Changing Between Virtual Reality and Real-World Adversely Affects Memory Recall Accuracy

Maarten H. Lamers * *and Maik Lanen*

Leiden Institute of Advanced Computer Science, Leiden University, Leiden, Netherlands

Context-dependency effects on memory exist, whereby people's context influences their ability to accurately recall items from memory. This effect was not previously studied when considering virtual reality as an environmental context. We show that adverse effects on recall of memorized items exist when changing between virtual and real environments. The effect was not present when memorizing and recall were both done in VR; it appears to be caused by the change of environmental context. This previously unknown effect may impact how we use VR for memorization tasks, particularly when accurate recall of memorized information in a real environment is important. In a memory-recall experiment (n = 51) participants that underwent a context change involving VR after memorizing performed significantly worse on 24-h later item recall than those who did not change context (17% lower accuracy, $p < 0.001$). In particular memorizing in VR as opposed to a real environment lowers accuracy of recall in a real environment (24% lower, $p = 0.001$).

Keywords: virtual reality, memory, context-dependency, recall, transfer

INTRODUCTION

It is possible that you have had to memorize words to meet an educational requirement. Sitting at home you memorize the given words until you are able to recall them well. The next day at school, you are quizzed on your ability to recall the words and you find yourself forgetting a portion of the words. However, upon your return home, when sitting behind your desk again, somehow you can recall most of the "forgotten" words. How is that possible?

This could be caused by a memory effect known as *context-dependency*, wherein your context influences the ability to recall items from your memory. For example, Godden and Baddeley (1975) let two groups of divers memorize words in different contexts, namely under water and on land. Later, all participants were asked to recall the words while being in either the same or the other context. It was found that subjects asked to recall in the other context than the one in which they had memorized the items did so significantly worse than those who were asked to recall them in the same context. Apparently, a change in context negatively affected the ability to recall items memorized. This effect was later found to exist for other context changes also. A good overview, in our opinion, is provided by Smith and Vela (2001).

The relation of context-dependent memory and virtual reality is an interesting topic to research as the role of VR has become more prevalent in society. For example, consider a surgeon who studied a 3D visualization of the location of tumors. Will they recall these positions in reality as well as in training? Or will the change of context from VR to reality affect their recall ability? An analogous case could be argued in light of increasing use of VR in

education. Our study aims to uncover knowledge about context-dependent memory and its relation to VR.

To uncover whether the change of context from VR to real environment and vice-versa affects our ability to recall items memorized, we hypothesize that such a change negatively affects our ability to recall when compared to an unchanged context, analogous to effects shown in prior studies. To test our hypothesis, we constructed and realized an empirical study ($n = 51$), based on methods and findings from prior studies. Our analyses take into account potential effects caused by the use of VR as a context and focus on the effects of context change.

We firstly explore the most relevant related work done to-date, specifically in relation to memory, context-dependency, and VR. Based on the review of prior studies we construct the experiment, of which all choices are described in *Methods and Materials Section*. Analyses of the results are presented and discussed in light of our hypothesis and potential implications of our findings.

This work was undertaken as the M.Sc. thesis work of Lanen (2018) in Spring 2018. It was presented and published in a more concise form at the 15th EuroVR International Conference in October 2018 (Lanen and Lamers, 2018).

RELATED WORK

Here we concisely sketch an overview of prior results that are of interest to our hypothesis. Selected prior research is mentioned, not aiming to be exhaustive.

Memory and Context

In a famous study, Goodwin et al. (1969) performed an experiment whereby participants performed four different memory tasks while being sober or being under the influence of alcohol. After 24 h, the participants were tested under both the conditions. Results showed that participants were better in recalling memorized items when being tested in the same state they memorized in. Their ability to recognize had not been altered by the different states; thus, not all sub-forms of memory were affected. The memorization and recall tasks applied by Goodwin et al. (1969) form the basis of our own experimental design, as is explained in a further section.

As mentioned in our Introduction, Godden and Baddeley (1975) showed that the context-dependency effect on memory extends to the external environment also, i.e., under water vs. on land. From this it was concluded that the disruption in moving from one environment to another negatively influences the ability to recall memory memorized in the first environment.

This outcome firmly underlies our own study, as we also study potential context-dependency effects on memory caused by moving from one environment to another. Specifically, we consider disruption caused by moving from a real to a similar but virtual environment, and vice-versa.

Other prior results include an outcome that no context-dependency was found in a memory-recall task that considered different tastes as contexts–one context being created by chewing mint-flavored gum, the other by chewing flavorless gum (Johnson and Miles, 2008). Also, a change of mood (happy/sad) was found to adversely affect ability to recall memorized items (Bower et al., 1978), as opposed to when mood was unchanged. When given a task that relies less on contextual information, but more on introspective thought for example, then context-dependency effects on memory-recall tasks are less strong (Smith and Vela, 2001).

Memory and Virtual Reality

The validity of using VR to assess learning and memory skills in brain-injured and healthy individuals was studied by Matheis et al. (2007). The authors found a significant correspondence between their VR-based assessment of memory and a standard neuropsychological measure. From this they conclude that VR "provides a viable medium for measuring learning and memory" (Matheis et al., 2007).

A study that compared potential effects on memory between active and passive participation in a virtual environment (Attree et al., 1996), found that participants with active participation tested as having a better memory for spatial layouts while participants who passively participated tested higher in object recall.

Naturally, how realistic a virtual environment is, is an interesting aspect to consider in relation to memory recall tasks. This was studied by Dinh et al. (1999) by way of a multi-modal experience wherein subjects could smell, feel and hear a virtual environment. Results showed that by offering tactile input, the quality of presence in VR was enhanced, making it easier for the participants to remember objects in the virtual environment. Auditory and olfactory stimuli only increased the feeling of being present in the virtual world, but had no effect on memorization.

