Minimalism or Creative Chaos? On the Arrangement and Analysis of Numerous Scatterplots in Immersive 3D Knowledge Spaces

Melanie Derksen, Torsten Kuhlen, Mario Botsch, and Tim Weissker

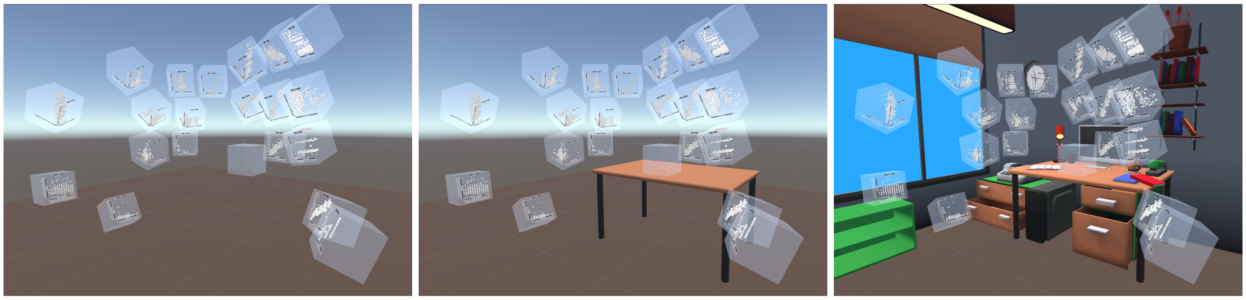

Fig. 1: We present a user study in which we investigate users' ability to memorize the location of a self-created scatterplot arrangement in three different virtual environments. While the *Empty* scenario (left) featured no environmental features at all, the *Office* scenario (right) offered an abundance of features that could be used as spatial references. The *Desk* scenario (middle) was situated between these extremes, offering only a single object for spatial orientation.

Abstract—Working with scatterplots is a classic everyday task for data analysts, which gets increasingly complex the more plots are required to form an understanding of the underlying data. To help analysts retrieve relevant plots more quickly when they are needed, immersive virtual environments (iVEs) provide them with the option to freely arrange scatterplots in the 3D space around them. In this paper, we investigate the impact of different virtual environments on the users' ability to quickly find and retrieve individual scatterplots from a larger collection. We tested three different scenarios, all having in common that users were able to position the plots freely in space according to their own needs, but each providing them with varying numbers of landmarks serving as visual cues: an *Empty* scene as a baseline condition, a single landmark condition with one prominent visual cue being a *Desk*, and a multiple landmarks condition being a virtual *Office*. Results from a between-subject investigation with 45 participants indicate that the time and effort users invest in arranging their plots within an iVE had a greater impact on memory performance than the design of the iVE itself. We report on the individual arrangement strategies that participants used to solve the task effectively and underline the importance of an active arrangement phase for supporting the spatial memorization of scatterplots in iVEs.

Index Terms—Virtual reality, 3D user interfaces, Head-mounted display, Immersive analytics, Spatial memory, Blind recall.

1 Introduction

In data analysis, scatterplots act as the first step in visualizing data, providing a solid visual basis for further data exploration, examining unusual patterns, finding correlations, identifying clusters, spotting outliers, and understanding general trends in the data [19, 22, 51]. With the increasing accessibility and popularity of Virtual Reality (VR) technology, the field of Immersive Analytics (IA) emerges as a more and more attractive and important area of research. Prior studies indicate that classic spatial analysis tasks, such as estimating distances [11], detecting outliers [60], and identifying clusters [30], can exhibit greater accuracy in immersive virtual environments (iVE) compared to traditional 2D desktop settings. Nonetheless, unlike with conventional 2D visualizations, standards and best practices for data analysis in iVEs still have to be established.

Besides the visual representation of the data items themselves, an important question to consider is how data items should be arranged in the surrounding virtual environment (VE) to enable the user to quickly locate the pieces of information relevant to them. Prior research shows that the ability to search for and remember the spatial arrangements of visualization elements within an information space is essential for carrying out visual analytical tasks and has a clearly positive impact on performance [6, 39]. Several papers also provide evidence that spatial aptitude is a strong predictor of performance in computer-based user interfaces (UIs) [16, 34, 56]. This emphasizes the significance of making appropriate design decisions to provide optimal support for individuals with varying levels of spatial abilities. A common strategy to support the spatial memorization of objects is the placement of salient reference objects, often referred to as landmarks [55, 58]. This strategy is particularly interesting for iVEs, given that these artificial worlds inherently provide less detail than the real world offers. However, an overuse of landmarks can also lead to visual clutter, which can distract the user and reduce their performance [44].

This paper investigates the strategies users employ to arrange numerous scatterplot objects for analysis tasks and explores whether different designs of iVEs influence these strategies and user's memory performance. Our work starts by presenting a design space encompassing the different factors investigated in prior studies on spatial memory, as well as using spatial cues in order to remember a list of items in iVEs. Based on this overview, we identify a research gap regarding the influence of the presence or absence of landmarks on scatterplot analysis tasks. We then present the results of an empirical between-subject user study with 45 participants to close this gap, in which we compared a completely *Empty* environment as well as an environment only containing a virtual *Desk* to a fully-modeled *Office* environment with an abundance of visual landmarks. In summary, our contributions are as follows:

- *Melanie Derksen and Mario Botsch are with TU Dortmund University. E-mail: {melanie.derksen | mario.botsch}@tu-dortmund.de*
- *Torsten Kuhlen is with RWTH Aachen University. E-mail: kuhlen@vr.rwth-aachen.de*
- *Tim Weissker is with RWTH Aachen University. E-mail: me@tim-weissker.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

- the derivation of a design space classifying related studies on (spatial) memory in iVEs regarding their *Main Independent Variable*, *Items of Interest*, and *Item Engagement*,
- quantitative results of a between-subject user study with 45 participants, indicating that the tested iVEs neither influenced memory performance nor task load significantly, and that the time participants spent arranging their plots is negatively correlated with error rates and positively correlated with recall time,
- qualitative results based on post-study interviews, showing that the assessment of complexity varied significantly, which may be attributed to differences in spatial reasoning skills, statistical skills, personal interests, and previous VR experience.

Our findings based on an exploratory follow-up analysis suggest that the time and effort users invest in arranging their plots within an iVE have a greater impact on memory performance than the design of the iVE itself.

2 Related Work

Previous research demonstrates the suitability of VR for data analysis and highlights its benefits over traditional 2D desktop setups, especially for spatial analysis tasks [11, 30, 41, 60]. For example, prior studies show that VR can facilitate the individual sensemaking process by allowing text and images to be arranged in 3D space [35, 36]. Furthermore, VR is shown to assist with the interpretation of more abstract data representations like plots, leading to a high engagement [1, 2, 59], an enhanced subjective perception of efficiency [59], and an increase of positive emotions [1]. Effectively conducting visual analytical tasks relies on the essential ability to locate and recall the spatial arrangements of visualization elements within an information space [6, 39], which is especially relevant when data can be distributed across the entire 3D space around the user. Given that spatial aptitude can vary across users and is shown to directly correlate with performance in computer-based UIs [16, 32, 34, 56], supporting the user's spatial memory is an essential component of interface design [53].

In the following, we start by highlighting the idea of introducing landmarks as a strategy to support spatial memory in both real and virtual environments (Sec. 2.1). Then, we present a more specific structured summary of prior research findings on spatial memory as well as on the use of spatial cues in order to remember abstract information in or with the help of iVEs (Sec. 2.2) to motivate our research focus of this paper (Sec. 2.3).

2.1 Landmarks for Supporting Spatial Memory

One of the most prominent strategies to support spatial memory is the placement of visually salient objects in the environment, often referred to as *landmarks* [48, 53], with respect to which the position of other objects can be memorized more easily. Vinson [58] underscores the importance of incorporating multiple landmarks into VEs and emphasizes that they should be (1) recognizable by differing greatly from one another and (2) discriminable from objects representing data. On the other hand, Quinn et al. [44] caution against the overuse of landmarks, finding that the extraneous visual clutter can distract users and decrease their performance. Fitzmaurice [14] introduces the concept of situated information spaces, where a data set is integrated into the real-world space such that landmarks and references from the real world can be used to assist with spatial memorization. Similarly, Büschel et al. [3] investigate the usage of spatial interaction with mobile devices for 3D data visualizations and propose using physical landmarks like a table to provide a general frame of reference to the users. Liu et al. [40] observe that their participants showed higher accuracy in recalling abstract spatial patterns and gave more favorable subjective ratings in an AR scenario when the room contained furniture, compared to an empty room. Uddin et al. [55] explore the use of anchor marks and a semi-transparent image as a way to improve people's spatial memory in 2D grid menus. The results indicate benefits especially in larger menus, with the simple anchor marks being faster and less error-prone than the visually richer images.

