Harman, Joel, Brown, Ross, & Johnson, Daniel
(2016)
The role of immersion during situated memory recall within virtual worlds.
In
28th Australian Conference on Computer-Human Interaction (OzCHI 2016), 29 November - 2 December 2016, Launceston, Tas.

https://doi.org/10.1145/3010915.3010945

The Role of Immersion during Situated Memory Recall within Virtual Worlds

Joel Harman
Queensland University of Technology (QUT)
ja.harman@qut.edu.au

Ross Brown
Queensland University of Technology (QUT)
r.brown@qut.edu.au

Daniel Johnson
Queensland University of Technology (QUT)
dm.johnson@qut.edu.au

ABSTRACT

Eliciting accurate knowledge from individuals is a non-trivial challenge. In this paper, we present the evaluation of a virtual reality-based approach to knowledge elicitation informed by situated cognition theory. This approach places users into 3D virtual worlds representing real-world locations and asks them to recall information about tasks completed in those locations. Through an empirical A/B evaluation of 64 users, we investigate whether the situated context provided by these virtual environments is adequate to assist with memory recall, and explore whether the added immersion provided by a head-mounted display (HMD) may meaningfully improve user memory recall capability when compared with a desktop display experience. Results suggest that those provided with a HMD may be able to recall more information about a sequenced task than those provided with a desktop display.

Author Keywords

3D Virtual Worlds, Situated Cognition, Virtual Reality, Knowledge Elicitation, Human-computer Interaction

ACM Classification Keywords

H5.m. Information interfaces and presentation, HCI: Miscellaneous.

INTRODUCTION

The ability to accurately gather knowledge is a non-trivial challenge with relevance to a variety of domains (Davis et al., 2006). Furthermore, if this information is not gathered effectively, there is the potential for many complications to arise. For example, when attempting to formalise business practices, working with incorrect, or incomplete, information may lead to higher construction costs, longer development times or poor-quality products (Smith, 2001).

Numerous approaches have been developed which aim to elicit accurate information from stakeholders (Zowghi & Coulin, 2005; Pits & Browne, 2007). Most of these methods, however, produce a trade-off between the quality of the information gathered, and the time required to do so (Davis et al., 2006). Questionnaires can be administered quickly, but often yield inadequate information. Role-plays, on the other hand, have been shown to be an effective way of gathering accurate knowledge (Costain & McKenna, 2011), but are used less often than interviews in a business setting (Zowghi & Coulin, 2005). Davis et al. (2006) suggests that this may be due to their potentially high setup costs and issues related to multi-user co-location. Despite these detractors, we have chosen to employ a modified role-play approach in this study due to the benefits it affords when attempting to elicit accurate and complete information.

When performing a role-play, the individual steps in the task are performed in a set sequence. The concept of memory chaining proposes that we are able to recall a piece of information more effectively if we are already thinking about the task it precedes (Mace, Clevinger & Martin, 2010). This would suggest that even potentially mundane tasks, which would usually be ignored in other elicitation approaches (e.g. interviews), may initiate memory chains which could potentially yield relevant information.

In addition to the potential benefit provided by memory chains, we believe the rich context provided during a role-play elicitation session affords sufficient context to achieve a situated cognition response. Situated cognition is the theory that all knowledge is, to some extent, tied to the situations and contexts in which it was learned (Brown, Collins & Dugoid, 1989). For example, Carraher et al. (1985) found that several students were able to correctly perform arithmetic while selling produce on the street, but were unable to answer the same questions on a formal school test. In this example, it is believed the students were unable to adequately manage the information while in the test setting, because they were not situated within the context in which the information was learned. We believe that during a role-play session, the person is exposed to the necessary context, thereby allowing them to better recall and manage their knowledge of the task.

