.5% of participants favoring the Juxtaposed technique because of its stronger support for discussion and joint exploration.

Over the last 30 years, the WIM metaphor has proven its usefulness in virtual reality. Our work continues the story of this small but effective metaphor by investigating comparative visualization techniques based on WIM multiples in single- and multi-user settings. Beyond urban planning, the Multi-WIM paradigm introduces a promising approach for enhancing insight and collaborative analysis in social VR contexts such as medical imaging, environmental simulations, and architectural design.

Acknowledgments

This work has mainly received funding from the German Federal Ministry of Research, Technology and Space (BMFTR) under grant FKZ 16SV9200 (Bauhaus Participation Lab). This work was also partially funded by the German Research Foundation (DFG) under Germany's Excellence Strategy within the Cluster of Excellence 'Imaginamics: Practices and Dynamics of Social Imagining' (EXC 3078).

References

[1] N. Ahmadpoor and S. Shahab. Spatial knowledge acquisition in the process of navigation: A review. *Current Urban Studies*, 7(1):1–19, 2019. 2

[2] A. Bangor. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies*, 4(3), 2009. 5

[3] L. Berger and K. Wolf. WIM: Fast Locomotion in Virtual Reality with Spatial Orientation Gain & without Motion Sickness. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, MUM '18, pp. 19–24. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3282894.3282932 2

[4] N. L. P. Berwaldt, G. Di Domenico, and C. T. Pozzer. Virtual MultiView Panels for Distant Object Interaction and Navigation in Virtual Reality. In *Proceedings of the 25th Symposium on Virtual and Augmented Reality*, SVR '23, pp. 88–95. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3625008.3625023 2

[5] J. Brooke. SUS: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, 1995. 4

[6] T. T. Brunyé, A. Gardony, C. R. Mahoney, and H. A. Taylor. Going to town: Visualized perspectives and navigation through virtual environments. *Computers in Human Behavior*, 28(1):257–266, 2012. 2

[7] H.-M. Chang, T.-W. Hsu, M.-H. Tsai, S. V. Babu, and J.-H. Chuang. Design and Evaluation of an Asynchronous VR Exploration System for Architectural Design Discussion Content. *Virtual Worlds*, 3(1), 2024. doi: 10.3390/virtualworlds3010001 2

[8] V. Chheang, F. Heinrich, F. Joeres, P. Saalfeld, R. Barmaki, B. Preim, and C. Hansen. WiM-Based Group Navigation for Collaborative Virtual Reality. In *2022 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pp. 82–92, 2022. doi: 10.1109/AIVR56993.2022.00018 2

[9] D. Coffey and D. F. Keefe. Biowim: A multisurface, multitouch interface for exploratory visualization of biomedical data. 2010. 2

[10] D. Coffey, N. Malbraaten, T. Le, I. Borazjani, F. Sotiropoulos, and D. F. Keefe. Slice wim: a multi-surface, multi-touch interface for overview+ detail exploration of volume datasets in virtual reality. In *Symposium on Interactive 3D Graphics and Games*, pp. 191–198, 2011. 2

[11] K. Danyluk, B. Ens, B. Jenny, and W. Willett. A Design Space Exploration of Worlds in Miniature. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, pp. 1–15. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445098 2

[12] G. D. Domenico, F. D. C. Becker, B. B. Schirmer, N. L. P. Berwaldt, G. M. d. Freitas, A. C. Neto, C. T. Pozzer, L. M. Fontoura, and R. C. Nunes. Evaluation of navigation in immersive large-scale vr environments using minimap-assisted teleportation. In *Proceedings of the 26th Symposium on Virtual and Augmented Reality*, pp. 270–274, 2024. 2

[13] R. M. Downs and D. Stea. *Image and environment: Cognitive mapping and spatial behavior*. Transaction Publishers, 2017. 2

[14] D. Englmeier, W. Sajko, and A. Butz. Spherical World in Miniature: Exploring the Tiny Planets Metaphor for Discrete Locomotion in Virtual Reality. *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 345–352, 2021. doi: 10.1109/VR50410.2021.00057 2

[15] S. K. Feiner and C. Beshers. Worlds within worlds: Metaphors for exploring n-dimensional virtual worlds. In *Proceedings of the 3rd Annual ACM SIGGRAPH Symposium on User Interface Software and Technology*, UIST '90, pp. 76–83. Association for Computing Machinery, 1990. doi: 10.1145/97924.97933 2