Perhaps less related to our work are studies in which personal demographics such as age (Plancher et al., 2010), medical conditions such as non-progressive brain injury (Sweeney et al., 2010) or mental parameters such as depression (Gould et al., 2007) act as different contexts for memorization and spatial recall tasks in VR.

An excellent and well-structured review of virtual reality in episodic memory research is provided by Smith (2019). It reviews topical relevant works along multiple factors of interest and established terminology for distinguishing subtypes of VR systems. Within the framework set by Smith, our work is viewed best as what he terms research into "real-life vs. VR training", using "stereoscopic headset-VR" technology.

Virtual Reality and Context-dependent Memory, an Unexplored Area

We have shown that prior studies have compared VR and real environments with respect to their effects on memory. Moreover, VR was shown to be suitable for memory-recall assessment within the VR environment itself. However, no prior study was done to uncover possible context-dependency effects on memory when changing between virtual and real environmental contexts. It is exactly that knowledge gap which our study aims to fill.

METHODS AND MATERIALS

Our experimental design considers four randomly assigned groups of participants who must memorize and recall items in either the same or a different context. Condition-group assignment was based on the "random sequence generator" tool of random.org. Participants were recruited on voluntary basis from the personnel of a Leiden-based sports retail company and student population of the Communication and Multimedia Design course of the Rotterdam University of Applied Sciences. Twenty-four hours after memorizing, participants are assessed on their ability to recall the memorized items [cf. work by Godden and Baddeley (1975) and other studies]. Since in practice, assessing all participants after exactly 24 h may be difficult, the exact memorizing and recall times are noted and used to check for possible confounding effects of retention duration.

The study was undertaken at Leiden University's Computer Sciences department in Spring 2018. Being unaccustomed to human-subject experiments, the institute had no ethics committee in place at the time. Although we acknowledge that formal ethical clearance would be preferable, we trust that the experiment's rather "lightweight" design regarding ethical questions holds up to common research ethics standards. No minors and otherwise vulnerable participants are involved. No impact on lives and concerns of participants is expected and their safety was at all times guaranteed. No simulator sickness was expected (nor observed among participants), since exposure to VR was of short duration (Kennedy et al., 2000), participants were seated at all times during the experiment [a factor shown to reduce simulator sickness by Merhi et al. (2007)] and no virtual locomotion was involved. Informed consent was gathered on paper only, and not copied or distributed. Other data collected from participants were kept to a minimum, anonymized, and removed as far as they were not used. Digital data were stored on university systems and accessible only to the authors; local copies for use by the authors were removed after data processing.

Real and Virtual Environments

We aim to make the real and virtual environments, which form the contexts for our study into context-dependent memory, as similar in experience as possible. By doing so we focus on uncovering potential effects of a change between a real and similar virtual environment, as opposed to effects caused by unnecessary large discrepancies between both environments. Naturally, in the hypothetical case that the virtual environment is indistinguishable from the real environment, we would not expect such effects to exist. Although visual fidelity in VR was shown to enhance episodic memory for elements within the virtual environment [an overview is provided by Smith (2019)], its effect on transfer of memory from virtual to real environments appears to be only studied for spatial memory (Mania et al. 2003). To be relevant with regards to the current state of VR technology, we create a likeness between real and virtual environments that is exemplary of said current state.

As the scenario that forms the context of our memory-recall study, participants are seated in an office chair behind a desk. A book is placed centered on the desk in front of them. A potted plant stands on the left side of the desk surface. Wooden divider boards surround the desk's left, right and opposing ends. A poster is hung on the right divider board, at eye's height. The plant and office chair are the same in both the real and virtual scenarios and the divider boards and poster as similar as possible. In both the real and virtual contexts participants are seated in front of the real desk on the real office chair (**Figure 1**).

In the virtual context, participants are outfitted with an HTC Vive headset and a single HTC Vive controller, enabling them to look around and interact with the virtual book on the desk. Turning a page of the virtual book is done by moving the virtual hand toward the top right corner of the book, clicking and holding the trigger on the controller and making an overturning movement to thereafter release the trigger. In VR, the participant can see their own avatar's arm that is dressed in a blue sleeve and has light toned skin (as we expected most test participants to have light toned skin). No audio was incorporated into the virtual context and no audio headphones were worn by participants (ears were uncovered).

The study was designed such that participants performed the same tasks in both the real and virtual environment. In both environments they are only allowed to look around and turn over pages of the book in front of them. The experiment is executed in two different locations, namely in the office environment of a sports retail company in the city of Leiden and in an educational environment within a building of the Rotterdam University of Applied Sciences. Both memorizing and recall occur at the same experimental location for each participant.

Language and Demographics

Participants are given two language options from which they can choose to perform the experiment in: Dutch and English. We provide both language options because during a small pilot study some Dutch participants were observed to have difficulties performing the experiment in English. Since a lack of fluency in a language could affect memory recall or understanding of the task, participants can only participate in the experiment when they are fluent in Dutch or English, an assessment that is made by participants themselves.

Several demographic variables are collected via self-report from participants, namely gender, age, prior experience with VR, and education level. Demographic questions are formulated by the standards of the PGA Group Consulting Psychologists (2018), questions about education are formulated by the European Qualifications Framework (European Commission, 2012).

