A Virtual Reality Memory Palace Variant Aids Knowledge Retrieval from Scholarly Articles

Fumeng Yang, Jing Qian, Johannes Novotny, David Badre, Cullen D. Jackson, and David H. Laidlaw

Abstract—We present exploratory research of virtual reality techniques and mnemonic devices to assist in retrieving knowledge from scholarly articles. We used abstracts of scientific publications to represent knowledge in scholarly articles; participants were asked to read, remember, and retrieve knowledge from a set of abstracts. We conducted an experiment to compare participants' recall and recognition performance in three different conditions: a control condition without a pre-specified strategy to test baseline individual memory ability, a condition using an image-based variant of a mnemonic called a "memory palace," and a condition using a virtual reality-based variant of a memory palace. Our analyses show that using a virtual reality-based memory palace variant greatly increased the amount of knowledge retrieved and retained over the baseline, and it shows a moderate improvement over the other image-based memory palace variant. Anecdotal feedback from participants suggested that personalizing a memory palace variant would be appreciated. Our results support the value of virtual reality for some high-level cognitive tasks and help improve future applications of virtual reality and visualization.

Index Terms—Virtual Reality, Mnemonic Devices, Natural Language Documents, Human Memory, Spatialization, Spatial Memory

1 Introduction

Our memory is imperfect. We easily forget the names of people we meet and the content of papers we read [1]. In complicated research activities involving large amounts of information, it is difficult to remember analytic stages, find valuable information, or manage computer-based documents effectively [2].

Modern technology can help our memory via *spatialization*, setting non-spatial information in a landscape of some sort and hence invoking spatial memory—often quite good—to compensate for other, fickle types of memory [3], [4]. Spatialization has been used in various domains to address problems like memorability [5], sense-making [6], [7], cluttering [8], and layout [9].

An organized way of using spatialization to aid memory is with a *memory palace* (or *method of loci*). A memory palace builds connections between information and the loci in the mind (see Fig. 1) [10], [11], [12], [13]. This mnemonic device is superior to many other methods (e.g., peg, link) [14], especially for serial recall [15], [16]. It is commonly used to memorize a list of items (e.g., words [17], [18], [19], [20], names [21], faces [22], [23], and graphical marks [24]).

In this paper, we were inspired by the memory palace method to explore the value of a memory palace in realistic tasks such as retrieving semantic knowledge from scholarly articles. The previously cited studies showed that a memory palace leads to promising improvements in human memory. Yet remembering a list of items is relevantly elementary; it does not provide good insights for knowledge workers, who usually face much longer and more complicated documents. A few studies proposed to use a memory palace for improving students' learning performance [25], [26], [27] and second language learning [28], [29]. They did not provide more specific experimental results for intricate scenarios such as retrieving knowledge from scholarly articles. Recall from sentences and paragraphs is theoretically and physiologically different from recall of a word list in many aspects; it requires a higher level of long-term knowledge [16], involves syntactic and semantic processing, and activates different areas of the brain [30]. Compared to recall a word list, application of a memory palace to scholarly articles is not trivial; its performance should be tested explicitly.

Recent work attempted to address the fact that a memory palace is difficult to build and use. Building a memory palace often requires a set of personally intimate loci, necessitates hours of training [19], and demands significant cognitive load and attention [20], [31]. It would be hard to apply a memory palace to remembering scientific knowledge in scholarly articles. A memory palace variant can mitigate these difficulties and address the efforts of utilizing personally intimate loci. For example, a memory palace variant could use a list of fictional loci (e.g., locations in a story [18], [32]), or a 3D virtual scene on a desktop [19], [33] or in virtual reality [22], [23], [24], [34], [35]. All of the cited publications used unfamiliar spatial cues, showing that a variant of a memory palace is comparable to a conventional one. Thus, in our work, we only considered variants of a conventional memory palace which do not require a set of personally intimate loci.

Among all these current techniques, virtual reality may offer the best way to augment a memory palace. Virtual reality is a replication and extension of physical reality and a technique to exploit spatial information. This technique supports cognitive tasks and accesses personal experience

- *Fumeng Yang, Jing Qian, Johannes Novotny, and David H. Laidlaw are with the Department of Computer Science, Brown University, Providence, RI, 02906. E-mails: {fy, jing_qian}@brown.edu and {novotny, dhl}@cs.brown.edu*
- *David Badre is with the Department of Cognitive, Linguistic & Psychological Sciences, Brown University, Providence, RI, 02906. E-mail: david_badre@brown.edu.*
- *Cullen D. Jackson is with Beth Israel Deaconess Medical Center, Harvard University, Boston, MA, 02215. E-mail: cdjackso@bidmic.harvard.edu.*

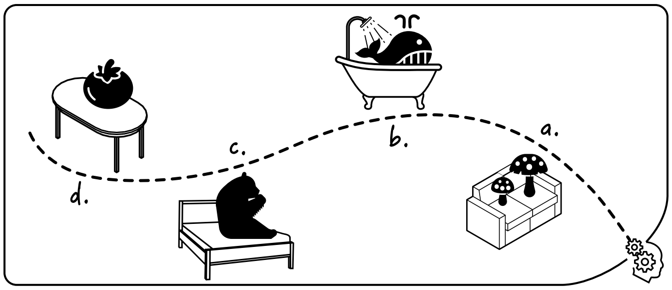


Fig. 1. Building a memory palace has three steps: (1) listing items to remember (e.g.,"mushroom", "whale", "bear", and "tomato"); (2) defining a route with a set of imaginary loci (e.g., couch, bathtub, bed, and table in an apartment); (3) making a connection between a locus and each item, usually via a vivid visualization. To recollect the list, one imagines walking along the route past the loci and picking up the connections. (All icons used in this paper are from The Noun Project [45].)

(e.g., daily life [36], learning [37], [38], [39], education [40], [41], and memory rehabilitation [31]). Virtual reality offers a space for people to move and think. It aids sense-making and builds an externalization of the reasoning process [7], [42]. These properties substantiate that virtual reality might be a suitable environment for using spatialization and mnemonics to aid knowledge retrieval from scholarly articles.

Here we report the results from a study that combined spatialization with virtual reality to help people remember scientific knowledge in scholarly articles. Our experiment focused on conceptual knowledge (e.g., "the interrelationships among the basic elements") and involved cognitive processes of "remembering" and "understanding" [43], [44]. We used abstracts from scientific publications as a representation of scholarly articles and conducted a human-subjects experiment to quantify the effects of a memory palace variant. The total length of the abstracts was about one page of a TVCG article. We report both quantitative and qualitative results and discuss our insights from the experiment. Last, we show how our experiment and results are connected to the literature. Specifically, our research provides three main contributions:

(1) We found that a "memory palace" variant can help retrieve scientific knowledge from scholarly articles;
(2) We demonstrated that virtual reality techniques (i.e., head-mounted displays) provide an effective virtual reality-based variant of a memory palace and improve memory of scientific knowledge;
(3) We showed that virtual reality techniques can support high-level cognitive tasks at least as well as traditional media such as screens.

The experiment and analysis code, materials, data, and videos are available at `http://github.com/Fumeng-Yang/VRMP_for_Knowledge_TVCG`.

2 Study Design

Here we start with an overview of the experiment and design justification. Then we describe the detailed experimental procedure and materials.

2.1 Experimental Design

With the goal of studying the effects of the memory palace method and virtual reality on assisting in knowledge retrieval, we had three experimental conditions:

(1) a control condition (denoted by CONTROL ∅) that tested baseline individual memory ability without any pre-specified strategies.
(2) a mnemonic condition (denoted by IMAGE) in which participants used the spatial cues from a picture and a story to build a memory palace variant. This story-based procedure was used in the literature as an variant of a conventional memory palace [18], [32] to address the issue that a memory palace is often hard to learn and build. To accurately record participants' data, this condition was conducted on a computer using a monitor to present the picture and the story.
(3) a mnemonic condition (denoted by VR) in which participants used spatial cues in virtual reality to build a variant of memory palace. This condition was to measure the effectiveness of virtual reality techniques for knowledge retrieval.

We would like to note again that the two mnemonic conditions are based on the literature of using a memory palace variant, where participants effectively and efficiently used a set of external spatial cues to aid recall; this approach is sometimes distinguished from a conventional "memory palace," which requires hours of training and is built on personally familiar loci. We used the term "memory palace" to follow the literature of building a memory palace variant and for simplicity, but our approach only utilized a variant of a conventional memory palace.

Our experiment was a mixed design. Each participant first took part in the control condition. After 72-96 hours (3-4 days), the participant returned and was randomly assigned to one of the two mnemonic conditions (i.e., CONTROL ∅ → IMAGE or CONTROL ∅ → VR). Thus, "condition" refers to each visit (CONTROL ∅, IMAGE, or VR), and "group" refers to the participants who committed to two visits.

We used abstracts (i.e., passages) from scientific publications to represent scientific knowledge in scholarly articles. Each participant saw all the passages (12) in randomized order and viewed different passages in the two visits. They were asked to read the passages and remember the main ideas (i.e., the gist), but not memorize the passages word for word. This process emulated making sense of scientific concepts and remembering knowledge from scholarly articles. Here we contrast our task with free reading, an approach that many people use with reading abstracts.