[16] M. Gleicher. Considerations for visualizing comparison. *IEEE transactions on visualization and computer graphics*, 24(1):413–423, 2017. 2

[17] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011. 2

[18] C. Harms and F. Biocca. Internal consistency and reliability of the networked minds measure of social presence. 2004. 5

[19] P.-C. Hsu, S. V. Babu, and J.-H. Chuang. Comparative Evaluation of Differing Levels of Information Presentation in 3D Mini-Maps on Spatial Knowledge Acquisition in VR. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, Saint Malo, France, 2025. doi: 10.1109/VR59515.2025.00076 2

[20] S. Imatani, K. Tobitani, and K. Akabane. Touchwim: Object manipulation in ar spatial design with world in miniature and hybrid user interface. *IEEE Access*, 2025. 2

[21] JASP Team. JASP. 5

[22] M. Jeon, S. Park, and S. Oh. Portallite: A bare-hand 3d portal creation and manipulation technique for remote object interactions in virtual reality. *IEEE Access*, 2025. 2

[23] J. Kehrer and H. Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE transactions on visualization and computer graphics*, 19(3):495–513, 2012. 2

[24] K. Kim, J. V. Carlis, and D. F. Keefe. Comparison techniques utilized

in spatial 3d and 4d data visualizations: A survey and future directions. *Computers & Graphics*, 67:138–147, 2017. 2, 3

[25] J. Lee, Y. Chen, M. Gillies, and X. Pan. The effect of child and adult avatars on spatial perception in urban-scale virtual environment. In *2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*, pp. 19–26. IEEE, 2025. 2

[26] A. Li, J. Liu, M. Cordeil, J. Topliss, T. Piumsomboon, and B. Ens. Gesture-Explorer: Immersive Visualisation and Exploration of Gesture Data. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, pp. 1–16. Association for Computing Machinery, 2023. doi: 10.1145/3544548.3580678 2

[27] Y. Li, C.-W. Fu, and A. Hanson. Scalable wim: Effective exploration in large-scale astrophysical environments. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):1005–1012, 2006. 2

[28] J. Liu, A. Prouzeau, B. Ens, and T. Dwyer. Design and Evaluation of Interactive Small Multiples Data Visualisation in Immersive Spaces. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 588–597. doi: 10.1109/VR46266.2020.00081 2

[29] M.-J. Lobo, E. Pietriga, and C. Appert. An Evaluation of Interactive Map Comparison Techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pp. 3573–3582. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2702123.2702130 3

[30] J. Matejka, M. Glueck, E. Bradner, A. Hashemi, T. Grossman, and G. Fitzmaurice. Dream Lens: Exploration and Visualization of Large-Scale Generative Design Datasets. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 1–12. Association for Computing Machinery, 2018. doi: 10.1145/3173574.3173943 2

[31] R. Morey and J. Rouder. BayesFactor. 5

[32] J. W. Nam, K. McCullough, J. Tveite, M. M. Espinosa, C. H. Perry, B. T. Wilson, and D. F. Keefe. Worlds-in-Wedges: Combining Worlds-in-Miniature and Portals to Support Comparative Immersive Visualization of Forestry Data. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 747–755, 2019. doi: 10.1109/VR.2019.8797871 2

[33] K. Özacar, Y. Ortakcı, and M. Y. Küçükkara. Vrarcheducation: Redesigning building survey process in architectural education using collaborative virtual reality. *Computers & Graphics*, 113:1–9, 2023. 2

[34] D. Paes, E. Arantes, and J. Irizarry. Immersive environment for improving the understanding of architectural 3D models: Comparing user spatial perception between immersive and traditional virtual reality systems. *Automation in Construction*, 84, 2017. doi: 10.1016/j.autcon.2017.09.016 2

[35] D. Paes, J. Irizarry, and D. Pujoni. An evidence of cognitive benefits from immersive design review: Comparing three-dimensional perception and presence between immersive and non-immersive virtual environments. *Automation in Construction*, 130:103849, 2021. 2

[36] I. Paraschivoiu, R. Steiner, J. Wieser, and A. Meschtscherjakov. Crafting cities together: Co-located collaboration with augmented reality for urban design. *Computer Supported Cooperative Work (CSCW)*, 34(1):249–291, 2025. 2

[37] R. Pausch, T. Burnette, D. Brockway, and M. E. Weiblen. Navigation and locomotion in virtual worlds via flight into hand-held miniatures. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, SIGGRAPH '95. Association for Computing Machinery, New York, NY, USA, 1995. doi: 10.1145/218380.218495 2