Orientation Task

After being introduced to their given context (real or VR), participants are verbally instructed to execute a simple orientation task within the given context. It is intended to make participants familiar with the given context and to ensure that the participant is proficient enough technologically to participate in the experiment. The orientation task consists of
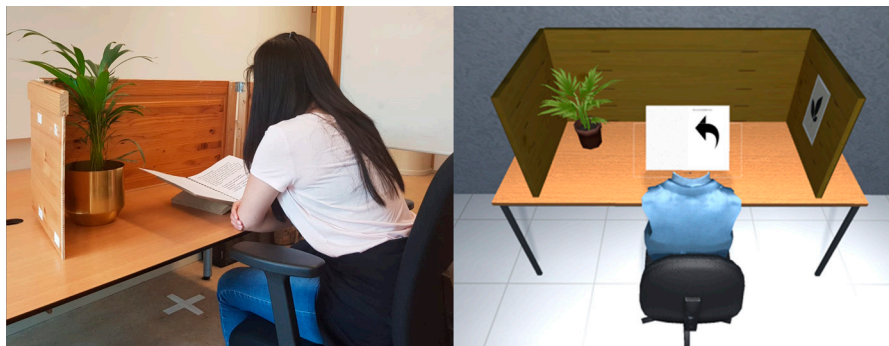
FIGURE 1 | Photograph of a participant in the real-world experimental setting **(LEFT)** and overview from an observer's point-of-view of the virtual experimental environment created in Unity **(RIGHT)**.

TABLE 1 | Examples of the four types of English and Dutch five-word sentences/lists used in the rote-learning task.

Type	English	Dutch
Normal sentence	I Walk to the station	Ik loop naar het station
Anomalous sentence	Colourless green ideas sleep furiously	Kleurloze groene ideeën slapen woedend
Anagram list	Drawer, redraw, reward, warder, warred	Mentors, stormen, stromen, `n stomer, `t morsen
Word list	Flatness, iron, harbor, crab, thief	Vlakheid, IJzer, haven, krab, dief

1) looking around at different objects (plant, poster, and book) and 2) executing an example task wherein the participant interacts with the book on the table by turning one page, either in reality or VR. On the day of recall, participants were asked to do the same orientation task as on the day of memorizing.

The orientation task is considered successfully completed if the participant is able to perform the requested interactions. If not, then the supervisor will try to help the participant to master the interactions. If the participant is then not able to master the interactions, then they are excluded from the experiment.

Rote-Learning Task

After the introduction and orientation task, participants must complete three tasks which were based on studies by Goodwin et al. (1969) and Marks and Miller (1964). These tasks are a rote-learning task, an association task and a recognition task. The resulting recall scores from the three tasks are averaged to form an overall "recall score" per participant. Here we start with describing the rote-learning task.

Memorizing

The rote-learning task consists of saying four 5-word sentences/lists (**Table 1**) with varying meaningfulness out loud, repeatedly, for the duration of 2 min [cf. Johnson and Miles (2008)]. Participants are asked to remember the sentences as they will be assessed on their ability to recall them after 24 h.

There are four types of sentences/lists: a normal sentence arbitrarily obtained from a magazine, an anomalous sentence obtained from a paper by Chomsky (1956), an anagram list obtained from online anagram generator litscape.com, and a

word list obtained from online tool textfixer.com. All were translated from English to Dutch, except for the anagram list, which was obtained from Dutch online tool mijnwoordenboek.nl. The sentences/lists are displayed to participants in the real and virtual book.

Recall

Participants were asked to recall the sentences/lists memorized within 2 min [cf. (Godden and Baddeley, 1975)]. Participants must say the sentences out loud while their voice is being recorded. The recorded answers were transcribed after the session to obtain the data, after which the recordings were deleted. Participants are prohibited to ask the experiment supervisor for hints regarding the memorized sentences.

Assessment

Recall performance is measured in terms of the number of sequence and omission errors. It is possible that participants recall a sentence or list differently than memorized, but the meaning is nearly the same. To score the similarity of recalled sentences to those that were memorized, we applied an online tool that was provided by Explosion.AI and based on the spaCy.io open-source library for natural language processing in *Python*. By using its word vector based "similarity" function and the large English language model "en_core_web_lg", it returns a similarity score in range 0–1. E.g., comparing "*I walk to the station*" and "*I walked to the station*" results in a similarity score of 0.97. For the word list and anagram list we use the fraction of correct answers. When a participant recalls a word that is close to the memorized word, then we assess this as explained through the following example. If a participant recalls

TABLE 2 | Ten English and Dutch low-association words used in the association task.

English	Dutch
Chance, cruel, lazy, melt, narrow, money, now, size, time, tall	Kans, wreed, lui, smelten, smal, geld, nu, maat, tijd, lang

English words were obtained from a study by Burke et al. (1987) and translated to Dutch.

the word list (**Table 1**) as "*Flat, Iron, Harbor, Crab, Thief*", we asses "*Flat*" as a semi-correct recall (the memorized word was in fact "*Flatness*"). Since the other four words were recalled correctly, the resulting score for this recall sequence is 0.9. It is also possible that a participant might say the recalled sentence with slight variance several times. We decided to only assess the last verbal submission to prevent lucky guesses. Every recalled sentence/list yields a score between zero and one. The final score for the rote-learning task is the average of the scores for all four memorized sentences/lists.

Association Task
Memorizing
Participants are asked to say out loud the first word that comes to mind in reaction to each of ten given low-association words obtained from a study by Burke et al. (1987). For Dutch speaking participants, these words were translated to Dutch (**Table 2**).

Recall
Given the same stimulus words as provided during memorizing, participants must recall the self-associated word and say them out loud.

Assessment
Only the final submission for each given stimulus word is assessed–previous submissions are not assessed. Scoring is done conform the assessment procedure for the rote-learning task.