Even when the objects to memorize are not linked to a fixed spatial position by default, the concept of landmarks can still assist with their memorization. Using the so-called Method of Loci (MoL) [62], and even more specifically the concept of Memory Palaces (MP), a list of objects to memorize is mentally connected with landmarks along a route in a well-known environment, which facilitates later recall [31, 61]. Traditionally, the MoL is shown to work best in unique environments that are not repetitive and where there is plenty of space between the objects to be remembered [62]. The results of Legge et al. [33] indicate that participants find using a briefly presented VE in order to use it as basis for the MoL easier and more effective than the conventional MoL in which they are asked to use a very familiar environment, and find evidence that their virtual adaption may be more effective for memory enhancement or compensation training than the traditional MoL protocol.

Motivated by the successful implementation of landmarks in prior research as well as the cognitive benefits of landmark-based approaches like the MoL, our work investigates how these benefits can be applied to the recall of scatterplots in 3D iVEs and how VEs need to be designed to effectively support the spatial memorization of scatterplot objects.

2.2 Specific Findings on Spatial Memory in iVEs

A large body of related work has already investigated the influences of various variables on different forms of spatial memory in iVEs. To provide a structured overview of these papers and to situate our research presented in this paper, we follow the methodology formalized in Zwicky's General Morphological Analysis [49] to identify systematic differences within the presented user studies and derive the classification scheme shown in Tab. 1. In particular, we categorize papers based on three higher-level categories representing the *Main Independent Variable* under investigation (Sec. 2.2.1), the *Items of Interest* that are to be memorized (Sec. 2.2.2), and the level of *Item Engagement* that participants are provided with (Sec. 2.2.3).

2.2.1 Main Independent Variable

The category *Main Independent Variable* encompasses the primary subject of investigation, detailing what is either systematically varied or measured as uncontrolled factors to explore correlations with participant performance. We identify nine subcategories, all of which will be detailed in the following.

Environmental Features relates to the general appearance of the environment [18, 28, 31], also including the presence or absence of landmarks [39, 45] as well as anchors and background images in grid structures [17]. In addition to the brief introduction and the benefits of landmarks already presented in Sec. 2.1, the results presented here explicitly refer to research regarding landmarks in iVEs. Ragan et al. [45] control three independent variables: presentation layout (spatial vs. non-spatial), presence of landmarks, and the user's field-of-view (FoV). For the investigation of landmarks, they contrast the presentation of items on a solid background with a checkered environment where the items are presented on top of pillars. Their participants employ more visualization strategies during the memorization task when landmarks are included in the spatial presentations. Jund et al. [28] compare the aforementioned iVEs, labeling them as egocentric condition, with another iVE being a navigable virtual apartment, calling it the allocentric condition. They find that participants have better memory recall accuracy in their egocentric setting than in the allocentric one. The studies of Gao et al. [17] reveal that the integration of 3D pins as well as anchors can facilitate the retrieval and recall of multiple targets when compared to simple grid interfaces. Liu et al. [39] incorporate artificial landmarks into both flat and full-circular grid structures. Because flat layouts inherently provide natural landmarks at their corners, the added artificial landmarks aimed to make the two layouts more comparable. However, the flat layout consistently outperforms the full-circular one regardless of the inclusion of artificial landmarks. Han and Cho [18] observe that Augmented Reality (AR) participants outperform those in VR in terms of spatial memory performance, likely because the AR environment offers spatial cues that are absent in the VR setting. Krokos et al. [31] compare two VEs and call them virtual MPs (vMP),

Table 1: The design space categorizes studies regarding spatial memory and the association of information with spatial cues to increase memory performance in and with the help of iVEs. A filled cell indicates the option used in the respective work.

Work	Main Independent Variable								Items of Interest						Item Engagement			
	Environmental Features	Arrangement/ Layout Structure	Level of Immersion	Travel Metaphor	Awareness of Memorization Strategy	Knowledge of the Environment	FoV	Users' Perception	Data Points	2D Visual Entities	Words	Texts	Combinations	3D Objects	Passive	Hybrid	Active	N/A
Han and Cho [18]	■	□	□	■	□	□	□	□	■	□	□	□	□	□	■	□	□	□
Gao et al. [17]	■	□	□	□	□	□	□	□	□	■	□	□	□	□	■	□	□	□
Liu et al. [39]	■	■	□	■	□	□	■	□	□	■	□	□	□	□	■	■	□	□
Jund et al. [28]	■	■	□	■	□	■	□	□	□	■	□	□	□	□	■	□	□	□
Krokos et al. [31]	■	□	■	□	□	□	□	□	□	■	□	□	□	□	■	□	□	□
Yang et al. [61]	□	□	■	□	□	□	□	□	□	□	□	■	□	□	■	□	■	□
Häfner et al. [26]	□	□	■	■	□	□	□	□	□	□	□	□	□	■	■	□	□	□
Huttner et al. [24]	□	□	□	□	□	□	□	■	□	□	□	□	■	□	□	■	□	□
Huttner et al. [23]	□	□	■	□	■	□	□	□	□	□	□	□	■	□	□	■	□	□
Moll and Sykes [43]	□	□	■	□	□	□	□	□	□	□	□	□	■	□	□	□	■	□
Friedrich et al. [15]	□	□	□	■	□	□	□	□	□	■	□	□	□	□	□	□	■	□
Ragan et al. [45]	■	■	□	□	□	□	■	□	□	□	□	□	■	□	■	□	□	□
Huttner, Robra-Bissantz [25]	□	□	■	□	□	□	□	□	□	□	■	□	□	□	□	□	□	■
Vindenes et al. [57]	□	□	■	□	□	□	□	□	□	□	□	□	□	■	□	□	■	□
Ours	■	□	□	□	□	□	□	□	□	□	□	□	□	■	□	□	■	□

being an ornate palace and a medieval town in their special case. They do not find a statistically significant effect on recall due to the scenes.

Arrangement/Layout Structure comprises the differences in spatial arrangement [28, 45] of the *Items of Interest*, as well as the differences in grid curvature and its outer form [39] if a grid structure is used in the respective papers. Ragan et al. [45] demonstrate that spatial presentations in which items are presented at various positions around the user significantly enhance memory performance compared to non-spatial ones where the items are shown one at a time at the same fixed position in front of the user. Similarly, the participants of Jund et al. [28] benefit from the spatial presentation of items in the egocentric condition. In the study of Liu et al. [39], participants show a greater recall accuracy in flat and semicircular grid layouts than in full-circular layouts. Generally, studies prove that participants tend to prefer world-fixed over body-fixed displays [37], and semicircular or curved arrangements over full-circular and in some cases even over flat arrangements [37–39].

Level of Immersion refers to the comparison between systems that offer varying degrees of user immersion. This involves comparing the absence of any visual or virtual representation, allowing participants to rely solely on their own memorization strategy with a 2D image and with an iVE [61], the comparison of MoL with an iVE [23, 43], the comparison of MoL with a 2D desktop application and with an iVE [57], and the comparison of a 2D desktop application with an iVE [25, 26, 31]. Krokos et al. [31] discover that their vMPs offers superior memory recall when experienced with an head-mounted display (HMD) compared to a desktop setup. They conclude that iVEs provide great potential to enhance productivity through better recall of large amounts of information organized using the idea of vMPs. The vMP tested by Yang et al. [61] shows a moderate improvement for recall accuracy and precision over their 2D image condition. Huttner and Robra-Bissantz [25] investigate the impact of an increased level of immersion (by either offering a 2D desktop or an HMD to investigate the vMP their participants should make use of) on the effectiveness of the MoL. The VR group shows higher compliance and greater learning success compared to the desktop group. The findings of Moll and Sykes [43] indicate that a vMP experience can be optimized to enable participants to learn the MoL technique with minimal training time, potentially leading to substantial enhancements in recall performance. In the experiment of Huttner et al. [23], the group using the MoL outperforms the other two groups using an immersive vMP in terms of memory performance. However, they mention that some limitations of the experimental design may have contributed to the superior performance of the MoL group. The different levels of immersion in the work of Häfner et al. [26], being a CAVE and a desktop, do not influence memorization performance significantly. Vindenes et al. [57] let participants create personalized vMPs with varying immersion levels. While the MoL group outperforms the vMP groups, uneven spatial skills across groups make it unclear if the results are due to immersion or skill imbalance.