[38] X. Qiu, L. Wen, C. Wu, Z. Yang, Q. Wang, H. Li, and D. Wang. Impact of learning methods on spatial knowledge acquisition. *Frontiers in Psychology*, 11:1322, 2020. 2

[39] J. C. Roberts. State of the Art: Coordinated & Multiple Views in Exploratory Visualization. In *Proceedings of the Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization*, CMV '07, pp. 61–71. IEEE Computer Society, USA, 2007. doi: 10.1109/CMV.2007.20 2

[40] J. N. Rouder, R. D. Morey, P. L. Speckman, and J. M. Province. Default Bayes factors for ANOVA designs. 56(5):356–374, 2012. doi: 10.1016/j.jmp.2012.08.001 5

[41] S. M. Saßmannshausen, J. Radtke, N. Bohn, H. Hussein, D. Randall, and V. Pipek. Citizen-centered design in urban planning: How augmented reality can be used in citizen participation processes. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*, pp. 250–265, 2021. 2

[42] K. A. Satriadi, B. Ens, M. Cordeil, T. Czauderna, and B. Jenny. Maps Around Me: 3D Multiview Layouts in Immersive Spaces. *Proc. ACM Hum.-Comput. Interact.*, 4(ISS):201:1–201:20, 2020. doi: 10.1145/3427329 9

[43] P. Schuchardt and D. A. Bowman. The benefits of immersion for spatial understanding of complex underground cave systems. In *Proceedings of the 2007 ACM symposium on Virtual reality software and technology*, VRST '07. Association for Computing Machinery, New York, NY, USA, 2007. doi: 10.1145/1315184.1315205 2

[44] A. W. Siegel and S. H. White. The development of spatial representations of large-scale environments. *Advances in child development and behavior*, 10:9–55, 1975. 2

[45] M. Slater, S. Wilbur, et al. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and virtual environments*, 6(6):603–616, 1997. 2

[46] M. Spur, V. Tourre, E. David, G. Moreau, and P. Le Callet. Exploring multiple and coordinated views for multilayered geospatial data in virtual reality. *Information*, 11(9):425, 2020. 3, 9

[47] M. C. Stellingwerff. Virtual context-investigating the characteristics and opportunities of digital visualisation media for situated approaches to architectural design in an urban environment. 2005. 2

[48] R. Stoakley, M. J. Conway, and R. Pausch. Virtual reality on a WIM: interactive worlds in miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95. ACM Press/Addison-Wesley Publishing Co., USA, 1995. doi: 10.1145/223904.223938 2

[49] S. Truman and S. Von Mammen. An Integrated Design of World-in-Miniature Navigation in Virtual Reality. *International Conference on the Foundations of Digital Games*, pp. 1–9, 2020. doi: 10.1145/3402942.3402994 2

[50] J. van Doorn, D. van den Bergh, U. Böhm, F. Dablander, K. Derks, T. Draws, A. Etz, N. J. Evans, Q. F. Gronau, J. M. Haaf, M. Hinne, Š. Kucharský, A. Ly, M. Marsman, D. Matzke, A. R. K. N. Gupta, A. Sarafoglou, A. Stefan, J. G. Voelkel, and E.-J. Wagenmakers. The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, 28(3):813–826, 2021. doi: 10.3758/s13423-020-01798-5 5

[51] D. Welti, M. Lutfallah, L. Cheng, and A. Kunz. Towards cross reality: A comparison of world-in-miniature and virtual reality switch. In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 1334–1335. IEEE, 2025. 2

[52] C. A. Wingrave, Y. Haciahmetoglu, and D. A. Bowman. Overcoming world in miniature limitations by a scaled and scrolling wim. In *3D User Interfaces (3DUI'06)*, pp. 11–16. IEEE, 2006. 2

[53] L. Zhang, A. Agrawal, S. Oney, and A. Guo. VRGit: A Version Control System for Collaborative Content Creation in Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, pp. 1–14. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3581136 2

[54] L. Zhang, M. Prather, D. Haynes, F. Lu, Y. Chen, and J. Li. Saving the Polar Bears: A Collaborative VR Storytelling Experience Enhanced by Multi-Scale and Multi-View WiM Designs. In *2025 Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2025. 2

[55] Y. Zhang, B. Ens, K. A. Satriadi, A. Prouzeau, and S. Goodwin. TimeTables: Embodied Exploration of Immersive Spatio-Temporal Data. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 599–605, 2022. doi: 10.1109/VR51125.2022.00080 2