To measure participants' memory rate, we used both recall and recognition tasks. The recall and recognition procedures are used commonly in learning and cognition studies (e.g., [46]). A recall procedure involves actively searching for a piece of information; a recall task is a reconstruction of items to be remembered and should not be considered a "hit rate" as in other fields like pattern recognition. A recognition procedure involves identifying previously learned information [47], and a recognition task is a discrimination between items to be remembered and other similar items [48]. While recall and recognition are related [46], recognition is considered easier [49]. We asked participants to recall the passages they read and recognize a set of sentences (10) from the passages. We measured participants' memory rate based on the recalled passages

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3009003, IEEE Transactions on Visualization and Computer Graphics

Fig. 2. In VR, we had (a) a participant walking around to see the spatial cues in virtual reality, read and remember the given set of passages; (b) one of the loci in the scene; and (c) a sample of text rendered in our virtual reality system while the participant is casting a ray ("laser pointer"). All three images here are first-person view screenshots, cropped to fit the manuscript. In particular, image (a) is a screenshot when viewed from a distance, edited to create a third-person view screenshot for illustrating the condition. (The 3D models used in this paper are all under the Royalty Free License.)

and their answers to the recognition questions.

2.2 Justification of Design Decisions

The first design decision we justify was the use of three experimental conditions: a baseline (CONTROL ∅), an image-based mnemonic condition (IMAGE), and a virtual reality-based mnemonic condition (VR); both mnemonic conditions were inspired by the memory palace method. The reasons are as follows.

- CONTROL ∅ measured an individual's memory ability for the known large variance in learning, reading, and individuals' ability to remember [50], [51], [52]. This condition provided a baseline for observing memory improvements, compared to the other two conditions.
- IMAGE intended to estimate the effects of using an image-based memory palace variant for remembering scientific knowledge in scholarly articles. IMAGE followed the literature of building a memory palace variant [18], [32] and weakened the imagining process and mental activities via providing participants a picture and a story (see Fig. 3a), while a conventional memory palace usually utilizes familiar locations. The picture and the story here invoke participants' imagination about a coffee shop; participants did not have to use the exact visual information offered in the picture. We explicitly instructed participants to take time and imagine themselves walking through the scene in the picture at their own pace until they felt confident that they knew the route and the given loci. In order to urge participants to focus on mental activities and reduce the interference from interactions, we minimized the potential interactions with a computer and only used a static picture.
- VR tested the effectiveness of using external and immersive spatial cues to create a memory palace variant, reduce mental demand, and improve memory performance. We used the same underlying loci from IMAGE to maximize comparability between the two conditions; we also selected the loci that would likely be familiar to the potential participants (see Section 2.3 below for more explanation about comparability).

A second design decision was to start with CONTROL ∅ and follow up with IMAGE or VR. Because learning effects might influence the condition in the second visit, the baseline memory rate must be tested before a participant has learned any specific strategies. Once participants have learned a strategy, they cannot "unlearn" it. To mitigate the impact of the learning effects, we required an interval of 3-4 days between the two visits. CONTROL ∅ was used as a reference point for both mnemonic conditions because the learning effects should be similar.

A third design decision was to assign one mnemonic condition to each participant, instead of using a complete within-subjects design. This was meant to address both the learning effects in participants' familiarity with the memory palace method and the recruiting difficulties. If we had used a within-subjects design, participants would experience both mnemonic conditions. They would learn spatial cues in the first mnemonic condition and bring these into the later one. Additionally, a complete within-subjects design would create recruiting difficulties because participants would have to commit to additional visits at similar intervals.

Finally, although augmented reality can support a memory palace as well [20], [53], [54], we used virtual reality because it provides a unique enclosed and unchangeable environment. Augmented reality relies on the real world to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3009003, IEEE Transactions on Visualization and Computer Graphics

provide spatial cues, and changes in the real world may interfere with the memory process. An enclosed virtual reality environment also helps reduce external interference and control variance in the experiment.

2.3 Improving Comparability

The primary means to improve comparability between the two mnemonic conditions is that we normalized participants' "memory palaces." We asked all participants to use the same set of spatial cues: IMAGE used loci from a picture rendered from the 3D model and ordered them as a short story (see Fig. 3a), and VR used loci from the same 3D model (see Fig. 2). Building a memory palace relies on individual and internal processes [19]. Individuals may select very different loci or find it difficult to come up with a set of loci in the experiment. Conversely, the literature suggests that a set of fictional or artificial loci can be used in the memory palace method; familiar and personal spatial cues are not always necessary (e.g., [17], [19], [21], [24], [32], [55], [56]). Therefore, it is possible to use the loci in a picture, a story [32], or an unfamiliar scene [19], [22] to facilitate the remembering process, allowing for comparison between our two mnemonic conditions.

We further made the two mnemonic conditions more comparable in several ways.

- **Assuring familiarity with the loci** We designed the experiment so that participants would have a moderate level of familiarity with the loci: we gave participants a generic coffee shop as their "palace" since the potential participants (college students) were very likely to be familiar with such a coffee shop, and there were three coffee shops with a similar interior within one mile of the experiment location.
- **Using the same rendering process** The textures of the passage objects in VR were screenshots of the text on the screen in IMAGE to eliminate formatting factors [57] (Fig. 2c vs. 3b).
- **Aligning the interaction fidelity** Moving and grabbing objects other than the passages and instructions was not allowed in virtual reality so that we could control the loci used. We had each participant read one passage at a time on screen by blurring all passages except the one under the cursor [58] (Fig. 3b). This is because, in the virtual reality system, participants only saw one passage at a time due to the limited resolution and field-of-view (Figs. 2bc). We implemented a laser pointer in virtual reality (Fig. 2c), where participants could use a controller to cast a ray to help them follow the text, much like a cursor on screen.

To further improve comparability, we balanced participants' gender between the two mnemonic conditions and included only graduate and undergraduate students with native or bilingual proficiency in English. These criteria constrained age, reading experience, and familiarity with the coffee shop scene.

2.4 Hierarchical Memory Palace

At a conceptual level, the memory palace method was originally developed to remember a list of items (e.g., names or words), not a series of passages and the descriptive information contained. We introduce the concept of a *hierarchical memory palace* as a richer way to remember knowledge that is more complex than word lists: each passage is associated with a locus, and each of the main ideas in a passage is associated with the spatial information near the locus. This is similar to building a spatial *concept map* [59], and this hierarchical procedure also aligns with the human brain's language processing [30]. Observations from our pilot study support this speculation about a hierarchical memory palace, as two out of four participants who tried a mnemonic condition claimed that they used a similar strategy. We urged participants to use a hierarchical memory palace, but they could use the memory palace method to read and remember the passages in any way they wanted, as long as their methods were intended to mentally visualize the information in the passages and associate it with the loci. To record the participants' methods, we asked participants to report in the post-hoc questions.

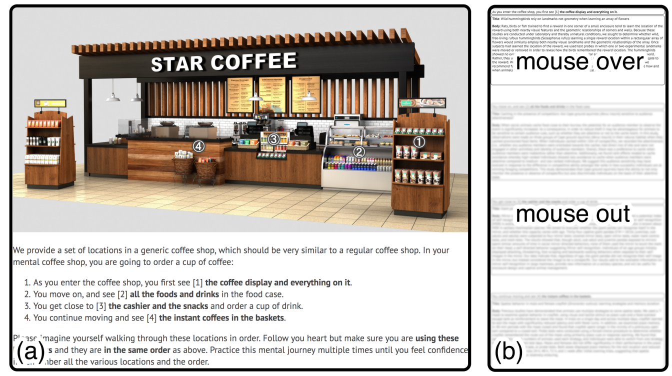


Fig. 3. In IMAGE: (a) participants first familiarized themselves with the given loci based on a picture and a story (not drawn to the original scale); (b) then the picture and the story were removed, and participants read and were asked to remember a series of on-screen passages by imagining all the loci. To urge participants to focus on one passage each time, a passage was readable only when their mouse was covering the passage.

2.5 Experimental Procedure

To recap, we used three experimental conditions (CONTROL ∅, IMAGE, and VR) and both recall and recognition tasks to measure participants' memory performance. Here we present the experimental procedure, describing CONTROL ∅ first followed by the corresponding modifications in IMAGE and VR.

Each of the three conditions consisted of five sessions: training, practice, main, recall, and recognition. Prior to the first session, participants started with an overview of the condition content; following the last session, participants answered the post-hoc questions. For the full experiment details, please refer to our supplementary materials and the videos.

Training Session

- In CONTROL ∅, participants were shown a sample passage, a recall guideline, and an example of a recalled passage. The instructions emphasized that participants should focus on remembering the main ideas in the passages.
- In IMAGE and VR, in addition to the training for CONTROL ∅, each participant learned about the memory palace method by reading an article and taking a follow-

up quiz about building a memory palace. They had to answer all three quiz questions correctly before they could proceed to the practice session. The instructions in this condition and the quiz were designed to encourage participants to build a hierarchical memory palace, as introduced in Section 2.4.

Practice Session

- In CONTROL ∅, participants were asked to read one passage, recall the passage, and finish a practice recognition task with three questions. Feedback was given at the end of the session.
- In IMAGE, participants first familiarized themselves with the given loci based on the rendered picture of an office room with a desk and a chair (see the supplementary materials and the videos). They then read the practice passage using the memory palace method.
  In VR, participants were first trained to use the virtual reality system. Participants walked around in the practice scene (the same office room), scaled and moved the passage objects, and checked the experimental instructions shown in the virtual reality system. After this training, they navigated through the same scene and read the same practice passage using the memory palace method. The experimenter introduced the virtual reality system without describing the memory palace method to ensure that the two groups of participants received the same amount of training.