Recognition Task
Memorizing
Participants are asked to memorize twenty different pictures, displayed in randomized order each on an individual page of the book. Ten pictures have emotional content (5 female and five male cover models from erotically themed magazines) and ten pictures have neutral content (5 female and five male mail-order catalog models), cf. the study by Goodwin et al. (1969). Participants are allowed to browse the images for a maximum duration of 5 min.

Recall
Participants are asked to select maximally twenty memorized pictures from forty pictures shown in the book, after browsing through the pictures for maximally 10 min. The twenty newly added pictures are of similar nature and distribution as the memorized pictures.

Assessment
The recognition task is scored as the fraction of memorized pictures correctly recalled. If participants change their mind

during the recall period, then their last submission is considered the final submission.

RESULTS

Within seven weeks in Spring 2018, 57 participants voluntarily did the memorizing tasks, of which 51 also returned for the recall tasks. The six non-returning subjects (two memorized in VR, four in the real environment) were not contacted and no reason for their absence is known. Their data was discarded and not expected to introduce selection bias. All participants opted for the Dutch language tasks, and all successfully completed the orientation tasks before both memorizing and recall.

In the below analyses and remainder of this paper, statistical significance is assumed when $p < 0.05$ and indicated with *. Student's T-tests are all two-tailed and assume equal variances. Normality is assessed through Shapiro-Wilk testing with $p < 0.05$ suggesting deviation from normality. Identification of continuous variable outlier values is done using the common "Q1−(1.5 · IQR)" and "Q3 + (1.5 · IQR)" boundaries. Removal of suspicious outliers is only decided after subjective scrutiny of the values. Following this strategy, two outlier values were identified but not removed, as discussed at the appropriate point within the remainder of this Section.

Table 3 and **Figure 2** compare the mean recall scores across various demographic sub-samples of the total sample ($n = 51$). From the appropriate p-values we conclude that participant gender, experimental testing location, occupational status, prior experience with VR, and age did not significantly affect the mean recall score.

Not all participants could perform the recall tasks exactly 24 h (1,440 min) after the memorizing tasks. **Figure 3** illustrates the correlation between retention duration in minutes (*mean* = 1,455, $SD = 104$, normally distributed) before recall and recall score (*mean* = 0.650, $SD = 0.134$, normally distributed). No evidence of significant correlation was found, leading us to conclude that variation in retention duration as found in our sample does not affect recall score.

As mentioned earlier, participants were randomly assigned to one of four environmental context conditions: "R + R" when both memorizing and recalling in the real context, "R + V" when memorizing in the real context and recalling in the virtual context, "V + R" when memorizing in the virtual context and recalling in the real context, and "V + V" when both memorizing and recalling in the virtual context. By joining groups R + R with V + V and groups R + V with V + R we obtain two new groups of participants: one for whom the environmental contexts during memorizing and recall were the same, and one for whom the context changed.

TABLE 3 | Outcomes of four independent samples Student's T-tests and a one-way ANOVA test (*p*-values and appropriate test statistics reported), comparing recall scores across various demographic sub-samples.

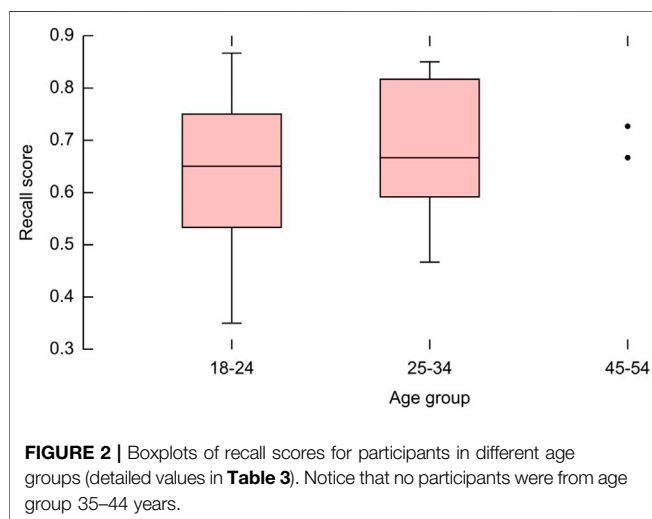| Sub-group	N	Shapiro-Wilk test	Mean (SD) recall score	T-test/ANOVA
Females	21	$p = 0.264$	0.635 (0.146)	$p = 0.516$ ($t_{49} = -0.655$)
Males	30	$p = 0.627$	0.660 (0.126)	
Tested at retail offices	19	$p = 0.388$	0.638 (0.157)	$p = 0.641$ ($t_{49} = -0.470$)
Tested at education location	32	$p = 0.520$	0.657 (0.121)	
Students	41	$p = 0.199$	0.646 (0.128)	$p = 0.722$ ($t_{49} = -0.358$)
Waged staff	10	$p = 0.384$	0.664 (0.163)	
Prior VR experience	27	$p = 0.353$	0.647 (0.130)	$p = 0.893$ ($t_{49} = 0.136$)
No VR experience	24	$p = 0.600$	0.653 (0.142)	
Aged 18–24	33	$p = 0.376$	0.633 (0.138)	$p = 0.479$ ($F_{2,48} = 0.747$)
Aged 25–34	16	$p = 0.193$	0.679 (0.132)	
Aged 35–44	0	n/a	n/a	
Aged 45–54	2	n/a	0.697 (0.042)	

FIGURE 2 | Boxplots of recall scores for participants in different age groups (detailed values in **Table 3**). Notice that no participants were from age group 35–44 years.

Figure 4, **5**, and **Table 4** compare mean recall scores across combinations of these groups. Participants whose environmental context changed score on average significantly lower ($p < 0.001$) than those for whom the context did not change; 17% lower in fact. This result confirms our main hypothesis that context-dependency effects as found in prior research exist when changing between environments of real and virtual nature also.