Travel Metaphor This subcategory refers to the locomotion technique that participants are provided with to traverse the VE. While participants may be encouraged to remain stationary by placing the *Items of Interest* only within a certain radius, other setups explicitly encourage physical walking by allowing to place items only beyond a certain radius [15]. Furthermore, virtual navigation can be enabled to navigate beyond the physically attainable space [18, 26, 28]. Han and Cho [18] investigate three 3D user interaction techniques and could show that physical walking is best to support spatial memory, followed by the direct manipulation of the objects instead of navigating towards them. Jund et al. [28] integrate a travel metaphor in their allocentric condition in order to let their participants navigate through the VE. However, the navigation technique shows no influence on the measurements. Friedrich et al. [15] and Häfner et al. [26] could not find a significant effect between motion types and memorization performance. Liu et al. [39] conclude that the main factor influencing participant performance is, even though being a side effect of their grid layouts, the type of physical navigation that comes along with the curvature, since a full-circular layout requires rotation while a flat layout requires physical walking.

Awareness of Memorization Strategy means whether or not the degree of knowledge regarding a memorization strategy is varied [23]. The memory performance in the study of Huttner et al. [23] is even without the conscious awareness of the application, or even the existence of the MoL not attenuated. They conclude that it is not necessary to introduce or educate the MoL before people enter a vMP. Contrary to that, the study of Legge et al. [33] shows that even with very little training, participants using either conventional MoL or virtual MoL significantly outperformed participants who were not instructed to use a particular strategy.

Knowledge of the Environment refers to the participants' familiarity of the environment [28]. Jund et al. [28] show that having prior knowledge of the architectural layout can help with a recall task but the recall performance still remains lower in the allocentric condition with the known environment than with their egocentric condition. Legge et al. [33] ask one of their groups to imagine a very familiar environment in order to use the MoL. Their results suggest that the VE is not significantly more difficult to use with the MoL than a personally familiar environment, and to the contrary, the MoL may be easier to use with the virtual MoL protocol. Caplan et al. [4] show that familiarity with an environment's structure can lead to better memory recall compared to other MP designs.

FoV The FoV can be artificially limited to experimentally isolate the influence of undesired side effects [39, 45]. Liu et al. [39] restrict the FoV to rule out grid overview opportunities as a factor, since flat layouts offer an overview by nature, unlike circular ones. Despite this restriction, the flat layout still outperforms the full-circular one, indicating that the difference in performance is not primarily due to the overview provided by the flat layout. Ragan et al. [45] argue that spatial perception is influenced by display factors and therefore vary the FoV, but they could not measure an effect on performance in their setup.

Users' Perception especially refers to the work of Huttner et al. [24], who explore correlations between users' attitude, their feeling of immersion and, among others, their learning success. They find significant correlations between the learning success and key factors of the users' intention to use a vMP.

2.2.2 Items of Interest

Items of Interest indicate the type of objects or information that are to be recalled by the users of the respective system. *Data Points* merely refer to points in a 3D scatterplot [18]. *2D Visual Entities* comprise two-dimensional rectangular shapes, which can be abstract uniform colored grid cells [39], pictograms [17] or images [15, 28, 31]. While *Words* represent individual unrelated words [25], the category *Texts* represents continuous texts that carry coherent information [61]. *Combinations* are a mix of varying kinds of items of the aforementioned types, like words plus images [23, 24, 43], or digits plus colored shapes [45]. *3D Objects* comprise virtual representations of real world objects [26] or so-called memory cubes, which are cubes with pictures projected to each side [57].

2.2.3 Item Engagement

Item Engagement denotes the degree of user involvement regarding the interaction with *Items of Interest*. *Passive* means that users are restricted to passively viewing the items without any possibility to interact with them [17, 18, 26, 28, 31, 39, 45, 61]. *Hybrid* denotes that there are only specific item interactions; some studies work with the concept of uncovering the items so that their actual content becomes visible by either point and click [23, 24] or by getting close enough to them [39]. *Active* means that users are able to position the items of interest freely in the corresponding VE [15, 43, 57, 61]. *N/A* implies that there are no virtual representations of those items spatially arranged in the VE itself and thus are presented separately [25].

Reviewing past work reveals a wide diversity in *Subjects of Investigation*. Some topics, like *Environmental Features* and *Level of Immersion*, received more attention than others, such as *Awareness of Memorization Strategies*, *Environmental Knowledge*, and *Users' Perception*. It became apparent that most of the studies focus on a *passive Item Engagement* with only four out of 14 works allowing for an *active Item Engagement*. Similar to the *Subjects of Investigation*, the *Items of Interest* highlight the heterogeneity of the research field. This raises questions about whether study outcomes would differ with other *Items of Interest*, how *Items of Interest* affect memorability, and the interplay between these items and the choice of *User Engagement*. Overall, the design space reveals that many combinations of the dimension variables offer opportunities for further research.

2.3 Discussion of Research Gaps

Based on our scoping review of prior studies on spatial memory performance in iVEs, we identified two research gaps that we will address in the following user study.

First, our review indicates that spatial memory research in the specific field of IA is relatively uncommon, even though the ability to locate and remember spatial arrangements of visualization elements within an information space is crucial for performing visual analytical tasks and can have a positive impact on performance [6, 39]. While one of the surveyed papers examines spatial memory in the context of a single prominent data plot [18], no research to our knowledge has yet focused on investigating spatial memory in the use case of working with multiple scatterplots in iVEs, a gap our work explicitly addresses.

Second, the aforementioned research indicates that spatial arrangements [28, 45] and the use of landmarks [17] have the potential to enhance memory performance, but also that an overuse of landmarks might be counterproductive [44]. However, the research papers discussed earlier mostly consider the presence and absence of certain types of landmarks as binary variables [18, 28, 45] without systematically varying the number of environmental features in a structured way. To approach this gap, our study explicitly introduces a condition between these two extremes to investigate whether a restricted set of only a single landmark is already sufficient to support spatial memory.

3 User Study

We conducted an empirical between-subject user study, in which we analyzed the effects of different types of *Environmental Features* on the recall performance of scatterplots and investigated the individual arrangement strategies. Participants were assigned to one of three VEs and were asked to actively (*Item Engagement*) distribute 2D and 3D scatterplot objects (*Items of Interest*) within that space. They were then required to recall the locations of certain plots in both a blind and a visible recall phase, which served as the main dependent variable for this experiment. The last row of Tab. 1 situates our study within the previously established design space of related work.

3.1 Hardware Setup

The application for our experiment is based on Unity version 2022.3.15f1. For the study, we used the *Meta Quest 3*, which has a resolution of 2064×2208 pixels per eye and an update rate of 120 Hz. Participants were sitting on a swivel chair and had the default interaction space of the Meta Quest's stationary mode (1 m $\times$ 1 m). For the creation of scatterplots, we used the Immersive Analytics Toolkit (IATK) [8], a visualization toolkit for Unity that forms a solid basis for the individual creation of multidimensional data visualizations in VEs.

Our focus on a stationary, seated VR experience was motivated by the daily work routines of data analysts that commonly take place in a desk-based office environment with limited movement space. Furthermore, prior research underlined that seated VR experiences generate less fatigue than standing experiences [5] and also tend to induce lower levels of cybersickness [42, 47]. In addition, prior work has not identified significant interactions between motion types and memorization performance [15, 26].

3.2 Conditions

The between-subjects factor of our experiment was the number of *Environmental Features* that the user could make use of to arrange and later recall the scatterplots. We decided to test three different levels, resulting in three different VEs ranging from the complete absence of environmental features (*Empty* condition) to the availability of a single virtual desk as feature (*Desk* condition) to a fully-modeled office space with ample features (*Office* condition). The furniture models in the

Desk and *Office* conditions were taken from a free asset pack called *Office Softpack*[1].

Empty The *Empty* condition (Fig. 1, left) serves a the baseline and merely consists of a ground plane surrounded by the default skybox of Unity. This condition helps to judge the potential improvements introduced by the other conditions by studying if the user-selected arrangement of scatterplots in empty space is already sufficient to support easy memorization.

Desk The *Desk* condition (Fig. 1, center) features a single desk object on the ground plane that can be used as a reference object for arranging scatterplots. It was inspired by previous work suggesting that a single prominent landmark might already be sufficient to support easy memorization of multiple objects [3, 59].

Office The *Office* condition (Fig. 1, right) resembles the everyday work scenario of data analysts by featuring a desk with additional objects like a monitor screen, a keyboard, some wall decorations, folders, etc. This condition was inspired by previous work on landmark-based memorization benefits and the finding that familiar environments can improve memory recall [4].