Main Session

- In CONTROL ∅, participants read and remembered four passages together. They had up to 30 minutes to read the passages. After this time, all the passages were blurred, and participants were not able to read them again (see Fig. 3b).
- In both IMAGE and VR, participants were instructed to use the memory palace method. In IMAGE, participants first familiarized themselves with the given loci using the rendered picture of the coffee shop (see Fig. 3a). They were instructed to imagine themselves walking inside the picture and remember the given series of loci until they felt confident that they remembered the route and the loci; there was no time constraint for familiarizing themselves with the loci. Participants then read four passages and associated each passage to a given locus (see Fig. 3b) using the memory palace method. They had up to 30 minutes to read the passages; after this time, all the passages were blurred.
  In VR, participants first familiarized themselves with the given loci in the same coffee shop scene (Fig. 2 without any passage); again, there was no time constraint for familiarizing themselves with the loci. After participants felt familiar with the scene and confident that they remembered the route and the loci, the experimenter made the passages visible to them (Fig. 2). They could walk to access passages in order, move and scale passages, and check the instructions in the virtual reality system. Participants read and remembered four passages using the memory palace method with a 30 minute time constraint. Because participants could not see a clock in the virtual reality system, they were free to ask the experimenter how much time was left, and the experimenter reminded each participant when 5 minutes were left. After 30 minutes, participants were asked to stop reading and take off the headset if they had not finished the task early. Participants then came back to sit in front of the computer and continued performing the experiment.

Recall Session

- In CONTROL ∅, the recall task was performed on a desktop computer immediately after the main session. Participants were instructed to recall and record the passages as separate entries in the order that the passages were read. Each participant had up to 30 minutes for this session.
- In both IMAGE and VR, participants were instructed to recall by imagining the coffee shop again, walking through the loci, and picking up the connections they built between the loci and the passages. They had up to 30 minutes. After this time, they were not able to input any further recall text.

Recognition Session

The recognition session was performed on the desktop; it was the same for all the conditions. Participants saw ten sentences and answered "Yes" or "No" to indicate whether they had read each sentence in the exact wording in the reading session; they rated their confidence in each answer on a 7-point Likert scale.

2.6 Experimental Materials

Apparatus We used an HTC Vive (2017 model) [60] and set it up in a 21′5″ (6.5m) by 8′11″ (2.7m) area. We chose an HMD over a CAVE-style virtual reality environment [61] for the flexibility of experimental setup. Our experimental surroundings were similar to a regular reading environment (e.g.,library, coffee shop, etc.) which might not be quiet at all times; participants were free to use the earplugs provided. All participants received the same HMD, and the locations of the two base stations remained the same throughout the experiment.

Palace and Loci To invoke participants' familiarity with a coffee shop, we used life-size 3D models and a nearly photorealistic rendering quality (only in the main session). We used global illumination, normal maps, reflection, refraction, and ambient occlusion; all were pre-rendered (i.e., baked) into different texture channels using the photorealistic rendering engine V-Ray [62]. We selected the loci in the palace based on three criteria: (1) the constraints of the available physical space and devices; (2) the distances between the loci; and (3) the visual appeal of the loci. We chose the loci from the customer area. The loci were consistent between IMAGE and VR with one necessary modification due to the physical space constraint: the front of the coffee display shelf (the first locus) was used in IMAGE (Fig. 3a), and the back was used in VR (Fig. 2a). The 3D objects were the same when viewed from the back and front.

Passages We used abstracts from the research field "animal cognition," since this topic is likely to be unfamiliar but accessible to a general audience. We gathered over 400 abstracts from articles published in *Animal Cognition* (Springer) between 2013 and 2018. This set was quartered by considering the length of each passage ($\mu = 213.83$,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3009003, IEEE Transactions on Visualization and Computer Graphics

$\sigma = 10.46$ words) and the readability of the title. The resulting abstracts were read by one author, filtered based on readability, and confirmed by a second author. The final set contained 12 abstracts of similar length, each describing a different animal species. These passages had an average Flesch-Kincaid grade level [63] of 13.56 ($\sigma = 2.07$, from [64]), meaning that they could be read by an average college student. To illustrate the Flesch-Kincaid grade level, this paragraph "Passages" has a score of about 11 (slightly easier). Each participant saw all 12 passages split randomly and equally over the two visits; four were targets and two were distractors. These twelve abstracts are provided in Appendix C.

Sentences Ten sentences used in the recognition task were created based on the passages. Four sentences were taken directly from the four passages shown in the reading session. Two were taken from two other passages that participants had not seen. The remaining four sentences were distractors; they were revisions of the sentences in the passages from the reading session created by (1) reversing a conclusion or result (e.g., "helpful" to "not helpful") or (2) changing the numbers in a passage (e.g., "10 cats" to "26 cats"). None of the sentences reversed any obvious facts (e.g., "cats catch mice") nor revised only wording (e.g.,"helpful" vs. "good"). All the sentences created for the recognition task are provided in Appendix C.

Participants Twenty-six participants (16 female and 10 male) were recruited from the campus and nearby institutions and received \$10 per hour as compensation. Participants had to be at least 18 years old in order to take part in the study. They were graduate or undergraduate students with native or bilingual proficiency in English (age $\mu = 21.92, \sigma = 2.48$); they had 25 different majors including some double or triple majors; computer science was most common (7 participants). All participants claimed to be experienced and comfortable with reading scientific publications. They were randomly assigned to the two experimental groups, and gender was balanced across conditions. The number of participants was decided based on a planned recruiting ending date.

Scoring We adapted a scoring method based on "idea units" to grade the recall passages and quantify the amount of knowledge retrieved [65], [66]. An idea unit is usually a proposition and consists of a predicate [67]. An idea unit from a recall could be correct, wrong, or new (elaboration), and an incomplete idea unit was allowed (0.5) [65]. We made one change in the original scoring method: we considered only relevant ideas and discarded unrelated ones (e.g., "*I forget the name of the fish*"). Each passage contained about 30 idea units, and each idea unit consisted of two to five English words. Two experienced raters parsed and scored the idea units in the original passages. Conflicts were resolved by discussion. The grading was to simply check if an idea appeared in a recall. A single rater compared the recalled and original passages twice, filling in the grading template without knowing the source of recalled passages (i.e.,which experimental condition). In addition, no indication of memorizing the passages (i.e.,word-for-word) was found.

3 Analyses and Results

To recap, we had two goals for this study: (1) to evaluate the use of variants of a memory palace for retrieving scientific knowledge from scholarly articles; and (2) to assess the effects of using virtual reality techniques to facilitate a memory palace variant for this process of retrieving scientific knowledge. Toward these two goals, we designed our experiment and collected data to compare among the control condition (CONTROL ∅), an image-based memory palace variant (IMAGE), and a virtual reality-based memory palace variant (VR). Here we first present the results from our pre-specified analyses followed by two post-hoc exploratory analyses to compensate for some of the unexpected results. Last, we discuss our insights about the results.

3.1 Pre-specified Analyses

To guide our analyses, we frame our research questions as follows.

RQ1 How does participants' performance change in any mnemonic conditions for retrieving scientific knowledge from scholarly articles?

RQ2 How does the effectiveness of the two mnemonic conditions (IMAGE and VR) differ from each other?

To answer these two questions, we compared the control condition and two mnemonic conditions by estimating the differences in memory performance; we used multilevel regression analyses to quantify the differences between the two mnemonic conditions. To address the limitations of null hypothesis significance testing, we followed the interval estimate method recommended by Cumming [68] and Dragicevic [69]; this method is also more suitable for our exploratory-type of research [68]. We report the 50% and 95% adjusted bootstrap percentile (BCa) confidence intervals and the effect size (Cohen's d). The bootstrap method does not assume an underlying data distribution and performs well on a small sample size. Note that the interpretation of confidence intervals is *nondichotomous* (i.e.,inconclusive, similar, small, moderate, or large effects). This is different from the interpretation of a significance test (i.e.,significant or not).

3.1.1 Data Preparation

We dropped one participant from the IMAGE group, since she was erroneously assigned to the passages she had already read. We thus had 12 participants from the IMAGE group and 13 participants from the VR group. We discarded a recalled passage if it was obviously incomplete from timeouts and used the remaining three passages. This led to three discarded passages from CONTROL ∅ and one from IMAGE (four different participants) out of a total of 200 recalled passages (25 participants $\times$ 2 visits $\times$ 4 passages). We discarded one of the 50 recognition scores because the participant misunderstood the instructions in CONTROL ∅ and answered "Yes" to all the questions.

We treated the four passages each participant recalled as one set for two reasons: (1) participants started with different passages (e.g., the first one or the last one) so that

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3009003, IEEE Transactions on Visualization and Computer Graphics

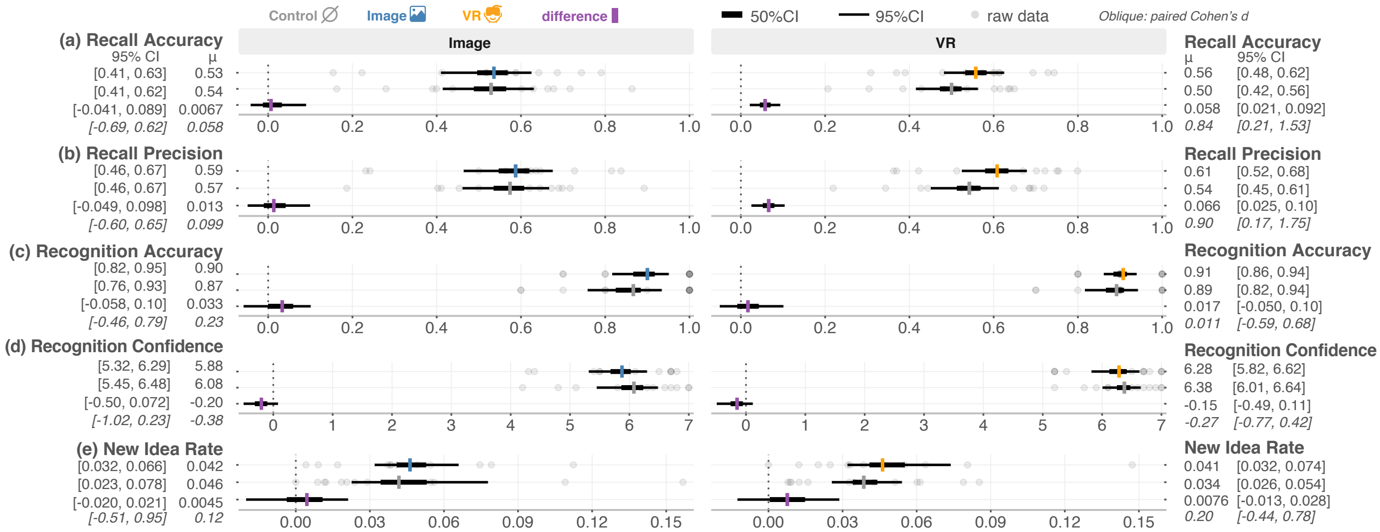


Fig. 4. Comparing the control and two mnemonic conditions: (a) Recall accuracy improved over CONTROL ∅ for participants in VR but remained inconclusive for participants in IMAGE. (b) The results of recall precision are similar to those of recall accuracy. (c) Recognition accuracy improved slightly; however, an effect is nearly missing. (d) Recognition confidence dropped slightly regarding the mean values, but this effect is small. (e) New idea rate increased slightly over the control condition for both mnemonic conditions, but the effect is small.

different levels of serial-position effect may have occurred if passages were treated separately; (2) given that each passage contains different numbers of ideas and the order of passages was randomized, treating them as one set avoids Simpson's paradox, that is, a global trend (e.g., an increase in memory rate) may disappear when data is separated into groups (e.g., passages).