Two recall score values were identified as outliers within their respective condition groups. These are visible in **Figure 4** in context-condition groups R + R (under the lower boundary) and V + R (above the upper boundary). Scrutiny of the data could not account for their suspicious values. It was decided to not remove them from the analyses, since their values are within reasonable range with respect to the complete data sample, indicating not unreasonable participant efforts. Moreover, given their value relative to other values, their inclusion would at worst lead to *under*estimation of the hypothesized effect and is therefore the conservative choice regarding hypothesis testing and effect estimation.

When comparing recall scores for context-conditions R + R and V + R the relative decrease in recall accuracy is 24% when memorizing in the virtual as opposed to the real environment, and statistically significant ($p = 0.001$, **Table 4**). This result is particularly striking in face of current interests in VR as an environment for learning–a task of which item memorization may surely not be the most important element, but certainly one with potential impact.

Through a two-way ANOVA test, we compared mean recall scores using the memorizing and recall contexts as independent fixed factors, and including their interaction effect in the model. A Q-Q plot of the residuals (**Figure 6**) and Shapiro-Wilk test ($W = 0.990$, $p = 0.946$) suggest that the residuals are normally distributed. Outcomes of the 2×2 ANOVA test (**Table 5**) tell us that the contexts in which memorizing and recall are done do not by themselves affect the recall score in a statically significant manner, and that their effect size can be considered small ($\eta^2 < 0.05$). However, the interaction between both factors does show a statistically significant effect on mean recall scores ($p < 0.001$) with substantial effect size ($\eta^2 = 0.205$), supporting our hypothesis that recall scores are affected by context change and not by the memorizing or recall environmental context itself.

When we consider the separate memory tasks (**Table 6**), then we observe that a context change negatively affects all mean task-specific recall scores. However, statistical significance of this outcome is found for the rote-learning task and the recognition task, but not for the association task. Deviation from normality of the recognition scores was visually scrutinized and found to be caused by their tendency to be on average near the maximum score, yet with substantial variance.

DISCUSSION

This study investigates the previously unexplored area of context-dependency effects on memory when changing between virtual reality and real-world environments. The outcomes of our experiment clearly show that significant and substantial
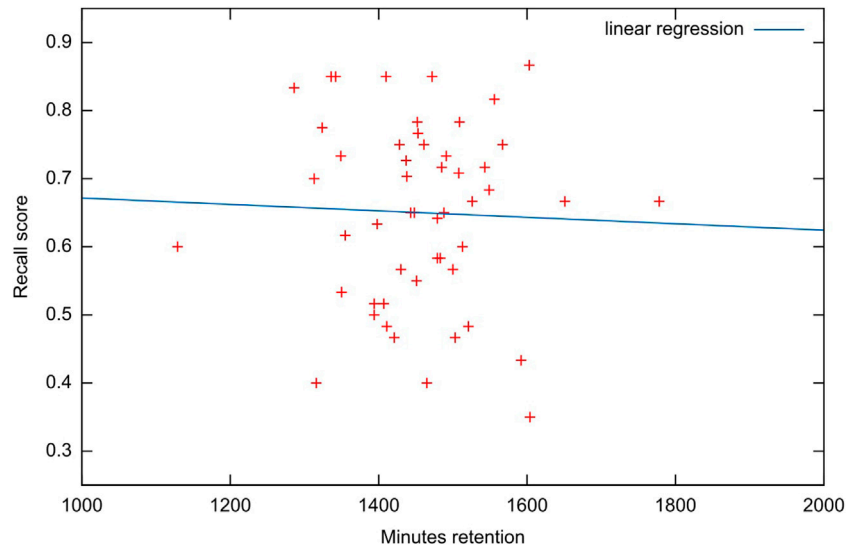
FIGURE 3 | Scatter plot correlating memory retention duration (in minutes) and recall score of all participants ($n = 51$, Pearson's $r = -0.037$, $p = 0.797$). Note that 24 h equals 1,440 min.
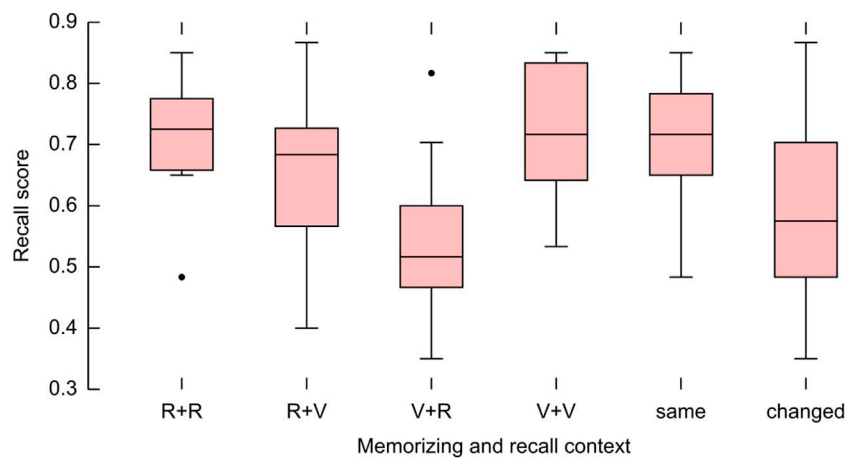

FIGURE 4 | Boxplots of recall scores for different sub-samples that are defined by their memorizing and recall context conditions (detailed values in **Table 4**).

adverse effects on memory recall accuracy exist when changing environments, compared to when not changing environments. Reviewing the results, we conclude that they are valid to base conclusions on. Memory recall scores do not appear to be influenced by demographic factors in our sample and prior experience with VR. All participants were native Dutch speakers, meaning that no language effects are expected to have influenced our results.