3.3 Procedure

The host institution did not require an ethics approval for the conducted study. Users were seated and did not make use of virtual navigation techniques associated with the elicitation of sickness symptoms. Participants came to our lab, were informed about the purpose of the study, and agreed to participate voluntarily. Then they completed a digital version of the so-called Corsi block-tapping test[2] [54], a psychological test that assesses visuo-spatial short-term working memory [9]. In this test, participants had to memorize a gradually increasing sequence of blinking blocks among nine randomly ordered blocks. The Corsi span is defined as the longest sequence a participant can correctly repeat, whereby an average Corsi span is somewhere between five and seven blocks for normal human subjects [29]. We then balanced the participants across the conditions based on their gender as well as their Corsi span. To ensure everyone had the same fundamental knowledge required for participating in the study, we had participants review some informative slides on scatterplots, allowing them to ask questions at any time. After explaining how to use the VR system, the VR experience started with a training of the controls, followed by the actual study that consisted of the three distinct phases *Arrangement*, *Blind Recall*, and *Analysis*. The participants were only given instructions for the arrangement and analysis phases. The blind recall phase was not announced in advance, ensuring that participants arranged their layouts specifically for the analysis tasks. During a training stage, participants arranged six plots and answered two example analysis questions based on classic scatterplot tasks [51], such as spotting outliers and identifying correlations in order to familiarize them with the tasks they would perform in the actual study. After each of the three phases in the actual study, we asked participants to fill in the Raw-TLX questionnaire to quantify perceived task load [20, 21] as well as the Discomfort Scale, which consists of the one question *"On a scale of 0 to 10, 0 being how you felt coming in, 10 is that you want to stop, where you are now?"*, to quantify their overall well-being [12, 46].

Phase 1: Arrangement In the first phase, participants were asked to arrange a total of 20 scatterplots (ten with two axes, ten with three axes) in the space around them. To do so, they pressed a button on their controller, which spawned the next scatterplot at a predefined position in front of them. The participants had two controllers with identical functionality. Each controller featured a ray pointer, enabling users to drag and drop plots. Additionally, they could perform two-handed rotations. Participants were instructed to drag each of these plots to a location in 3D space, creating an arrangement of plots that would help them to quickly locate and retrieve relevant plots in later analysis tasks. This experimental paradigm is commonly referred to as *placement-retrieval* and is a well-established method in the literature on measuring spatial memory [6, 15, 27, 50].

The scatterplots for the training were based on data from Crawford's *Cereals* dataset[3], and for the actual study from Crawford's *Camera* dataset[4] as well as Quinlan's *Auto MPG* dataset[5] in which the units were converted to match the conventions of our host institution. It is likely that no one was an expert in the specifics of the datasets, but it cannot be ruled out that someone may have encountered some of the datasets before, as these are publicly available standard datasets. All participants saw the plots in the same order.

Each scatterplot could be dragged and dropped in a world-fixed manner as done in previous studies [10, 37]. Since users were observed to consider rearranging earlier objects during a task [52], we explicitly allowed the repositioning of earlier scatterplots in our study as well. The next plot could only be spawned if existing plots were neither overlapping nor situated in the predefined plot spawning area. Invalid positions were indicated by a semi-transparent red coloring of the bounding box of the respective plots. We measured the total time participants took to complete this phase.

Phase 2: Blind Recall After participants were satisfied with their arrangement of all scatterplots, the second phase began, in which each plot was replaced with a solid black box. This phase was not announced beforehand. Participants were then shown a total of ten plots (five with two, five with three axes) and asked to select the black box that corresponds to the position at which they placed this plot in the *Arrangement* phase. The literature refers to this experimental paradigm as *blind recall*, which is relevant to eliminate visual search strategies and, therefore, to measure participants' spatial recall performance in isolation [15, 27, 53].

Similar to the *Arrangement* phase, users could press a button to show the next scatterplot to be recalled in a predefined spawning position. They then used simple raycasting to select the box where they assumed the original plot was located. The chosen box briefly turned green with a white check mark if the selection was correct or red with a white cross if it was incorrect. Only one attempt per plot was permitted. For each task, we measured the time between the spawning of the reference plot to the corresponding click on a box (Retrieval Time). In the case of an error, we followed previous research [15, 27] and measured the straight-line distance from the center of the selected plot to the correct plot (Euclidean Error Distance) as well as the number of scatterplot boxes that were closer to the correct one (Items Closer to Correct One, similar to [15, 18]).

Phase 3: Analysis Although blind recall tasks are beneficial in terms of experimental control, the resulting experimental procedure also appears more artificial and, therefore, decreases ecological validity. To overcome this limitation, the third phase of the study confronted participants with more realistic data analysis tasks, in which the black boxes were removed again to show the initially placed scatterplots.

We asked participants a total of five questions on the underlying datasets that could be answered with one of the previously arranged plots. We alternated in asking for plots with either two or three attributes as well as for the two datasets. The general kind of analysis questions were derived from classic scatterplot analysis tasks [51]. Questions were displayed in text form and required participants to deduce which of the plots is required, to find it in the arrangement, and to investigate it for the correct answer. Once the correct answer was given verbally, the experimenter logged it in and initiated the display of the next task in the iVE. While participants were allowed to rearrange plots while searching for the correct answer, the arrangement was reset to the one of the *Arrangement* phase after each question. For each question, we measured the time between task assignment and verbal answer.

[1] https://nappin.itch.io/office-props-softpack

[2] https://www.psytoolkit.org/experiment-library/corsi2.html

[3] https://www.kaggle.com/datasets/crawford/80-cereals

[4] https://www.kaggle.com/datasets/crawford/1000-cameras-dataset

[5] https://archive.ics.uci.edu/dataset/9/auto+mpg

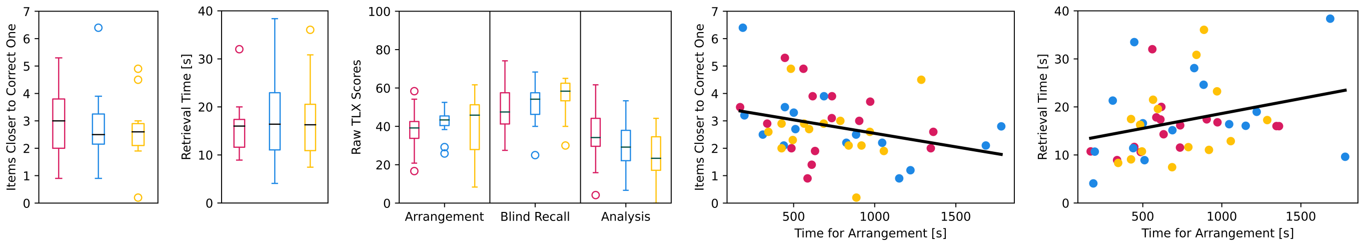

Fig. 2: The boxplots (from left to right) illustrate the distribution of Items Closer to Correct One, the Retrieval Time in *Blind Recall*, the task load in the different phases separated by condition. The two scatterplots (from left to right) illustrate the time spent in the *Arrangement* in relation to the average Items Closer to Correct One per participant, as well as the average Retrieval Time per participant. In both plots, a fitted black regression line illustrates the overall trend. The colors represent the ■ *Empty*, ■ *Desk*, and ■ *Office* condition.

Study Conclusion At the end of the study, participants were asked to describe in their own words how challenging they found every stage as well as their strategies for arranging the plots, and if they have further comments. They then provided demographic data before receiving an expense allowance of 12 Euros for their efforts. The entire study procedure took approximately 60 min to complete.

3.4 Hypotheses

Our primary analysis goal is to investigate the impact of different iVEs and arrangement strategies on spatial memory in the specific use case of working with multiple scatterplots, focusing on how variations in the level of features within these VEs affect the outcome. Previous research strongly suggests that the environment will have an impact as, for example, using vMPs can lead to improved memory performance [31, 61], and landmarks and visual anchors can aid recall [17]. However, an overuse of landmarks can also cause visual clutter with a negative impact on spatial memory [44]. Therefore, we formulated undirected hypotheses as follows:

H1: The iVE will have an impact on recall performance, this means on the Items Closer to Correct One (H1a) and Retrieval Time (H1b) in the *Blind Recall* stage.

Considering the importance of spatial anchors for the development of spatial memory [53], we assume that the *Empty* condition will lead to the highest task load among the three conditions as it is characterized by the explicit absence of visual anchors:

H2: The task load in the *Blind Recall* as well as in the *Analysis* phase will be highest in the *Empty* condition.

Despite the hypothesized dependence of the results on the iVE, we also assume that longer arrangement times might also enhance the recall performance, since longer times might be due to the development and adjustment of a sophisticated arrangement strategy that reacts appropriately to new information as new plots are added:

H3: Participants who take more time in the *Arrangement* phase will have lower error rates, this means fewer Items Closer to Correct One (H3a), and will be faster in recalling the plots in the *Blind Recall* phase (H3b).

3.5 Participants

A total of 45 participants (20 female, 25 male), between 18 and 37 years of age ($M = 25.4$, $\sigma = 4.6$) with a Corsi span ranging from 4 to 9 ($M = 6.4$, $\sigma = 1.3$), were recruited from the local university campus and through dedicated mailing lists for our user study. This provided 15 participants per condition, balanced in terms of both gender and Corsi span. In the *Empty* condition, there were 7 females and 8 males, with a mean Corsi span of 6.47. The *Desk* condition also comprised 7 females and 8 males, with a mean Corsi span of 6.33. Finally, the *Office* condition included 6 females and 9 males, with a mean Corsi span of 6.4. Based on a 5-point Likert scale, where 1 represents *very little* and 5 represents *very much*, the self-reported mean prior experiences were as follows: VR experience averaged at 2.5, gaming experience at 3.6, data visualization experience at 2.9, and scatterplot experience at 2.6. Notably, the full range of the scale was utilized across all categories.