3.1.2 Measures

We adapted the measures used in educational psychology [70], [71] and information retrieval [72] and therefore had five measures defined as follows.

- **Recall accuracy** is the ratio of the number of correct idea units to the number of idea units in the original passage;
- **Recall precision** is the ratio of the number of recalled idea units to the number of idea units in the original passage;
- **Recognition accuracy** is the ratio of the number of correct answers to the number of questions (ten);
- **Recognition confidence** is the average of confidence ratings for all the ten recognition questions.
- **New idea rate** is the ratio of the number of new idea units to the number of idea units in the original passage.

Among these five measures, our primary interest is in **recall accuracy.** Recall precision is highly correlated with recall accuracy both in our data ($r = .97$) and in the literature [71]. Therefore we anticipated that the results of recall accuracy and precision would be similar. In addition, recognition is considered an easy task [49] in which people usually perform very well [71]; we anticipated that it would display few effects. The new idea rate measured whether participants introduced *false memories* (e.g.,"memories of events that took place within experiments but which do not correspond to experimentally presented stimuli" [73]).

To clarify, previous work used the order of recalled items as a measure (e.g., [15], [19], [23], [24], [74]). We found no indication of confusion in the correct order of the four passages, and thus did not consider the order of recalled passages as a measure.

3.1.3 *RQ1: Comparing* CONTROL *with* IMAGE *and* VR

To answer the research question about how participants' performance differs when using a memory palace variant, we compare the results from the control condition to either of the two mnemonic conditions. We report the mean values, 95% BCa confidence intervals of the mean values, differences between the control and two mnemonic conditions, effect size (paired Cohen's d), and the 95% BCa confidence intervals of effect size in Fig. 4. The results are as follows.

- Recall accuracy increased by 0.058 [0.021, 0.092] for VR. The results strongly suggest that recall accuracy improved over CONTROL ∅, and this effect could be large (d: 0.84 [0.21, 1.53]). The effect for IMAGE is very small (d: 0.058 [-0.69, 0.62]).
- Recall precision has similar results to recall accuracy, except that the effect size is slightly larger.
- Recognition accuracy improved slightly in term of mean values for both groups (0.033 [-0.058, 0.10], 0.017 [-0.050, 0.10]). Overall, recognition accuracy remained very similar; the results may not suggest an effect.
- Recognition confidence dropped slightly in terms of mean values for both groups (-0.20 [-0.60, 0.072], -0.15 [-0.49, 0.11]); this effect is small (e.g., d: -0.38 [-1.03, 0.23]).
- New idea rate increased slightly in terms of mean values for both groups (0.0045 [-0.020, 0.021], 0.0076 [-0.013, 0.028]). Overall, this rate remained very similar between the control and either of the two mnemonic conditions.

3.1.4 *RQ2: Comparing* IMAGE *with* VR

To quantify the differences between the two mnemonic conditions, we used mixed-effects models, since the experiment had both within- and between-subjects components. We modeled the experimental conditions as a fixed effect and participants as random intercepts. The fixed effect quantified the differences between conditions, and the random intercepts accounted for the correlation between the observations from the same participant [75]. Using this model, we captured the differences between the experimental conditions (the between-subjects component) and the correlation

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3009003, IEEE Transactions on Visualization and Computer Graphics

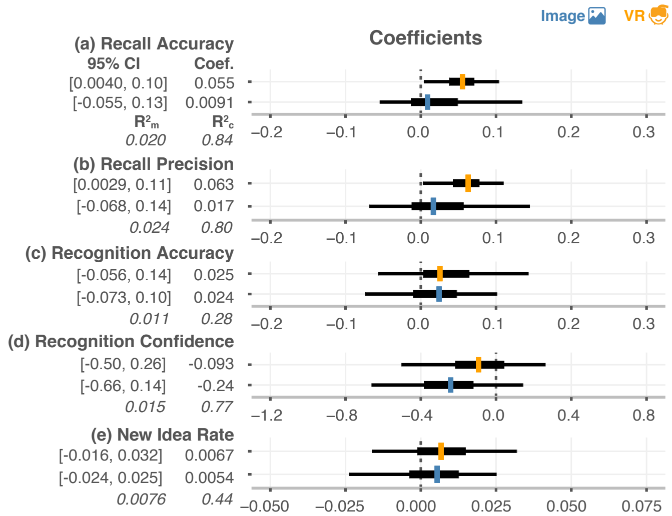


Fig. 5. Comparing the two mnemonic conditions: the five mixed-effects models for each measure, respectively. VR shows a moderate improvement over IMAGE for recall accuracy and precision. The effects of other measures and comparisons remained inconclusive.

within the same participants (the within-subjects component). It allows us to compare the two different groups of participants together.

We therefore built a mixed-effects model for each measure. We report the coefficients, 95% confidence intervals, and R^2 (marginal and conditional) in Fig. 5. The coefficients and confidence intervals represent the differences compared to CONTROL ∅. The results are as follows.

- Recall accuracy improved for VR (0.055 [0.0040, 0.10]) and did not improve substantially for IMAGE (0.0091 [-0.055, 0.13]). These results support that VR moderately improves recall accuracy over IMAGE, and the effect is moderate.
- Recall precision has similar model coefficients and confidence intervals to those of recall accuracy; it improved for VR, although the effect is moderate.
- Both IMAGE and VR have inconclusive results regarding recognition accuracy, recognition confidence, and new idea rate since the confidence intervals largely overlap with zero.

3.1.5 Summary

To summarize, the results show that using virtual reality techniques with the memory palace method as a memory palace variant reliably improved both recall accuracy and precision, compared to the control condition where no pre-specified strategy was given. This effect could be large, but we were not able to estimate its real size from this experiment. A virtual reality-based memory palace variant shows a moderate improvement for recall accuracy and precision over the image-based memory palace variant. However, for the image-based memory palace variant and other measures, the results do not seem to suggest an effect. Recognition confidence dropped slightly for both mnemonic conditions.

3.2 Post-hoc Analyses

Here we present our post-hoc analyses to account for the observed individual differences between the two groups

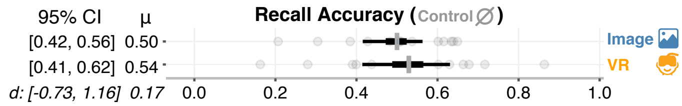


Fig. 6. Recall accuracy for the two groups in CONTROL ∅.

and unexpected results. We first explore the effects of covariates like demographic factors and locomotion, using recall accuracy as the only measure to simplify the analysis. We also report qualitative observations based on anecdotal feedback.

3.2.1 Individual Differences

One observation from the pre-specified analyses is that the performance in CONTROL ∅ seemed different between the two groups: IMAGE seemed to contribute higher and more scattered observations (replotted in Fig. 6). Yet the effect size of this difference is very small (Cohen's d: 0.17 [-0.73, 1.16]).

We then investigated the observed covariates (see Fig.7a), including participants' gender, age, interest in the reading topic [65], verbal ability, visuospatial ability, and text difficulty. We recoded each variable to the scale of [0, 1]. We sampled verbal and visuospatial abilities twice (two visits) by asking participants to self-rate at four levels {*below average*, *average*, *above average*, *very good*}, and mapped them to {0, 0.33, 0.67, 1}. We found that the two sets of samples were consistent (Cohen's weighted κ: 0.84 [0.79, 0.89], 0.59 [0.14, 1.00]). Therefore we used average scores. We measured text difficulty using the Flesch-Kincaid grade level [63]. We also recoded gender to 0 (male) and 1 (female) and rescaled age, interest in the reading topic, and text difficulty to [0, 1].

We built a multiple regression model to quantify the effects of covariates in CONTROL ∅. We focused on explaining variance in data rather than building a predictive model. We used all the covariates above as well as the assigned groups (IMAGE or VR, encoded as 0 and 1, respectively) as the regressors for the models. We checked the collinearity between the variables using the variance inflation factor (all < 2). We report all the coefficients and their 95% confidence intervals in Fig. 7b.