We hypothesized that a change in environmental context from VR to real-world environment, and its inverse change, have a combined adverse effect on the recall accuracy of memorized items after 24 h. This hypothesis is supported by a highly significant ($p < 0.001$) difference in mean recall scores between participant sub-samples with and without a change of

environmental context. The group that changed environmental context during memory retention scored on average 17% lower on item recall accuracy, when compared to the group that did not change their environmental context. This main result is novel, in that context-dependency of memory recall was not yet studied using the environmental contexts of VR and real-world environments, and in that it uncovers significant effect size. Yet, these results correspond with prior research findings regarding changes of other environmental contexts such as being under water and on land (Godden and Baddeley, 1975) and changes of mental states such as being under the influence of alcohol and sober (Goodwin et al., 1969).

Another highly significant result is that the scored ability to accurately recall items in the real-world environment is 24%
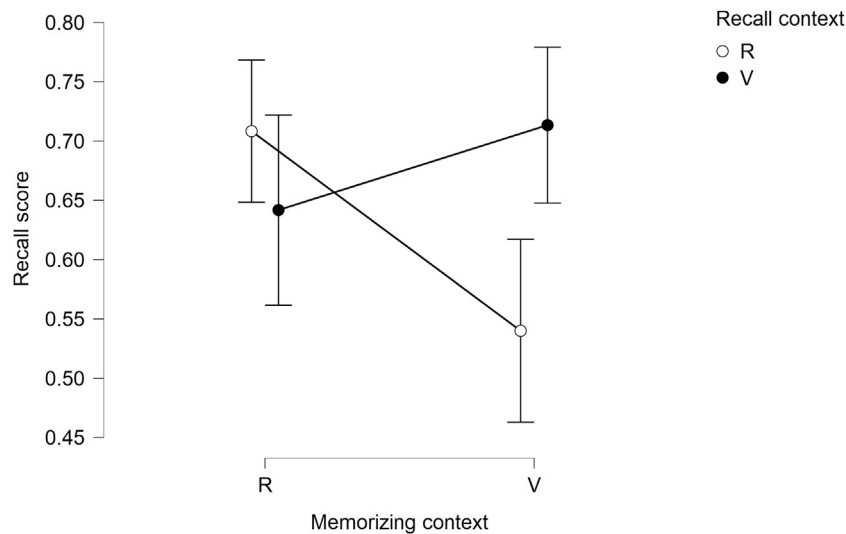
FIGURE 5 | Plot of mean recall scores with 95% confidence intervals for four different combinations of memorizing and recall contexts. Crossing of the connecting lines illustrates potential interaction of the memorizing and recall contexts on the recall score (detailed values in **Table 4**).

TABLE 4 | Outcomes of four independent samples Student's T-tests (p-values with t-statistic and degrees of freedom reported), each comparing mean recall scores of two sub-samples that are defined by their memorizing and recall context conditions.

Memorizing context	Recall context	N	Shapiro-Wilk test	Mean (SD) recall score	T-test
Real	Real	12	$p = 0.290$	0.708 (0.094)	$p = 0.901$ ($t_{23} = -0.125$)
Virtual	Virtual	13	$p = 0.263$	0.713 (0.109)	
Real	Virtual	13	$p = 0.944$	0.642 (0.133)	$p = 0.058$ ($t_{24} = 1.994$)
Virtual	Real	13	$p = 0.867$	0.540 (0.128)	
Real	Real	12	$p = 0.290$	0.708 (0.094)	$p = 0.001^*$ ($t_{23} = 3.724$)
Virtual	Real	13	$p = 0.867$	0.540 (0.128)	
Same context		25	$p = 0.278$	0.711 (0.100)	$p < 0.001^*$ ($t_{49} = 3.551$)
Changed context		26	$p = 0.794$	0.591 (0.138)	

lower when these items were memorized in VR when compared to memorization in a real environment ($p = 0.001$). This previously unknown effect may impact how we use VR for short-term memorization tasks, particularly when accurate recall of memorized information in a real environment is important.

One could conjecture that memorizing in VR is more difficult than memorizing in the real environment. Excitement for or novelty of VR might make it more difficult for participants to memorize in the VR context. However, if this conjecture were to hold, then one would hypothesize the V + V context group to score lower than the R + R group, which was not the case. Moreover, two-way ANOVA testing showed that memorizing and recall contexts separately do not significantly affect recall accuracy, but that their interaction does ($p < 0.001$) with substantial effect size ($\eta^2 = 0.205$). As such the alternative explanation for the findings, posed in this paragraph, is unsupported and the low recall performance under V + R condition is attributed to the change of environmental context.

We provided an example in our introduction of a surgeon who must memorize the position of tumors for later recall during surgery. Do the uncovered effects imply that it is unwanted to use VR for the memorization of information for this task? Or for other memory recall intensive tasks? We cannot answer this question with certainty. And seeing as how learning in a broader sense relies not on memorization alone, we cannot possibly dismiss VR as a learning environment. However, our findings do mean that we are less certain that VR can be effectively used for memorizing information whereby short-term accurate recall in reality is important.