4 Results

This section reports on the results of our data analysis, which we begin by an overall analysis of discomfort scores as an indicator for the validity of all other measurements. In particular, the discomfort scores reported after each phase (with a possible range from 0 to 10) ranged from 0 to 4 ($M = 0.93, \sigma = 1.25$). The majority of scores ($N = 88$) were between 0 and 1, with only a small number ($N = 7$) reaching a score of 4. Overall, these figures suggest that participants were in good shape to complete the study, which implies that the other measurements are likely not affected by discomfort as a confounding variable.

4.1 Hypotheses

Based on our formulated hypotheses before the experiment, we conducted our inferential statistical analyses using *Jamovi* (version 3.2.28). For the comparisons of the three iVEs, we ran one-way ANOVAs and tested the assumption of normality with the Shapiro-Wilk test and homogeneity of variances with Levene's test. We switched to Kruskal-Wallis tests if the assumptions were not fulfilled.

To prevent an over-reliance on p-values, we supplement our reports with the effect size η^2 for ANOVAs as suggested by the APA publication guidelines, applying the threshold values of $\eta^2 > 0.01$ (small), $\eta^2 > 0.06$ (medium), and $\eta^2 > 0.14$ (large) as suggested by Cohen [7] to quantify the corresponding effect magnitudes. Correlational analyses of the entire dataset were performed by computing Pearson's r, given that the combined sample size of $N = 45 > 30$ is sufficiently large to assume a normal sampling distribution based on the central limit theorem [13, pp. 170–172]. Plots illustrating the distributions are given in Fig. 2.

Recall Performance (H_1) The average number of Items Closer to Correct One were 2.64 in the Office condition ($\sigma = 1.09$), 2.77 in the Desk condition ($\sigma = 1.28$), and 3.00 in the Empty condition ($\sigma = 1.24$). A one-way ANOVA on the data did not reveal a significant difference between the conditions, $F(2,42) = 0.343, p = 0.712, \eta^2 = 0.016$ (small effect).

The average Retrieval Time was with 15.8 s the shortest in the Empty condition ($\sigma = 5.53$ s), 16.9 s in the Office ($\sigma = 8.33$ s) and longest 18.3 s for Desk ($\sigma = 9.53$ s). The average Retrieval Time was not significantly affected by the iVE, $\chi^2(2) = 0.200, p = 0.905, \eta^2 = 0.005$.

Task Load (H_2) The Empty condition induced on a average the lowest task load in the *Blind Recall* phase with 50.3 ($\sigma = 12.7$), followed by Desk 51.6 ($\sigma = 10.6$) and Office 55.8 ($\sigma = 9.77$). A one-way ANOVA showed that the mean task load in the *Blind Recall* phase was not significantly affected by the iVE, $F(2,42) = 0.995, p = 0.378, \eta^2 = 0.045$ (small effect).

In the *Analysis* phase the Empty condition induced on average the highest task load 35.6 ($\sigma = 14.0$), followed by Desk with 29.8 ($\sigma = 13.2$) and Office with 24.6 ($\sigma = 13.0$). A one-way ANOVA showed that the mean task load in the *Analysis* phase was not significantly affected by the iVE, $F(2,42) = 2.56, p = 0.089, \eta^2 = 0.109$ (medium effect).

Correlational Analyses (H_3) Correlational analyses revealed a significant negative linear relationship between the arrangement time and the average number of Items Closer to Correct One, $r = -0.311, p = 0.038$.

Furthermore, the analyses revealed a significant positive linear relationship between the arrangement time and the average recall time in the *Blind Recall* phase, $r = 0.296, p = 0.048$.

4.2 Discussion and Exploratory Follow-Up Analyses

Contrary to our expectations motivated by related work, our inferential analyses indicated that the number of landmarks present in each of the tested iVEs did neither significantly impact recall performance nor task load, leading us to reject H_1 and H_2 for our experiment. Furthermore, we found that a longer arrangement time resulted in smaller error rates, which confirms H3a. Surprisingly, the opposite effect predicted in H3b occurred: a longer arrangement time led to a longer recall time.

4.2.1 Arrangement Complexity

Given that our study could not detect significant influences of the iVEs on recall performance (H_1), we were interested in other factors that could have determined how well participants performed in the *Blind Recall* phase. To this end, the results of our correlational analyses provided us with indications that, independent of the iVE, the time spent in the *Arrangement* phase was a significant predictor of the number of Items Closer to Correct One with a medium effect of $r = -0.311$, even though higher arrangement times also resulted in higher recall times with $r = 0.296$ contrary to our expectations (H_3). The negative relationship between arrangement time and error suggests that better performing participants might have devised a more sophisticated strategy in the *Arrangement* phase, which took them longer to optimize before moving on to the *Blind Recall* phase.

To test this claim numerically, we decided to analyze the final scatterplot arrangements of all participants in more detail. Based on our observations of the study, we hypothesized that participants without a clear strategy mainly arranged plots based on their spawning order, while participants with a strategy clustered and ordered plots based on semantics rather than spawning order. To quantify this, we had a look at the 19 pairs of neighboring scatterplots in the spawning order and investigated their Euclidean distances in the final arrangement created by each participant. We then normalized this distance vector for each participant by dividing each value by the longest distance in order to account for individual variations in the overall spatial extent of each arrangement. As a final step, we computed the mean value of this normalized vector as a single numeric score for each participant. Correlational analyses revealed that this score had a significantly negative relationship with the average number of Items Closer to Correct One with a medium effect size, $r = -0.371$, $p = 0.012$. This indicates that arrangements deviating stronger from the spawning order were related to better results in the recall phase, confirming our assumption that the participants' arrangement strategy was a stronger predictor of recall performance than the iVE. In the following, we will summarize some of the most prevalent strategies participants mentioned to have used.

4.2.2 Arrangement Strategies

Overall, we observed that participants used a large variety of different strategies for arranging the scatterplots. All of them had in common that participants made extensive use of the ability to freely arrange scatterplots in the 3D space, resulting in layouts that were either flat or centered around the user in a curved manner, and were based on characteristics of the presented data (e.g., their respective dataset, their number of axes, their attributes, etc) rather than being connected or bound to environmental features. To improve the visibility of the plots in these user-centered arrangements, 30 participants slightly tilted the plots such that all axes were directly visible from the central position.

All participants clustered scatterplots based on the two underlying datasets (cameras and cars), 43 of which used a vertical split between the clusters (one horizontal, one slanted). 36 participants created a clear visual gap between the clusters, while the remaining nine participants juxtaposed both clusters without a clearly visible separation. Another factor for creating clusters was the number of axes in the scatterplots (2D or 3D), where 14 and six participants used vertical and horizontal separation, respectively.

23 participants created clusters based on the attributes visible in each scatterplot. Among these, ten participants arranged the plots such that identical attribute axes were all placed along the same line in 3D space. Six participants decided to rotate plots such that certain axes of neighboring plots were identical. Four participants included specific semantic interpretations of the attributes in the provided scene, for example, attributes representing "weight" being situated near the ground or attributes representing *maxima* (e.g., "maximal resolution" in the camera dataset) being situated at a higher level in the VE. One participant considered the alphabetical order of the attributes, while another one focused on grouping similar data distribution shapes together.

4.2.3 Exemplary Arrangements

To better demonstrate the variety of arrangement strategies as well as the resulting user performance in the blind recall task, we will discuss two exemplary participant arrangements in more detail.

Figure 3 shows the arrangement of participant P3, who achieved the overall lowest error rate with an average Items Closer to Correct One score of 0.2, just one error out of ten blind recalls, and an average Retrieval Time of 36.06 s. The user spent 14.9 min in the *Arrangement* phase and was in the *Office* condition. In the interview the participant explained they put the car plots on the right-hand side and the camera plots to the left-hand side, grouped by attributes in a grid-like manner and tried, whenever possible, to arrange the same attributes along a line. The resulting arrangement shows that the plots are spatially clearly separated based on the dataset they represent and grouped into flat clusters. Plots that share a common attribute are neighboring and often aligned along a line or arranged in clusters in the iVE, regardless of which plot axis displays that attribute. For example, plots containing "acceleration" (with their numbers in spawning order being 5 and 7) are placed horizontally at the top, plots containing "displacement" (1, 4, 7, 13, 18) are arranged vertically, and plots containing "model" (1, 4, 10) are positioned at the bottom, forming a triangular shape. Figure 3 also demonstrates that this participant's arrangement significantly deviated from the original spawning order; it rather exhibits a clear semantic organization. In the interview this participant called the *Arrangement* phase the most challenging one, which once again suggests that the effort invested during this stage paid off in the *Blind Recall*, leading to a low error rate.