The results show that the model explains some variance in the baseline recall accuracy (multiple $R^2 = 0.39$). The effects of most variables are inconclusive, as the confi-

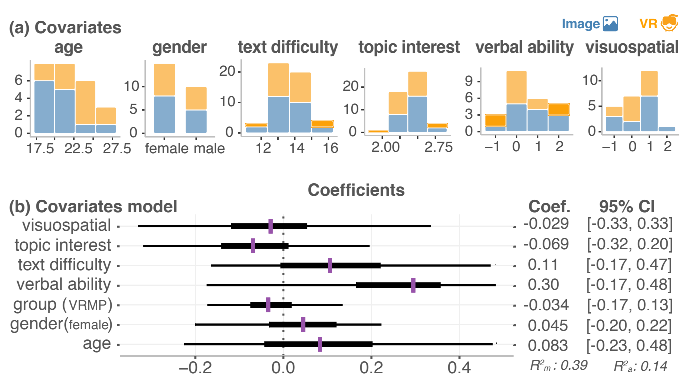


Fig. 7. **(a) Covariates –** age was slightly unbalanced. Otherwise, gender, text difficulty, and participants' interests in the topic were balanced between the two groups. **(b) Explaining variance –** The multiple linear model uses covariates to explain recall accuracy from CONTROL ∅. Overall, verbal ability displayed a small positive effect on recall accuracy; the effects of all the other covariates are inconclusive.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3009003, IEEE Transactions on Visualization and Computer Graphics

dence intervals are large and overlap with 0 (e.g., gender: 0.045 [-0.20, 0.22]). One exception is verbal ability, which displays a very small positive effect on recall accuracy.

In sum, the results show that the two groups performed similarly in CONTROL ∅; henceforth a further comparison between them is fair. We do not suggest generalizing these observations and inferring any effects of individual differences. These observations only support that these two groups of participants are comparable.

3.2.2 Locomotion and Memory Performance

We also investigated participants' movements in VR. Previous studies suggest that participants' distance moved and the view angle to a target during the experiment may reveal insights about their behavior [76], [77]. We were not able to recover the view targets in the experiment. We therefore analyzed participants' movements in VR for each device (i.e., headset, the left and right controllers).

We show the density of each device's position in Fig. 8. We show headset position from the top via its xz-coordinates. Because the left and right controllers were always around the headset, their xz-coordinates were very similar to those of the headset; their y-coordinates (height) seemed to vary more. Therefore, we show xy-coordinates for the two controllers here and provide a figure for the movements on each axis in Appendix D. We also calculated cumulative distance moved and distance moved per minute (speed). To compare these with participants' recall accuracy, we standardized each metric and plotted them in Fig. 8d.

We first noticed that participants generally located themselves near four positions, which were the loci given in the experiment. Then, we found that a few participants (e.g., 2, 3, 8, and 12) had visited more areas in the experiment, but their recall accuracy varied. We may have had a data logging issue with Participant 4, but his or her memory performance seems unaffected.

We then observed that participants generally used their hands at two height levels; we infer that the higher position is where they interacted with the passages, and the lower position is where they put down their hands when not actively using them. Additionally, some participants used one controller more than the other, possibly indicating a dominant hand.

Last, we investigated the correlation between cumulative distance moved, speed, and recall accuracy (Fig. 8d). As suggested above, participants' head movements were correlated with their hand movements; we also found that participants moved their heads less far and more slowly than their hands. We did not observe a strong correlation between distance moved and recall accuracy or between speed and recall accuracy; the correlation coefficients are small (distance: r = .32, .24, .27; speed: r = .11, .068, .10).

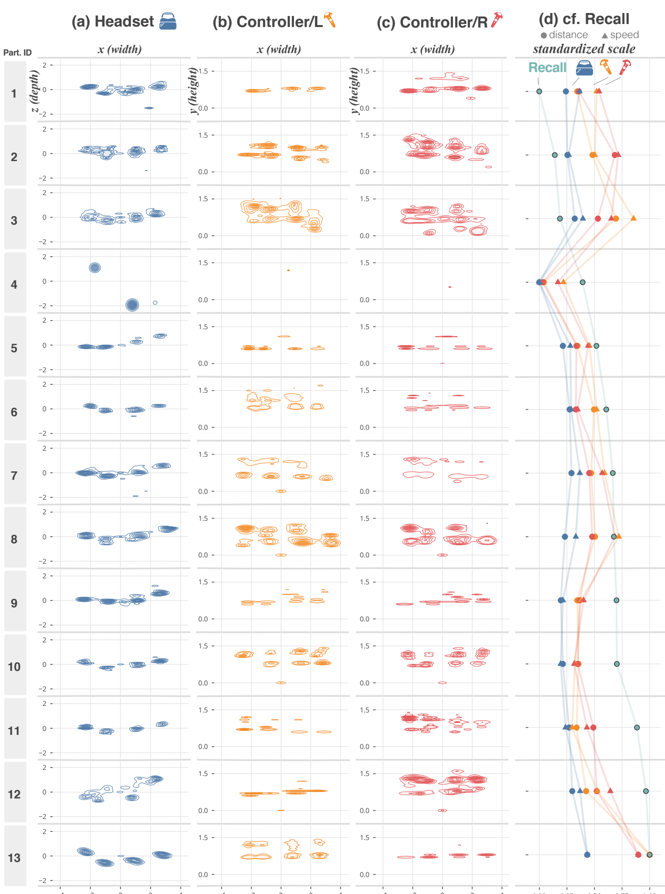


Fig. 8. **Locomotion and Memory Performance** —Here we show the density of participants' position in VR, cumulative distance moved, and recall performance. Since the two controllers are always around the headset, we show their height (y) instead of depth (z). We sorted participants by their recall accuracy and standardized distance moved, speed, as well as recall accuracy to facilitate graphical comparison.

3.2.3 Thematic Analysis

We analyzed participants' post-experiment comments using thematic analysis [78], a widely used qualitative analytic method. We focused on the questions that received more informative comments and discarded the rest; the omitted questions were designed for checking if participants followed instructions or only received a "Yes" or "No" answer from the majority of the participants. In CONTROL ∅, we analyzed the answers to the question about whether participants used any mnemonic methods and their free comments. In both mnemonics conditions (IMAGE and VR), we analyzed the answers to the question about the validity of selected locations and free comments. For each of the three conditions, two authors encoded all the comments and extracted themes independently; the two authors then sat together and merged the themes through discussion. The themes (denoted as "T") in participants' comments are reported as follows.

CONTROL ∅

$\mathbf{T_1}$: Most participants (16/25) used their own strategy in the main session for reading and remembering; some participants answered "no mnemonics" but reported a strategy, and there were at least four different strategies: mental imaging (4/25), a focus on order (8/25), using idea chunks (3/25), and a focus on numbers (2/25).

$\mathbf{T_2}$: The task was interesting (3/25) and difficult (2/25).

IMAGE

$\mathbf{T_1}$: The use of a coffee shop picture (4/12) or the order of loci (3/12) made sense to participants.

$\mathbf{T_2}$: Participants wanted to select a different location in the coffee shop picture (4/12) or use a different scene (6/12).

$\mathbf{T_3}$: Most participants did not think this image-based memory palace variant was helpful or useful (7/12).

VR

$\mathbf{T_1}$: Almost all the participants reported that the use of a coffee shop scene (10/13) was reasonable, and one participant explicitly mentioned that he was familiar with the coffee shop scene.

$\mathbf{T_2}$: Participants would like to use different locations in the scene (2/13) or a different scene (3/13). These three participants specifically mentioned the scenes they would like to use: a nature scene (2/3) or their home (1/3). Among the five participants, two participants agreed that potential participants (college students) were very likely to visit such a coffee shop, but they themselves did not visit a coffee shop very often.

$\mathbf{T_3}$: Participants were impressed by the virtual reality environment (3/13).

$\mathbf{T_4}$: Participants (5/13) thought that using virtual reality techniques had some drawbacks. For example, the virtual reality environment was overwhelming or distracting (2/13).

$\mathbf{T_5}$: Some participants (2/13) reported hardware issues such as the scene being jittery.

3.3 Other Observations

We observed that participants used different strategies and exhibited various behaviors during the experiment. We did not allow taking notes or using marks such as underlining, but some participants used the mouse to select and highlight sentences to help them read. Similarly, in virtual reality, some participants used the laser pointer as a cursor. One participant whispered when reading in both CONTROL ∅ and VR, while the other participants read silently.

To mitigate the learning effects, we designed the experiment so that a mnemonic condition always came after CONTROL ∅. Such a design can be subject to learning effects in the second session (IMAGE or VR), even with a 3-4 day"wash-out" period. We observed a small increased familiarity with the experimental process. For example, some participants got more efficient at using the 30 minute recall time in the later visit. To help compensate for potential unfamiliarity with the experimental process, we dropped incomplete passages and participants' data due to misunderstanding the instructions. The lower performance in IMAGE may also suggest that learning effects did not play an important role. None of the participants had used the memory palace method to remember scientific knowledge or articles before the study. Two participants from IMAGE and three from VR reported that they had heard of this method or tried it for remembering a list of words.

4 Discussion

In this section, we present our understanding and insights about the experimental results, followed by the limitations of this work. We also show how our experiment and results are connected to other studies combining memory palace variants with virtual reality techniques.

4.1 Experimental Setups

Improvements in performance may have been subject to a kind of ceiling effect. In cases where users had high baseline performance, there was little room for them to improve. Over half of the participants used their own strategies in the control condition, and some of the strategies were a mnemonic device similar to a memory palace (e.g., mental images). They might have unconsciously declined to use the memory palace method or assumed it is difficult to use. Therefore, the improvements of both mnemonic conditions might have been larger if all the participants had not used any strategies in the control condition.

In IMAGE, we observed clear themes in participants' comments, where they wanted to use a different scene, and they did not think such a mnemonic is helpful. This can be explained by the unexpected difficulty of concurrently thinking of spatial cues, reading the passages, and making connections between passages and loci based on a previously seen picture. Participants may not have been engaged in the experiment. Another counter-intuitive fact is that immediate serial recall is easier for sentences than word lists because of the additional support that meaningful material receives from long-term memory [16], which may also explain some of the ineffectiveness. We cannot conclude that an image-based memory palace variant is detrimental to knowledge retrieval from scholarly articles, but, at least in our setup of using a picture to invoke the spatial cues and for such a difficult task, we did not observe a strong positive effect.