When assessing the broader implications of the uncovered context-dependency effect, one should recognize that the distinction between real-world and VR environments is not the only distinction to consider. In many real-world scenarios, the real-world memorization context is forcibly different from the recall context. Imagine a rapidly developing search-and-rescue scenario in which rescue workers must memorize information about potential victims and spatial information about the scene
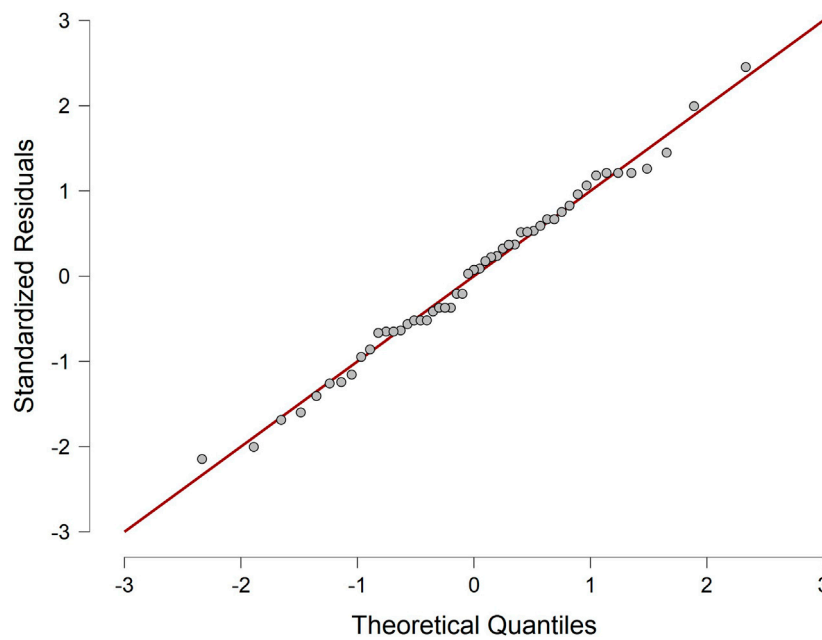
FIGURE 6 | Q-Q plot of the residuals ($n = 51$) from a two-way ANOVA test using memorizing and recall context-conditions as fixed factors and recall score as dependent variable. Visual inspection suggests that the residuals are normally distributed.

TABLE 5 | Outcomes of a two-way ANOVA test using recall score as dependent variable, with memorizing and recall context-conditions and their interaction (×) as model terms ($n = 51$; sub-sample sizes are reported in **Table 4**; degrees of freedom are one for each factor and 47 for the residuals).

Tested factor	ANOVA	Estimated effect size
Memorizing context	$p = 0.148$ ($F_{1,47} = 2.162$)	$\eta^2 = 0.033$
Recall context	$p = 0.111$ ($F_{1,47} = 2.645$)	$\eta^2 = 0.041$
Memorizing context × recall context	$p < 0.001^*$ ($F_{1,47} = 13.327$)	$\eta^2 = 0.205$

TABLE 6 | Comparison through independent samples Student's t-test (p-values with t-statistic and degrees of freedom reported) of mean recall scores between context-changed and context-unchanged sub-samples, separated by memory task.

Task	N	Same context Shapiro-Wilk test	Mean (SD) recall score	N	Changed context Shapiro-Wilk test	Mean (SD) recall score	T-test
Rote-learning	25	$p = 0.214$	0.473 (0.221)	26	$p = 0.133$	0.294 (0.225)	$p = 0.006^*$ ($t_{49} = 2.863$)
Association	25	$p = 0.074$	0.792 (0.144)	26	$p = 0.111$	0.715 (0.185)	$p = 0.106$ ($t_{49} = 1.646$)
Recognition	25	$p = 0.014^*$	0.868 (0.104)	26	$p = 0.003^*$	0.763 (0.214)	$p = 0.032^*$ ($t_{49} = 2.206$)

Mean recall scores per sub-sample are reported along with the standard deviation (SD). Significant outcomes of Shapiro-Wilk testing suggest that recall scores for the recognition task under both conditions deviate from normality.

at hand. It would be difficult to provide a real-world environmental context during memorization (R^M) that is similar to the real-world environment where rescue workers must rely on the accurate recall of information (R^R). However, a rough virtual approximation with low visual fidelity to be used for memorization (V^M) could potentially be automatically constructed from existing building plans. Using the notations of this paper, one should then not compare recall accuracy

between R + R and V + R, but rather between "$R^M + R^{R}$" and "$V^M + R^{R}$" where V^M corresponds more closely to R^R then R^M possibly could. In such a case, it would be expected that adverse context-dependency effects are stronger for $R^M + R^R$ than for $V^M + R^R$, making virtual reality the preferred memorization context.

Recall accuracy for the rote-learning task and the recognition task was significantly affected by a change of environmental

context, both in terms of effect size and statistical significance. However, the negative effect of context change on the association task recall accuracy was not strong enough to be statistically significant. In the association task, participants created self-generated content which was to be memorized. Perhaps the association task recall accuracy is influenced less by a change of environmental context because it is more focused on generating content than on memorizing given content. It is possible that self-generated memory is less susceptive to a context-dependency effect, a finding that may relate to effects of introspective thought found by Smith and Vela (2001). We can conclude that not all sub-forms of memory are equally sensitive to the effect, confirming prior findings such as by Goodwin et al. (1969).

A final matter worthy of further attention is that in this work we only considered short-term context-dependency effects on memory. One could easily imagine scenarios in which memory retention lasting more than 24 h is of importance. However, we did not take effects beyond a 24-h memory-recall period into account, which is something that further research could explore. Also, context-dependency effects caused by changing graphic and immersive qualities of the VR environment remain open for investigation. One could hypothesize that as VR immersion and realism approach that of the real world, that environmental context-dependent memory effects become less dominant, if not absent.

Despite the cautions and considerations raised here, it remains in our view a solid and interesting effect that was uncovered:

under the experimental conditions, a change from a real environment to virtual reality, or vice-versa, adversely affects our ability to recall items memorized after 24 h.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements at the time of the study. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTION

Research conceptualization and design: MHL and ML. Experiment construction and data collection: ML. Data analysis: MHL and ML. Writing: MHL and ML. All authors contributed to the article and approved the submitted version.