In contrast, Fig. 4 shows the arrangement of participant P12, who achieved the overall highest error rate with an average Items Closer to Correct One score of 6.4, eight errors out of ten blind recalls, and the fastest average Retrieval Time of 4.07 s. The user spent 3.1 min in the *Arrangement* phase, thus being the second-fastest, and was in the *Desk* condition. In the interview the participant mentioned they wanted to be able to see all plots by moving the head slightly up and down. Furthermore, they explained they tried to align plots having the same attributes along a line, which is similar to the strategy of participant P3. In the resulting arrangement in Fig. 4, the plots were separated horizontally depending on their dataset but, in contrast to participant P3, there is no clear spatial gap between the two different datasets. Furthermore, a deeper investigation shows that the intention to align plots with the same attributes along a line has been only partially realized. For example, for plots having the attribute "model" (1, 4, 10) this concept applies only to 4 and 10, while 1 is neither in the same line nor spatially close to 4 and 10. Similarly, plots having the attribute "# cylinders" (3, 7, 10, 16) neither share a single line nor are positioned in spatial proximity. 3, 7 and 16 are arranged along a vertical line but are not neighboring, only 7 and 10 are adjacent in the participant's arrangement, just to mention a few. In addition, the visualization of the spawning order together with the short arrangement time creates the impression that newly spawned plots were rather stacked on top of each other or had been successively added to the existing arrangement. It seems there was no effort made to

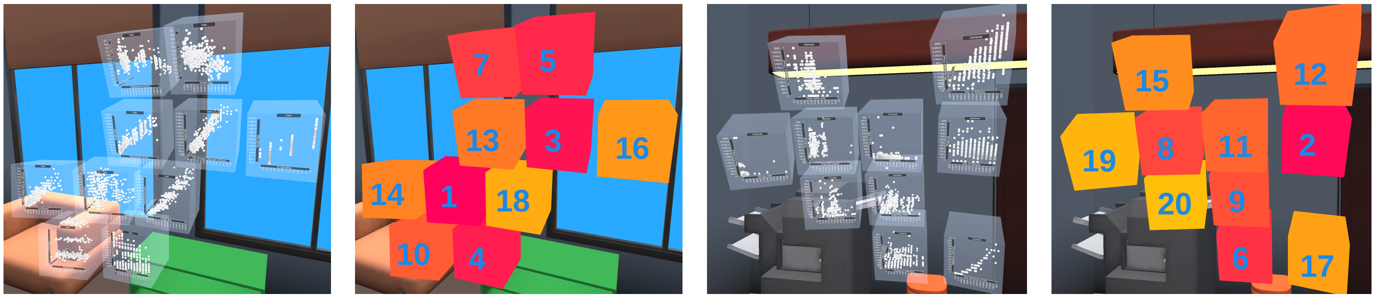

Fig. 3: The pictures (from left to right) show the arrangement of participant P3 of the plots representing the car dataset, their respective spawning order, the arrangement of the plots representing the camera dataset, and their respective spawning order. For demonstrating the spawning order, the color ■ represents the first plot (number 1), with interpolated colors towards ■ marking the sequence from plot 2 to 20.

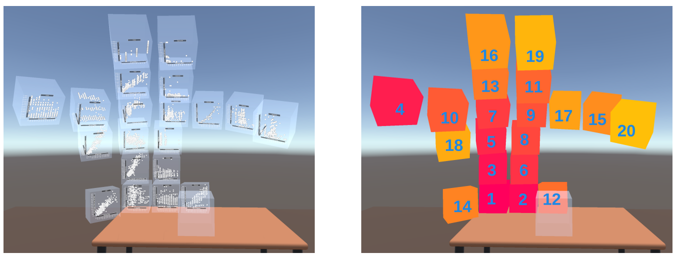

Fig. 4: The arrangement of participant P12 was strongly inspired by the spawning order of the plots instead of their semantics. See Fig. 3 for an explanation of the employed color coding.

rearrange the existing plots for refinement. Despite being told to have unlimited time for the *Arrangement* phase, the participant's seemingly hurried attempt led to an arrangement that only partially reflected their actual intentions. The brief arrangement time likely limited the time to examine and memorize individual plots, resulting in higher error rates.

4.2.4 Semi-Structured Interview

Below, we summarize the participants' descriptions, in their own words, regarding how challenging they found each stage.

Arrangement While two participants described the *Arrangement* as the most complex phase, five participants reflected that it was more complex than they initially realized and felt they should have spent more time on it. 13 found it quite difficult, one attributed this to their weakness in spatial reasoning, another mentioned their unorganized nature which they felt was a disadvantage in that case, and one even stated that creating a meaningful arrangement seemed impossible. 17 participants called it tricky/not so easy to develop a good system, some expressing an uncertainty about what to sort by and others lost the overview. One participant mentioned that they could have spent hours in arranging the plots because they found the task enjoyable. Five found it easy, and two considered it very easy. This widespread nature of oral responses regarding the complexity of this phase is also mirrored in the task load in Fig. 2.

When asking for further comments, one participant in the *Office* condition said that it would be nice if the VE's context matched the data presented in the plots in order to build environmental associations with the plot's content, and another one in the *Office* condition mentioned that the surroundings helped in estimating distances and groups, providing a better sense of space. Two participants appreciated the ability to position the plots freely in space, without being restricted by gravity. Our observations revealed that once participants understood the plots were not influenced by gravity, they began "ignoring" environmental features in the *Desk* and *Office* condition. Initially, some tried placing the plots on desks or shelves, but after recognizing the absence of gravity, they let the plots float freely within the iVE, independent of the surroundings.

Blind Recall Regarding the *Blind Recall* phase, 18 participants found it to be very complex, one of them pointing out that they were very happy they had devised a good system in the previous phase. 19 said it was quite difficult. Six classified it to be moderate in complexity. Two said it was easy but both clarified that this was because they simply guessed. Among those complexity estimations, 13 participants additionally stated that they could only vaguely remember the positions of the plots but not precisely, and one stating that it became pretty obvious in which areas they spent enough time thinking about their arrangement and in which they did not. The feedback goes hand in hand with the higher scores of the task load in the *Blind Recall* in Fig. 2.

Analysis The *Analysis* phase was stated to be complex by only two and to be moderately complex by 13 participants. 25 classified it as easy, some mentioning the intuitiveness of the VR system, some mentioning their arrangement was specifically designed to solve analysis tasks, while others described the process as relaxing despite being instructed to find the answers as quickly as possible. The remaining five directly compared it to the *Blind Recall* phase and found the *Analysis* to be easier. Among those opinions, three especially highlighted the fun aspect in the VR system during the analysis tasks. One of them, who had no statistical background, mentioned that they would have liked to continue with more tasks and found it enjoyable because of their good system they had created that made sense to them. These responses indicate that the *Analysis* phase seems to be the easiest one, which is also mirrored by the lower task load scores in Fig. 2.

Regarding further comments, it stood out that nobody complained about visual clutter, but one participant in the *Empty* condition mentioned that the skybox made the numbers in the plots hard to read. One participant in the *Office* condition mentioned that the environmental features present in that particular VE did not create visual clutter. However, they could imagine that other scenarios with poor color choices might do so. Two said that they found working with the camera dataset easier than with the car dataset across all phases due to a generally higher personal interest regarding cameras.

Overall, it stands out that the participants' opinions on the complexity of each phase show high variances, but that the overall trend is mirrored in the task load in Fig. 2. From additional comments, we conclude that this discrepancy (beyond the varying levels of experience with VR technology and scatterplots revealed by the quantitative data) may be attributed to differences in spatial reasoning skills, personal interests, and the amount of time and effort invested in the *Arrangement* phase, which served as a basis for subsequent phases.

4.3 Limitations

In the following, we would like to list a few limitations of our study design to motivate relevant aspects of future work.

First of all, with our study design we could not detect statistical significance for two of our three hypotheses. Nonetheless, we provided further exploratory analyses beyond our hypotheses to provide potential explanations for the observed data. We explicitly encourage the replication of our experiment by other researchers to see if these exploratory findings replicate with a different sample.

Furthermore, post-hoc analyses revealed that our participants had varying levels of experience across different areas, including VR technology, data visualizations, and scatterplots. Although all participants successfully completed the analysis tasks, the question remains as to how transferable our results are to professional data analysts since the understanding of the data and plots might have influenced the arrangements. In addition, we did not provide our participants with a specific arrangement or memorization strategy, as prior research is divided on whether introducing such strategies is necessary [23, 33]. Moreover, our participants actively positioned the items themselves, rather than being restricted to memorize a static, predefined arrangement. This approach differs from studies such as those by Ragan et al. [45], Krokos et al. [31], and Jund et al. [28], which also explore different VEs but use predefined item positions for participants to passively observe and memorize.