In VR, we observed various themes in participants' comments; half of the participants found it useful, while the other half experienced difficulties with it. Some participants unfamiliar with a coffee shop may have used the spatial cues that they had just learned in the virtual reality scene, while others might not. The various themes may indicate more variance in this condition, which may have weakened the observed effect size.

4.2 Mnemonic Conditions

Between the two mnemonic conditions, we "normalized" participants' "memory palaces" and urged them to use the same set of loci; we also gave participants an unbounded time to get familiar with the tasks, loci, and scene. Although both mnemonic conditions generally improved participants' memory performance (IMAGE was merely in the positive direction given the inconclusive results), the results of this normalization may be subject to individual differences. Some participants were familiar with the coffee shop scene, and therefore they may have found the memory palace method worked well for them in the experiment. However, other participants were not familiar with such a coffee shop. They had time to explore the scene in the experiment, but they might not have been able to utilize spatial cues effectively because they had to finish an extra task of remembering the given spatial cues. This normalization had the benefits of reducing the variance between conditions, improving comparability between conditions, and increasing statistical power. However, it may have weakened the improvement of memory rate, compared to an ideal case of using each participant's favorite scene or a set of loci that they have personal attachments to; that is, we could use a procedure more similar to a conventional memory palace.

There were also a number of differences between these two mnemonic conditions that may inspire future work. Each of our conditions vary in resolution, head tracking for navigation, and the level of interaction fidelity. While these differences could be varied independently in a different experiment, our design combined specific levels of these factors to create conditions that combine what we believe are the virtual reality characteristics that are most valuable for a memory mnemonic. Some of these differences are constrained by technology. For example, we cannot easily have a higher-resolution head-mounted virtual reality display. It is an open question which combination of factors causes the differences we found; exploring that further may give a better understanding of what makes virtual reality effective. One important factor could be the rich interactions available in a virtual reality environment. Being able to resize, move, and adjust passages may have created a personal attachment between a participant and the passages, and therefore facilitated engagement of the tasks. It is possible to use interactions on a desktop to support panning the picture to imitate some of the interactions. This open question suggests a future study to see if interactions can improve participants' memory performance. A further step could be to navigate in the scene on a desktop. These would be interesting research avenues to explore in the future, now that our initial findings support the value of mnemonic methods in virtual reality for knowledge retrieval. Furthermore, participants also mentioned that they would like the scene in virtual reality built on top of real-world locations such as their home. Real-world locations often change unpredictably, which would interfere with their use as a memory aid. Yet we can load any personalized scenes and put any readings in virtual reality. We can reconstruct physical locations [79] using techniques like Google ARCore [80] and load the models in virtual reality. Further work could extract and visualize the ideas from passages in virtual reality, similar to visualizing the words for memorization [24].

4.3 Virtual Reality and Human Memory

VR is effective for a number of reasons. First, it is a spatialization of the knowledge so that participants were able to use their spatial memory to aid their verbal memory. Second, it is an externalization of an internal representation to help participants cope with harder problems [81]. Third, it uses visual cues to help participants remember the information, compared to a conventional memory palace built in the mind. Last, it may aid sentence comprehension and activate the processes of semantic and syntactic.

Our virtual reality-based memory palace variant is an instance of spatialization. Using virtual reality techniques helps build a *cognitive map* [82] as a knowledge spatialization. Using a hierarchical memory palace, virtual reality helps construct a *fuzzy cognitive map* [83] that represents the understanding of knowledge and reasoning about the information flow in the passages, serving as a mental *concept map* for remembering structural materials. Furthermore, the physical navigation involved in using virtual reality devices may also improve participants' performance. Previous research on large displays [6], [7], [42] and 3D interfaces [84] shows that user performance improved in cognitive tasks requiring spatialization when utilizing physical navigation over virtual navigation. In our experiment, IMAGE led to mental navigation, while VR involved physical navigation. A physical space can offer people room to organize ideas and build information flow. Physical navigation may have better invoked participants' spatial memory [84], and therefore participants were better at utilizing spatial cues and building connections to the passages. We analyzed some of the movements data, and we anticipate that future work could continue exploring the relationship between users' movements and memory performance. Last, the potential personal attachments created between participants and the passages (see Section 4.2) may also have helped invoke spatial memory.

Virtual reality techniques also help externalize internal representations of information and map them to spatial cues. Externalization is the projection of internal characteristics onto the outside world [81], [85], widely used in problem-solving and diagrammatic reasoning [86]; visualization per se is an instance of externalization [87]. The externalization process reduces remembering and recalling efforts. Participants may not have to remember the actual knowledge but can mentally visualize the knowledge presented at the loci. They recall the knowledge by imagining the external representations associated with the loci.

Virtual reality techniques also cue participants with visual information unavailable in a conventional memory palace or on a large display [7]. This observation can be explained in the way that visual embellishments are easier to recall than a picture and a story, but they do not detract from graphical comprehension [88]. In addition, natural-looking objects and redundancy help visualization recall [89].

More specific to different types of human memory, using virtual reality techniques helps concretize the abstract concepts and map them to the vivid visual and spatial cues offered in an immersive environment. In this way, participants might have been using their episodic memory ("the memory of personally experienced events" [90]) for aiding their semantic memory ("knowledge about the world in the broadest sense" [91]). For the participants who utilized a picture and a story, they may have to concretize both the spatial cues and the abstract concepts in the passages. As such, they might have been building connections between abstract concepts (e.g., an abstract coffee mug for abstract knowledge).

Last, we speculate that such virtual reality-based memory palace variants may match the processes of natural language understanding and sentence recall. Although the relationship between memory and language is still an open area, virtual reality techniques offer detailed and vivid clues to invoke episodic memory needed in understanding natural materials; a hierarchical memory palace naturally aligns with the lexical, semantic, and syntactic representations of the sentences [16] and may also have a physiological basis [30].

4.4 Limitations and Biases

There were some limitations with our experimental setups. The first one is that the verbal and visuospatial scores are post-experiment subjective ratings. Using standard psychological tests before the experiment (e.g., [71], [92]) could stabilize the scores. Second, the picture used in IMAGE has a higher rendering quality than the 3D scene used in VR (Fig. 3a vs. 2). This difference is unlikely to explain the lower performance in IMAGE because the picture and the 3D scene were rendered from the same 3D model using V-Ray and both nearly at a photorealistic level. Third, our virtual reality setup lacked a real world clock; the time reminder from the experimenter at the end of the experimenter might have interrupted participants' cognitive processes, and they might have been less engaged. Last, there were a few distractors in the experiment. For example, the HMD was quite heavy (about 1.2 lbs without cables) and limited by cables; the experimenter had to walk around and move the cables away from participants as they moved. Our experimental environment was not consistently quiet, possibly distracting participants.

Self-serving bias [93] (i.e., interpreting ambiguous information to serve one's own interests) and response bias (i.e., altering one's responses to serve the interests of the experimenter) [94] may exist in our experiment. Potential participants who have a good memory or are interested in virtual reality might be more likely to participate. The participants assigned to IMAGE may be disappointed by not using virtual reality. Alternatively, participants using virtual reality might be more engaged in the experiment. One last possible bias is that participants interacted more with the experimenter in VR (e.g., by helping them put on virtual reality devices), and this may have altered participants' behavior.

4.5 Related Work and Connection to Our Study

Our experiment and discussion acknowledge that a memory palace is a well-known technique for memory enhancement (e.g., virtual reality [24], [74], conventional [19]), and a personalized memory palace is not always necessary (e.g., [32], [74]). All these cited publications used memorization tasks (i.e., word-for-word) and asked participants to memorize a list of items, usually words [19], [74].

One key difference in our study is that we tackle knowledge retrieval from scholarly articles instead of low-level memorization such as remembering a list of items. The tasks we used were not to simply memorize words that had been known to participants. Participants had to make sense of, organize, and remember the main ideas behind the passages. Each of our participants read about 800 words in total, while the task used in the literature was to memorize dozens of words [19], [24], [74]. Our results show that virtual reality can support high-level cognitive tasks. In addition, our tasks incorporated a reading process, suggesting that people are able to read articles in virtual reality with a state-of-the-art HMD (HTC Vive, 2017 model).

These two claims may contrast with some of the literature, which states that spatial information in virtual reality could lead to *insignificant* improvements over non-spatial or non-immersive environments for graphical learning and memorization [4], [22], [23], [24], [39], [95].

Our study is different in several ways. First, we gave participants a clear strategy to use—we guided participants to build a hierarchical memory palace and move along a predefined path—therefore participants were able to employ spatial cues and organize information efficiently. Second, the other cues in virtual reality, such as rich interactions, visual cues, and physical navigation, may also contribute; the physical space that participants used seems larger than other virtual reality environments used in the literature (e.g., $5m^2$ [96]); using a coffee shop scene and high rendering quality also adds familiarity and immersiveness. Third, our high-level cognitive task is verbal-centric, in contrast to the visual-centric tasks used (e.g., video games) in the literature that can be mixed with the rich visual cues in a virtual reality environment. Thus, our results are consistent with the findings that spatial cues in virtual reality can help verbal recall [24], [74], [95]. The reason for this improvement in recall could be that retrieval cues help the long-term store of verbal memory [97]. Last, there might be a misinterpretation of *insignificant* results [98] in the literature; an insignificant result means that we are not able to observe an effect given the data; however, we should not conclude that the effect does not exist.