REFERENCES

Attree, E. A., Brooks, B. M., Rose, F. D., Andrews, T. K., Leadbetter, A. G., and Clifford, B. R. (1996). "Memory processes and virtual environments: I can't remember what was there, but I can remember how I got there. Implications for people with disabilities," in *Proceedings of the first European conference on disability, virtual reality and associated technologies*. Editor P. M. Sharkey (Reading, UK: University of Reading), 117–121.

Bower, G. H., Monteiro, K. P., and Gilligan, S. G. (1978). Emotional mood as a context for learning and recall. *J. Verbal Learn. Verbal Behav.* 17 (5), 573–585. doi:10.1016/S0022-5371(78)90348-1

Burke, D. M., Peters, L., and Harrold, R. M. (1987). Word association norms for young and older adults. *Social Behav. Sci. Documents* 17 (2).

Chomsky, N. (1956). Three models for the description of language. *IEEE Trans. Inform. Theor.* 2 (3), 113–124. doi:10.1109/TIT.1956.1056813

Dinh, H. Q., Walker, N., Hodges, L. F., Chang Song, C., and Kobayashi, A. (1999). Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. *Proc. IEEE Virtual Reality*, 222–228. doi:10.1109/VR.1999.756955

European Commission (2012). Descriptors defining levels in the European Qualifications framework (EQF)—learning opportunities and Qualifications in Europe. Available at: https://ec.europa.eu/ploteus/en/content/descriptorspage (Accessed March 19, 2018).

Godden, D. R., and Baddeley, A. D. (1975). Context-dependent memory in two natural environments: on land and underwater. *Br. J. Psychol.* 66 (3), 325–331. doi:10.1111/j.2044-8295.1975.tb01468.x

Goodwin, D. W., Powell, B., Bremer, D., Hoine, H., and Stern, J. (1969). Alcohol and recall: state-dependent effects in man. *Science* 163 (3873), 1358–1360. doi:10.1126/science.163.3873.1358

Gould, N. F., Holmes, M. K., Fantie, B. D., Luckenbaugh, D. A., Pine, D. S., Gould, T. D., et al. (2007). Performance on a virtual reality spatial memory navigation task in depressed patients. *Am. J. Psychiatry* 164 (3), 516–519. doi:10.1176/ajp.2007.164.3.516

Johnson, A. J., and Miles, C. (2008). Chewing gum and context-dependent memory: the independent roles of chewing gum and mint flavour. *Br. J. Psychol.* 99 (2), 293–306. doi:10.1348/000712607X228474

Kennedy, R. S., Stanney, K. M., and Dunlap, W. P. (2000). Duration and exposure to virtual environments: sickness curves during and across sessions. *Presence: Teleoperators Virtual Environ.* 9 (5), 463–472. doi:10.1162/105474600566952

Lanen, M. (2018). Virtual reality as a context for memorizing: it can make a difference. Master's thesis. Leiden (NL): Leiden Institute of Advanced Computer Science, Leiden University.

Lanen, M., and Lamers, M. H. (2018). "Context-dependent memory in real and virtual reality," in *Virtual reality and augmented reality, lecture notes in computer science 11162*. Editors P. Bourdot, S. Cobb, V. Interrante, H. Kato, and D. Stricker (Berlin, Germany: Springer), 177–189. doi:10.1007/978-3-030-01790-3_11

Mania, K., Troscianko, T., Hawkes, R., and Chalmers, A. (2003). Fidelity metrics for virtual environment simulations based on spatial memory awareness states. *Presence: Teleoperators Virtual Environ.* 12 (3), 296–310. doi:10.1162/105474603765879549

Marks, L. E., and Miller, G. A. (1964). The role of semantic and syntactic constraints in the memorization of English sentences. *J. Verbal Learn. Verbal Behav.* 3 (1), 1–5. doi:10.1016/S0022-5371(64)80052-9

Matheis, R. J., Schultheis, M. T., Tiersky, L. A., DeLuca, J., Millis, S. R., and Rizzo, A. (2007). Is learning and memory different in a virtual environment? *Clin. Neuropsychol.* 21 (1), 146–161. doi:10.1080/13854040601100668

Merhi, O., Faugloire, E., Flanagan, M., and Stoffregen, T. A. (2007). Motion sickness, console video games, and head-mounted displays. *Hum. Factors* 49 (5), 920–934. doi:10.1518/001872007X230262

PGA Group Consulting Psychologists (2018). Standardized survey classifications—Individuals. Available at: http://www.pgagroup.com/standardized-survey-classifications.html (Accessed March 19, 2018).

Plancher, G., Gyselinck, V., Nicolas, S., and Piolino, P. (2010). Age effect on components of episodic memory and feature binding: a virtual reality study. *Neuropsychology* 24 (3), 379–390. doi:10.1037/a0018680

Smith, S. A. (2019). Virtual reality in episodic memory research: a review. *Psychon. Bull. Rev.* 26, 1213–1237. doi:10.3758/s13423-019-01605-w

Smith, S. M., and Vela, E. (2001). Environmental context-dependent memory: a review and meta-analysis. *Psychon. Bull. Rev.* 8 (2), 203–220. doi:10.3758/BF03196157

Sweeney, S., Kersel, D., Morris, R. G., Manly, T., and Evans, J. J. (2010). The sensitivity of a virtual reality task to planning and prospective memory impairments: group differences and the efficacy of periodic alerts on performance. *Neuropsychol. Rehabil.* 20 (2), 239–263. doi:10.1080/09602010903080531