Additionally, the *Items of Interest* in our study differ markedly to previous work. While earlier studies mostly used abstract items, like *2D Visual Entities* [15, 17, 28, 31, 39] or unrelated *Words* [25], only one study uses whole texts [61] which are, similar to ours, rich in content but, contrary to ours, not interrelated. In contrast, scatterplots inherently carry a high level of information, show a greater complexity, and, in our case, are interrelated. Presumably, these factors also influence the complexity of the arrangement process.

All in all, spatial memory is a complex cognitive construct influenced by many factors that cannot always be isolated optimally.

5 Conclusion and Future Work

In this work, we introduced a design space categorizing studies on spatial memory as well as the association of abstract information with spatial cues in order to increase memory performance in and with the help of iVEs. We then investigated how individual strategies for arranging multiple scatterplot objects across three different iVEs with varying amounts of environmental features influence spatial memory performance. Unexpectedly, our user study did not reveal a significant effect of the different iVEs on memory performance, but the time participants took for their individual arrangement was a significant predictor of user performance instead. In combination with an exploratory follow-up analysis, our study's findings indicate that the time and effort invested in arranging items seem to have a stronger influence on spatial memory performance than the environmental features. As a result, we conclude that giving users the freedom to create their own arrangement instead of providing them with pre-existing arrangements can be beneficial for supporting spatial memory in data analysis scenarios.

Future work might focus on how transferable our results regarding the impact of environmental features are when using less complex *Items of Interest* that are not interrelated and have a lower density of information. Furthermore, our interview revealed that some participants found working with one dataset easier than with the other one due to individual differences in interests regarding specific topics, which makes personal interests another relevant factor to examine in future studies. On top of that, it makes sense to investigate how transferable our results are to professional data analysts since the understanding of the data and plots might have influenced the arrangements. All in all, we believe that the investigation of spatial memory in abstract data analysis scenarios is a highly relevant area for future research. Our presented design space as well as our study results provide further insights into the large variety of potential factors influencing memory performance and, therefore, pave the way for relevant future studies in the field.

Acknowledgments

This work was supported by "Stiftung Innovation in der Hochschullehre" through the project "Hybrid Learning Center" (FBM2020-EA-690-01130). We are grateful to all subjects attending our user study and are sincerely grateful to Sven Wagner for his temporary assistance in conducting the study as an experimenter.

References

[1] D. Allcoat and A. von Mühlenen. Learning in virtual reality: Effects on performance, emotion and engagement. *Research in Learning Technology*, 26, 2018. doi: 10.25304/rlt.v26.2140 2

[2] A. Batch, A. Cunningham, M. Cordeil, N. Elmqvist, T. Dwyer, B. H. Thomas, and K. Marriott. There Is No Spoon: Evaluating Performance, Space Use, and Presence with Expert Domain Users in Immersive Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):536–546, 2020. doi: 10.1109/TVCG.2019.2934803 2

[3] W. Büschel, P. Reipschläger, R. Langner, and R. Dachselt. Investigating the Use of Spatial Interaction for 3D Data Visualization on Mobile Devices. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, ISS '17, pp. 62–71. Association for Computing Machinery, 2017. doi: 10.1145/3132272.3134125 2, 5

[4] J. B. Caplan, E. L. Legge, B. Cheng, and C. R. Madan. Effectiveness of the method of loci is only minimally related to factors that should influence imagined navigation. *Quarterly Journal of Experimental Psychology*, 72(10):2541–2553, 2019. doi: 10.1177/1747021819858041 4, 5

[5] M. R. Chester, M. J. Rys, and S. A. Konz. Leg swelling, comfort and fatigue when sitting, standing, and sit/standing. *International Journal of Industrial Ergonomics*, 29(5):289–296, 2002. doi: 10.1016/S0169-8141(01)00069-5 4

[6] A. Cockburn and B. McKenzie. Evaluating the Effectiveness of Spatial Memory in 2D and 3D Physical and Virtual Environments. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 8 pages, pp. 203–210, 2002. doi: 10.1145/503376.503413 1, 2, 4, 5

[7] J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Routledge, second ed., 2013. doi: 10.4324/9780203771587 6

[8] M. Cordeil, A. Cunningham, B. Bach, C. Hurter, B. H. Thomas, K. Marriott, and T. Dwyer. IATK: An Immersive Analytics Toolkit. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 200–209, 2019. doi: 10.1109/VR.2019.8797978 4

[9] P. M. Corsi. *Human memory and the medial temporal region of the brain.* PhD thesis, McGill University, 1972. 5

[10] B. M. Ens, R. Finnegan, and P. P. Irani. The Personal Cockpit: A Spatial Interface for Effective Task Switching on Head-Worn Displays. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 3171–3180, 2014. doi: 10.1145/2556288.2557058 5

[11] R. Etemadpour, E. Monson, and L. Linsen. The Effect of Stereoscopic Immersive Environments on Projection-Based Multi-dimensional Data Visualization. In *International Conference on Information Visualisation*, pp. 389–397, 2013. doi: 10.1109/IV.2013.51 1, 2

[12] A. S. Fernandes and S. K. Feiner. Combating VR Sickness Through Subtle Dynamic Field-of-View Modification. In *IEEE Symposium on 3D User Interfaces*, pp. 201–210, 2016. doi: 10.1109/3DUI.2016.7460053 5

[13] A. Field. *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications, 4^{th} ed., 2013. 6

[14] G. W. Fitzmaurice. Situated Information Spaces and Spatially Aware Palmtop Computers. *Commun. ACM*, 36(7):39–49, 11 pages, 1993. doi: 10.1145/159544.159566 2

[15] T. Friedrich, A. Prouzeau, and M. McGuffin. The Effect of Increased Body Motion in Virtual Reality on a Placement-Retrieval Task. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, 2021. doi: 10.1145/3489849.3489888 3, 4, 5, 9

[16] D. Gagnon. Videogames and Spatial Skills: An Exploratory Study. *Educational Technology Research and Development*, 33(4):263–275, 1985. doi: 10.1007/BF02769363 1, 2

[17] B. Gao, H. Kim, B. Kim, and J.-I. Kim. Artificial Landmarks to Facilitate Spatial Learning and Recalling for Curved Visual Wall Layout in Virtual Reality. In *IEEE International Conference on Big Data and Smart Computing*, pp. 475–482, 2018. doi: 10.1109/BigComp.2018.00076 2, 3, 4, 6, 9

[18] D. Han and I. Cho. Evaluating 3D User Interaction Techniques on Spatial Working Memory for 3D Scatter Plot Exploration in Immersive Analytics. In *IEEE International Symposium on Mixed and Augmented Reality*, pp. 513–522. IEEE Computer Society, 2023. doi: 10.1109/ISMAR59233.2023.00066 2, 3, 4, 5

[19] L. Harrison, F. Yang, S. Franconeri, and R. Chang. Ranking Visualizations of Correlation Using Weber's Law. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1943–1952, 2014. doi: 10.1109/TVCG.2014.2346979 1

[20] S. G. Hart. NASA-Task Load Index (NASA-TLX); 20 Years Later. *Pro-*

ceedings of the Human Factors and Ergonomic Society annual meeting, 50(9):904–908, 2006. doi: 10.1177/154193120605000909 5

[21] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, vol. 52 of *Advances in Psychology*, pp. 139–183. North-Holland, 1988. doi: 10.1016/S0166-4115(08)62386-9 5

[22] N. Heckert, J. Filliben, C. Croarkin, B. Hembree, W. Guthrie, P. Tobias, and J. Prinz. Handbook 151: NIST/SEMATECH e-Handbook of Statistical Methods, 2002. doi: 10.18434/M32189 1

[23] J.-P. Huttner, Z. Qian, and S. Robra-Bissantz. A Virtual Memory Palace and the User's Awareness of the Method of Loci. In *Proceedings of the European Conference on Information Systems*, 2019. 3, 4, 9

[24] J.-P. Huttner, K. Robbert, and S. Robra-Bissantz. Immersive Ars Memoria: Evaluating the Usefulness of a Virtual Memory Palace. *Proceedings of the Hawaii International Conference on System Sciences*, 2019. 3, 4

[25] J.-P. Huttner and S. Robra-Bissantz. An Immersive Memory Palace: Supporting the Method of Loci with Virtual Reality. In *Americas Conference on Information Systems*, 2017. 3, 4, 9

[26] P. Häfner, C. Vinke, V. Häfner, J. Ovtcharova, and W. Schotte. The Impact of Motion in Virtual Environments on Memorization Performance. In *IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications*, pp. 104–109, 2013. doi: 10.1109/CIVEMSA.2013.6617404 3, 4