5 Conclusion

This paper explored the use of virtual reality techniques inspired by a mnemonic device called a "memory palace" to assist in knowledge retrieval from scholarly articles. We hypothesized that an extended version of a memory palace, which we call a hierarchical memory palace, may better match human memory for remembering and retrieving scientific knowledge from scholarly articles. We found that an image-based memory palace variant did not improve knowledge retrieval and was not favored by participants. However, when using a virtual reality-based memory palace variant, participants improved their recall accuracy and precision by mentally visualizing the knowledge items, mapping them to loci, and navigating the space of loci. Our work corroborates the proposition that virtual reality supports high-level cognitive tasks such as reading and remembering concepts in complicated documents. More broadly, this work offers insights supporting the value of virtual reality for application design. For future work, our method might be enhanced by building personalized spatial cues and choosing more efficient loci.

Acknowledgments

We thank all the anonymous reviewers for their constructive comments. We thank Natasha Danas, Wenyue Xing, and Ailin Deng for their valuable feedback. We thank Katrina Avery, Cyrus Maden, and Sara Gramley for their help with the manuscript.

References

[1] M. Lamming and M. Flynn, "Forget-me-not: Intimate computing in support of human memory," 1994, pp. 125–128.

[2] G. Robertson, M. Czerwinski, K. Larson, D. C. Robbins, D. Thiel, and M. van Dantzich, "Data mountain: using spatial memory for document management," pp. 153–162, 1998.

[3] A. Skupin and S. I. Fabrikant, "Spatialization methods: a cartographic research agenda for non-geographic information visualization," *Cartography and Geographic Information Science*, vol. 30, no. 2, pp. 99–119, 2003.

[4] E. D. Ragan, D. A. Bowman, and K. J. Huber, "Supporting cognitive processing with spatial information presentations in virtual environments," *Virtual Reality*, vol. 16, no. 4, pp. 301–314, 2012.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3009003, IEEE Transactions on Visualization and Computer Graphics

[5] M. Tory, C. Swindells, and R. Dreezer, "Comparing dot and landscape spatializations for visual memory differences," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1033–1040, Nov 2009.

[6] C. Andrews and C. North, "The impact of physical navigation on spatial organization for sensemaking," *IEEE Transactions on Visualization & Computer Graphics*, vol. 19, no. 12, pp. 2207–2216, 2013.

[7] E. D. Ragan, A. Endert, D. A. Bowman, and F. Quek, "How spatial layout, interactivity, and persistent visibility affect learning with large displays," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, ser. AVI '12. ACM, 2012, pp. 91–98.

[8] M. Tory, D. Sprague, F. Wu, W. Y. So, and T. Munzner, "Spatialization design: Comparing points and landscapes," *IEEE Transactions on Visualization & Computer Graphics*, vol. 13, no. 6, pp. 1262–1269, 2007.

[9] O.-H. Kwon, C. Muelder, K. Lee, and K.-L. Ma, "A study of layout, rendering, and interaction methods for immersive graph visualization," *IEEE Transactions on Visualization & Computer Graphics*, vol. 22, no. 7, pp. 1802–1815, 2016.

[10] F. A. Yates, *The art of memory*. Routledge and Kegan Paul London, 1966.

[11] M. Carruthers and J. M. Ziolkowski, *The medieval craft of memory: an anthology of texts and pictures*. University of Pennsylvania Press, 2003.

[12] J. Lienhard, "Ricci's memory palace," 2005.

[13] G. H. Bower, "Analysis of a mnemonic device: Modern psychology uncovers the powerful components of an ancient system for improving memory," *American Scientist*, vol. 58, no. 5, pp. 496–510, 1970.

[14] A. Y. Wang and M. H. Thomas, "Looking for long-term mnemonic effects on serial recall: The legacy of simonides," *The American journal of psychology*, vol. 113, no. 3, p. 331, 2000.

[15] H. L. Roediger, "The effectiveness of four mnemonics in ordering recall," *Journal of Experimental Psychology: Human Learning and Memory*, vol. 6, no. 5, pp. 558–567, 1980.

[16] E. Jefferies, M. A. L. Ralph, and A. D. Baddeley, "Automatic and controlled processing in sentence recall: The role of long-term and working memory," *Journal of Memory and Language*, vol. 51, no. 4, pp. 623 – 643, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0749596X04000853

[17] H. F. Crovitz, "Memory loci in artificial memory," *Psychonomic Science*, vol. 16, no. 2, pp. 82–83, 1969.

[18] F. S. Bellezza and B. G. Reddy, "Mnemonic devices and natural memory," *Bulletin of the Psychonomic Society*, vol. 11, no. 5, pp. 277–280, 1978.

[19] 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*, vol. 141, no. 3, pp. 380–390, 2012.

[20] O. Rosello, M. Exposito, and P. Maes, "Nevermind: Using augmented reality for memorization," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, ser. UIST '16 Adjunct. ACM, 2016, pp. 215–216.

[21] J. Ross and K. A. Lawrence, "Some observations on memory artifice," *Psychonomic Science*, vol. 13, no. 2, pp. 107–108, 1968.

[22] E. Krokos, C. Plaisant, and A. Varshney, "Spatial mnemonics using virtual reality," in *Proceedings of the 2018 10th International Conference on Computer and Automation Engineering*. ACM, 2018, pp. 27–30.

[23] E. Krokos, C. Plaisant, and A. Varshney, "Virtual memory palaces: immersion aids recall," *Virtual Reality*, pp. 1–15, 2018.

[24] J.-P. Huttner, D. Pfeiffer, and R.-B. Susanne, "Imaginary versus virtual loci: Evaluating the memorization accuracy in a virtual memory palace," in *Proceedings of Hawaii International Conference on System Sciences (HICSS)*, 2018.

[25] S. Ruth Cohen, "Influence of organizing strategies, time, and grade point averages on retention performance," *The Journal of Educational Research*, vol. 70, no. 4, pp. 219–221, 1977.

[26] M. A. Mastropieri, T. E. Scruggs, and J. R. Levin, "Maximizing what exceptional students can learn: A review of research on the keyword method and related mnemonic techniques," *Remedial and Special Education*, vol. 6, no. 2, pp. 39–45, 1985.

[27] S. Foley, "Virtual environment for the navigation of ideas and concepts in education (V.E.N.I.C.E)," in *Proceedings of the International Conference on Advanced Visual Interfaces*, ser. AVI '10. ACM, 2010, pp. 428–428.

[28] M. Tavakoli and E. Gerami, "The effect of keyword and pictorial methods on efl learners vocabulary learning and retention," no. 19, pp. 299–316, 2013.

[29] A. Qureshi, F. Rizvi, A. Syed, A. Shahid, and H. Manzoor, "The method of loci as a mnemonic device to facilitate learning in endocrinology leads to improvement in student performance as measured by assessments," *Advances in Physiology Education*, vol. 38, no. 2, pp. 140–144, 2014.

[30] A. D. Friederici, "The brain basis of language processing: from structure to function," *Physiological reviews*, vol. 91, no. 4, pp. 1357–1392, 2011.

[31] T. Dalgleish, L. Navrady, E. Bird, E. Hill, B. D. Dunn, and A.-M. Golden, "Method-of-loci as a mnemonic device to facilitate access to self-affirming personal memories for individuals with depression," *Clinical Psychological Science*, vol. 1, no. 2, pp. 156–162, 2013.

[32] G. G. Briggs, S. Hawkins, and H. F. Crovitz, "Bizarre images in artificial memory," *Psychonomic Science*, vol. 19, no. 6, pp. 353–354, 1970.

[33] E. Fassbender and W. Heiden, "The virtual memory palace," *Journal of Computational Information Systems*, vol. 2, no. 1, pp. 457–464, 2006.

[34] T. Jund, A. Capobianco, and F. Larue, "Impact of frame of reference on memorization in virtual environments," in *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*, July 2016, pp. 533–537.

[35] J. Harman, "Creating a memory palace using a computer," in *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '01. ACM, 2001, pp. 407–408.

[36] G. Plancher, S. Nicolas, and P. Piolino, "Virtual reality as a tool for assessing episodic memory," in *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '08. ACM, 2008, pp. 179–182.

[37] M. Bricken, "Virtual reality learning environments: potentials and challenges," *ACM SIGGRAPH Computer Graphics*, vol. 25, no. 3, pp. 178–184, 1991.

[38] E. D. Ragan, A. Sowndararajan, R. Kopper, and D. A. Bowman, "The effects of higher levels of immersion on procedure memorization performance and implications for educational virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 19, no. 6, pp. 527–543, 2010.

[39] G. Mineev, "The impact of immersive virtual reality on effectiveness of educational games," Master's thesis, Utrecht University, 2017.

[40] Z. Pan, A. D. Cheok, H. Yang, J. Zhu, and J. Shi, "Virtual reality and mixed reality for virtual learning environments," *Computers & Graphics*, vol. 30, no. 1, pp. 20–28, 2006.

[41] B. Schmidt and S. Stewart, "Implementing the virtual reality learning environment: Second life," *Nurse Educator*, vol. 34, no. 4, pp. 152–155, 2009.

[42] R. Ball, C. North, and D. A. Bowman, "Move to improve: Promoting physical navigation to increase user performance with large displays," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '07. ACM, 2007, pp. 191–200.

[43] D. R. Krathwohl, "A revision of bloom's taxonomy: An overview," *Theory into practice*, vol. 41, no. 4, pp. 212–218, 2002.

[44] E. D. Ragan, "Supporting learning through spatial information presentations in virtual environments," Ph.D. dissertation, Virginia Tech, 2013.

[45] "The Noun Project," https://thenounproject.com.

[46] G. Gillund and R. M. Shiffrin, "A retrieval model for both recognition and recall," *Psychological Review*, vol. 91, no. 1, pp. 1–67, 1984.