[27] Y. Jansen, J. Schjerlund, and K. Hornbæk. Effects of Locomotion and Visual Overview on Spatial Memory when Interacting with Wall Displays. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 12 pages, pp. 1–12, 2019. doi: 10.1145/3290605.3300521 5

[28] T. Jund, A. Capobianco, and F. Larue. Impact of Frame of Reference on Memorization in Virtual Environments. In *IEEE International Conference on Advanced Learning Technologies*, pp. 533–537, 2016. doi: 10.1109/ICALT.2016.77 2, 3, 4, 9

[29] R. P. C. Kessels, M. J. E. van Zandvoort, A. Postma, L. Jaap Kappelle, and E. H. F. de Haan. The Corsi Block-Tapping Task: Standardization and Normative Data. *Applied Neuropsychology*, 7(4):252–258, 2000. doi: 10.1207/S15324826AN0704_8 5

[30] M. Kraus, N. Weiler, D. Oelke, J. Kehrer, D. A. Keim, and J. Fuchs. The Impact of Immersion on Cluster Identification Tasks. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):525–535, 2020. doi: 10.1109/TVCG.2019.2934395 1, 2

[31] E. Krokos, C. Plaisant, and A. Varshney. Virtual memory palaces: immersion aids recall. *Virtual Reality*, 23(1):1–15, 2019. 2, 3, 4, 6, 9

[32] W. S. Lages and D. A. Bowman. Move the Object or Move Myself? Walking vs. Manipulation for the Examination of 3D Scientific Data. *Frontiers in Information and Communication Technology*, 5:15, 2018. doi: 10.3389/fict.2018.00015 2

[33] E. L. Legge, C. R. Madan, E. T. Ng, and J. B. Caplan. Building a memory palace in minutes: Equivalent memory performance using virtual versus conventional environments with the Method of Loci. *Acta Psychologica*, 141(3):380–390, 2012. doi: 10.1016/j.actpsy.2012.09.002 2, 4, 9

[34] B. Leitheiser and D. Munro. An Experimental Study of the Relationship Between Spatial Ability and the Learning of a Graphical User Interface. In *Proceedings of the Americas Conference on Information Systems*, 1995. 1, 2

[35] L. Lisle, X. Chen, J. Edward Gitre, C. North, and D. A. Bowman. Evaluating the Benefits of the Immersive Space to Think. In *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops*, pp. 331–337, 2020. doi: 10.1109/VRW50115.2020.00073 2

[36] L. Lisle, K. Davidson, E. J. Gitre, C. North, and D. A. Bowman. Sensemaking Strategies with Immersive Space to Think. In *IEEE Virtual Reality and 3D User Interfaces*, pp. 529–537, 2021. doi: 10.1109/VR50410.2021.00077 2

[37] J. Liu, B. Ens, A. Prouzeau, J. Smiley, I. K. Nixon, S. Goodwin, and T. Dwyer. DataDancing: An Exploration of the Design Space For Visualisation View Management for 3D Surfaces and Spaces. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 17 pages, 2023. doi: 10.1145/3544548.3580827 3, 5

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

[39] J. Liu, A. Prouzeau, B. Ens, and T. Dwyer. Effects of Display Layout on Spatial Memory for Immersive Environments. *Proceedings of the ACM on Human-Computer Interaction*, 6:468–488, 2022. 1, 2, 3, 4, 9

[40] J. Liu, K. A. Satriadi, B. Ens, and T. Dwyer. Investigating the Effects of Physical Landmarks on Spatial Memory for Information Visualisation in Augmented Reality. In *IEEE International Symposium on Mixed and Augmented Reality*, pp. 289–298, 2024. doi: 10.1109/ISMAR62088.2024.00043 2

[41] J. P. McIntire and K. K. Liggett. The (Possible) Utility of Stereoscopic 3D Displays for Information Visualization: The Good, the Bad, and the Ugly. In *IEEE VIS International Workshop on 3DVis*, pp. 1–9, 2014. doi: 10.1109/3DVis.2014.7160093 2

[42] O. Merhi, E. Faugloire, M. Flanagan, and T. A. Stoffregen. Motion Sickness, Console Video Games, and Head-Mounted Displays. *Human Factors*, 49(5):920–934, 2007. doi: 10.1518/001872007X230262 4

[43] B. Moll and E. Sykes. Optimized virtual reality-based Method of Loci memorization techniques through increased immersion and effective memory palace designs: A feasibility study. *Virtual Reality*, 27(2):941–966, 2023. doi: 10.1007/s10055-022-00700-z 3, 4

[44] P. Quinn, A. Cockburn, Indratmo, and C. Gutwin. An Investigation of Dynamic Landmarking Functions. In *Proceedings of the ACM Working Conference on Advanced Visual Interfaces*, pp. 322–325, 2008. doi: 10.1145/1385569.1385623 1, 2, 4, 6

[45] E. D. Ragan, D. A. Bowman, and K. J. Huber. Supporting cognitive processing with spatial information presentations in virtual environments. *Virtual Reality*, 16:301–314, 2012. doi: 10.1007/s10055-012-0211-8 2, 3, 4, 9

[46] L. Rebenitsch and C. Owen. Individual Variation in Susceptibility to Cybersickness. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pp. 309–317, 2014. doi: 10.1145/2642918.2647394 5

[47] G. E. Riccio and T. A. Stoffregen. An Ecological Theory of Motion Sickness and Postural Instability. *Ecological Psychology*, 3(3):195–240, 1991. doi: 10.1207/s15326969eco0303_2 4

[48] K.-F. Richter and S. Winter. Landmarks. *Springer International Publishing*, 2014. 2

[49] T. Ritchey. Fritz Zwicky, Morphologie and Policy Analysis. In *16th EURO conference on operational analysis, Brussels*, 1998. 2

[50] G. Robertson, M. Czerwinski, K. Larson, D. C. Robbins, D. Thiel, and M. Van Dantzich. Data Mountain: Using Spatial Memory for Document Management. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pp. 153–162, 1998. 5

[51] A. Sarikaya and M. Gleicher. Scatterplots: Tasks, Data, and Designs. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):402–412, 2018. doi: 10.1109/TVCG.2017.2744184 1, 5

[52] 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, 2020. doi: 10.1145/3427329 5

[53] J. Scarr, A. Cockburn, and C. Gutwin. Supporting and Exploiting Spatial Memory in User Interfaces. *Foundations and Trends in Human–Computer Interaction*, 6(1):1–84, 2013. doi: 10.1561/1100000046 2, 5, 6

[54] G. Stoet. PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments. *Teaching of Psychology*, 44(1):24–31, 2017. doi: 10.1177/0098628316677643 5

[55] M. S. Uddin, C. Gutwin, and A. Cockburn. The Effects of Artificial Landmarks on Learning and Performance in Spatial-Memory Interfaces. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 13 pages, pp. 3843–3855, 2017. doi: 10.1145/3025453.3025497 1, 2

[56] K. J. Vicente, B. C. Hayes, and R. C. Williges. Assaying and Isolating Individual Differences in Searching a Hierarchical File System. *Human Factors*, 29(3):349–359, 1987. doi: 10.1177/001872088702900308 1, 2

[57] J. Vindenes, A. Ortiz de Gortari, and B. Wasson. Mnemosyne: Adapting the Method of Loci to Immersive Virtual Reality. In *Proceedings of Augmented Reality, Virtual Reality, and Computer Graphics*, pp. 205–213, 2018. doi: 10.1007/978-3-319-95270-3_16 3, 4

[58] N. G. Vinson. Design Guidelines for Landmarks to Support Navigation in Virtual Environments. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 278–285, 1999. doi: 10.1145/302979.303062 1, 2

[59] J. A. Wagner Filho, C. Freitas, and L. Nedel. VirtualDesk: A Comfortable and Efficient Immersive Information Visualization Approach. *Computer Graphics Forum*, 37(3):415–426, 2018. doi: 10.1111/cgf.13430 2, 5

[60] J. A. Wagner Filho, M. F. Rey, C. M. D. S. Freitas, and L. Nedel. Immersive Visualization of Abstract Information: An Evaluation on Dimensionally-

Reduced Data Scatterplots. In *IEEE Conference on Virtual Reality and 3D User Interfaces*, pp. 483–490, 2018. doi: 10.1109/VR.2018.8447558 1, 2

[61] F. Yang, J. Qian, J. Novotny, D. Badre, C. D. Jackson, and D. H. Laidlaw. A Virtual Reality Memory Palace Variant Aids Knowledge Retrieval from Scholarly Articles. *IEEE Transactions on Visualization and Computer Graphics*, 27(12):4359–4373, 2021. doi: 10.1109/TVCG.2020.3009003 2, 3, 4, 6, 9

[62] F. A. Yates. *Art of Memory*. Routledge, 2013. doi: 10.4324/9781315010960 2