[47] M. Jarvis and J. Russell, *Key ideas in psychology*. Nelson Thornes, 2002.

[48] B. Mullen and C. Johnson, *The Psychology of Consumer Behavior*. Psychology Press, 2013.

[49] C. Spielberger, *Encyclopedia of Applied Psychology*. Academic Press, 2004.

[50] M. Daneman and P. A. Carpenter, "Individual differences in working memory and reading," *Journal of Verbal Learning and Verbal Behavior*, vol. 19, no. 4, pp. 450–466, 1980.

[51] M. Denis, "Imaging while reading text: A study of individual differences," *Memory & Cognition*, vol. 10, no. 6, pp. 540–545, 1982.

[52] D. R. Goodenough, "The role of individual differences in field dependence as a factor in learning and memory," *Psychological Bulletin*, vol. 83, no. 4, pp. 675–694, 1975.

[53] Y. Ikei and H. Ota, "Spatial electronic mnemonics for augmentation of human memory," in *2008 IEEE Virtual Reality Conference*, March 2008, pp. 217–224.

[54] A. Colley, J. Rantakari, and J. Häkkilä, "Augmenting the home to remember: Initial user perceptions," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, ser. UbiComp '14 Adjunct. ACM, 2014, pp. 1369–1372.

[55] P. Isola, J. Xiao, A. Torralba, and A. Oliva, "What makes an image memorable?" in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 145–152.

[56] H. F. Crovitz, "The capacity of memory loci in artificial memory," *Psychonomic Science*, vol. 24, no. 4, pp. 187–188, 1971.

[57] K. Rayner, T. J. Slattery, and N. N. Bélanger, "Eye movements, the perceptual span, and reading speed," *Psychonomic Bulletin & Review*, vol. 17, no. 6, pp. 834–839, 2010.

[58] D. Lagun and E. Agichtein, "Viewser: Enabling large-scale remote user studies of web search examination and interaction," in *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2011, pp. 365–374.

[59] K. M. Markham, J. J. Mintzes, and M. G. Jones, "The concept map as a research and evaluation tool: Further evidence of validity," *Journal of research in science teaching*, vol. 31, no. 1, pp. 91–101, 1994.

[60] "HTC Vive," https://www.vive.com/us/.

[61] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, "Surround-screen projection-based virtual reality: The design and implementation of the cave," in *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '93. ACM, 1993, pp. 135–142.

[62] "V-Ray," https://www.chaosgroup.com/vray/revit.

[63] J. P. Kincaid, R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom, "Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel," Naval Technical Training Command Millington TN Research Branch, Tech. Rep., 1975.

[64] "Readability analyzer estimates the readability of a passage of text using the flesch reading ease, fog scale level, flesch-kincaid grade level, and other metrics." https://datayze.com/readability-analyzer.php.

[65] U. Schiefele and A. Krapp, "Topic interest and free recall of expository text," *Learning and Individual Differences*, vol. 8, no. 2, pp. 141–160, 1996.

[66] R. E. Mayer, "Structural analysis of science prose: Can we increase problem-solving performance?" *ACM SIGDOC Asterisk Journal of Computer Documentation*, vol. 19, no. 3, pp. 3–25, 1995.

[67] W. Kintsch, "Text comprehension, memory, and learning." *American Psychologist*, vol. 49, no. 4, pp. 294–303, 1994.

[68] G. Cumming, "The new statistics: Why and how," *Psychological Science*, vol. 25, no. 1, pp. 7–29, 2014.

[69] P. Dragicevic, *Fair Statistical Communication in HCI*. Springer International Publishing, 2016, pp. 291–330.

[70] R. F. Lorch, E. P. Lorch, and W. E. Inman, "Effects of signaling topic structure on text recall." *Journal of Educational Psychology*, vol. 85, no. 2, pp. 281–290, 1993.

[71] J. Brandt, "The Hopkins Verbal Learning Test: Development of a new memory test with six equivalent forms," *The Clinical Neuropsychologist*, vol. 5, no. 2, pp. 125–142, 1991.

[72] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies*, vol. 2, pp. 37–63, 2011.

[73] K. B. McDermott, "The persistence of false memories in list recall," *Journal of Memory and Language*, vol. 35, no. 2, pp. 212–230, 1996.

[74] J.-P. Huttner and R.-B. Susanne, "An immersive memory palace: Supporting the method of loci with virtual reality," in *Proceedings of Americas Conference on Information System (AMCIS)*. ACM, 2017.

[75] A. Tom, T. A. S. R. J. Bosker, and R. J. Bosker, *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. Sage, 1999.

[76] 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*, vol. 26, no. 1, pp. 536–546, 2019.

[77] Y. Yang, T. Dwyer, B. Jenny, K. Marriott, M. Cordeil, and H. Chen, "Origin-destination flow maps in immersive environments," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, p. 693703, Jan 2019. [Online]. Available: http://dx.doi.org/10.1109/TVCG.2018.2865192

[78] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[79] G. Vosselman, S. Dijkman *et al.*, "3D building model reconstruction from point clouds and ground plans," *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, vol. 34, no. 3/W4, pp. 37–44, 2001.

[80] "ARCore," https://developers.google.com/ar/discover/.

[81] D. Kirsh, "Thinking with external representations," *AI & Society*, vol. 25, no. 4, pp. 441–454, 2010.

[82] E. C. Tolman, "Cognitive maps in rats and men," *Psychological Review*, vol. 55, no. 4, pp. 189–208, 1948.

[83] B. Kosko, "Fuzzy cognitive maps," *International Journal of Man-machine Studies*, vol. 24, no. 1, pp. 65–75, 1986.

[84] A. Cockburn and B. McKenzie, "Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '02. ACM, 2002, pp. 203–210.

[85] J. Sandler, *Projection, Identification, Projective Identification*. Karnac Books, 1988.

[86] J. Zhang, "The nature of external representations in problem solving," *Cognitive Science*, vol. 21, no. 2, pp. 179–217, 1997.

[87] T. Munzner, *Visualization Analysis and Design*. CRC press, 2014.

[88] S. Bateman, R. L. Mandryk, C. Gutwin, A. Genest, D. McDine, and C. Brooks, "Useful junk?: The effects of visual embellishment on comprehension and memorability of charts," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '10. ACM, 2010, pp. 2573–2582.

[89] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister, "What makes a visualization memorable?" *IEEE Transactions on Visualization & Computer Graphics*, vol. 19, no. 12, pp. 2306–2315, 2013.

[90] E. Tulving, "What is episodic memory?" *Current Directions in Psychological Science*, vol. 2, no. 3, pp. 67–70, 1993.

[91] E. Tulving, "Episodic memory: from mind to brain," *Annual review of psychology*, vol. 53, no. 1, pp. 1–25, 2002.

[92] R. H. Benedict, D. Schretlen, L. Groninger, M. Dobraski, and B. Shpritz, "Revision of the Brief Visuospatial Memory Test: Studies of normal performance, reliability, and validity," *Psychological Assessment*, vol. 8, no. 2, pp. 145–153, 1996.

[93] D. T. Miller and M. Ross, "Self-serving biases in the attribution of causality: Fact or fiction?" *Psychological bulletin*, vol. 82, no. 2, pp. 213–225, 1975.

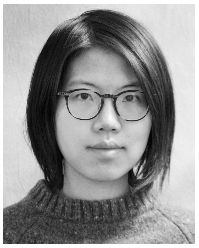
[94] N. Dell, V. Vaidyanathan, I. Medhi, E. Cutrell, and W. Thies, ""Yours is better!": Participant response bias in HCI," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012, pp. 1321–1330.

[95] E. D. Ragan, K. J. Huber, B. Laha, and D. A. Bowman, "The effects of navigational control and environmental detail on learning in 3D virtual environments," in *Virtual Reality Short Papers and Posters (VRW), 2012 IEEE*. IEEE, 2012, pp. 11–14.

[96] P. Häfner, C. Vinke, V. Häfner, J. Ovtcharova, and W. Schotte, "The impact of motion in virtual environments on memorization performance," in *Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 104–109.

[97] F. I. Craik and R. S. Lockhart, "Levels of processing: A framework for memory research," *Journal of verbal learning and verbal behavior*, vol. 11, no. 6, pp. 671–684, 1972.

[98] M. Kaptein and J. Robertson, "Rethinking statistical analysis methods for CHI," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012, pp. 1105–1114.



Fumeng Yang is a PhD student in the Department of Computer Science at Brown University. Her research interests include information visualization, virtual reality, and human-computer interaction.

Jing Qian is a PhD student in the Department of Computer Science at Brown University. His research interests include human-computer interaction, augmented reality, and mobile sensing.

Johannes Novotny is a PhD candidate in the Department of Computer Science at Brown University. His research interests include scientific visualization, virtual reality, and computer graphics.

David Badre is an Associate Professor in the Department of Cognitive, Linguistic, and Psychological Sciences. He received his Ph.D. from the Department of Brain and Cognitive Sciences at Massachusetts Institute of Technology, following a postdoctoral fellowship at the University of California, Berkeley. His research interest focuses on the cognitive neuroscience of memory and cognitive control with an emphasis on frontal lobe function and organization.

Cullen D. Jackson received his PhD degree in experimental psychology from Brown University. He was a postdoctoral research associate in the Computer Science Department at Brown University. He is currently an Instructor in Anaesthesia at Harvard Medical School, and the Director of Innovation in the Department of Anesthesia, Critical Care and Pain Medicine at Beth Israel Deaconess Medical Center. He is a member of the IEEE.

David H. Laidlaw is a professor in the Computer Science Department at Brown University. He received his PhD degree in computer science from the California Institute of Technology, where he also did post-doctoral work in the Division of Biology. His research centers on applications of visualization, modeling, computer graphics, and computer science to other scientific disciplines. He is a fellow of the IEEE.