In our work, we explore the potential of a modified in-situ role-play and how it may be used to assist in obtaining quality information during elicitation. Rather than conducting these role-play sessions within the real world, these sessions are instead conducted in a virtual world which closely matches the real world location. This approach may be preferable when the costs, risks or other constraints of a real-world role-play (such as co-location) would make a real-world role-play infeasible. As this work aims to provide context to improve memory recall performance, it is constrained to the domain of episodic memory. Memory which is not directly tied to contexts or

OzCHI '16, November 29 - December 02, 2016, Launceston, TAS, Australia

ISBN 978-1-4503-4618-4/16/11... $15.00
DOI: http://dx.doi.org/10.1145/3010915.3010945

situations, such as our knowledge of mathematics, is unlikely to be meaningly affected by the role-play approach we have described. Furthermore, while most literature related to role-play aims to assist with training and providing information to participants, this work is instead looking at eliciting knowledge the participants have already acquired prior to their participation in the study.

In a prior study (Harman et. al., 2016), we have found that when comparing the differences in recall ability between participants provided with virtual world stimuli and participants that were not, those provided with the virtual world stimuli were able to recall a larger amount of higher quality information. This suggests that a virtual world may provide a sufficient degree of context to assist memory recall ability. Furthermore, conducting these role-play sessions within virtual worlds also affords other benefits. Risky and expensive tasks can be simulated at low cost, without the concerns of participant safety (Joyner & Younger, 2006). It also allows the role-play to be conducted while participants are spatially separated and affords many of the benefits associated with simulations, such as the ability to capture detailed logs and directly manipulate the environment.

In this study, we have continued to explore this recall phenomenon, and have considered whether the immersion and embodiment provided by the virtual world viewing interface may affect the ability of participants to recall information. Specifically, we have examined how the use of a head-mounted display (HMD) may change the way in which people are able to recall and articulate their knowledge. Mania & Chalmers (2001) have previously examined this concept in some detail, but this work aims to extend the memory tests they have conducted to also consider the concepts of sequencing and potential choices. While this work has been motivated by our prior work in the domain of business process management (Harman et. al., 2016), our investigations have uncovered intriguing general results, which we report in this paper, that have import for the CHI community.

SITUATED COGNITION AND VIRTUAL WORLDS

Applying the concept of situated cognition within virtual worlds has been extensively explored in literature. When situated cognition was first proposed by Brown et al. (1989), it considered the concept of situated learning, previously discussed by Lave & Wenger (1991), to be a related concept. Situated learning postulates that students need to be taught information within the contexts in which it will be applied.

Situated learning research has primarily considered virtual worlds for their potential applications to distance learning (Dickey et al., 2005; Dede et al., 2004). The goal of these environments is to place students within virtual classrooms, in order to better situate them within an environment they associate with a student-teacher learning dynamic. Dickey (2003) suggests that this approach may be an improvement over typical distance-learning communication methods, such as conference calls.

Outside of the context of learning, there has been an absence of literature which has explored the theory of situated cognition and the appropriateness of the context and situations provided by virtual worlds in achieving an appropriate response. There are some theoretical papers, however, such as Carassa et al. (2005) which have explored the idea of situated cognition within virtual worlds at a conceptual level. Carassa et al. (2005) discusses the appropriateness of existing measures, such as immersion and presence, and conjectures that presence may have an important role in achieving effective situated cognition responses within virtual worlds. As this study discusses the potential for assisting memory recall within virtual worlds via situated cognition, we have decided to explore the concept of presence and whether there exists the conjectured positive relationship between presence and recall.

RESEARCH QUESTIONS

In this study, we have looked to explore how differences in viewing immersion may affect participant behaviour and memory recall ability. This has led to the formation of two distinct research questions:

RQ1: How do changes in viewing immersion affect memory recall performance of participants when they are asked to describe information while role-playing within a related virtual world?

RQ2: How do changes in viewing immersion affect the behaviour of participants when they are asked to describe information while role-playing within a related virtual world?

We have chosen to explore potential behavioural differences as prior work suggests that there may be a relationship between participant behaviour and recall performance (Harman et al., 2016). To examine changes in viewing immersion, we will be comparing a virtual reality setup, using a HMD, with a standard desktop display. From these two research questions, we have constructed two main hypotheses by examining existing theory within related literature. The first hypothesis is:

H1: Users asked to recall information while viewing a related virtual world within a HMD will recall more information than those viewing the virtual world on a desktop display.

The theory behind this hypothesis has primarily been motivated by the psychological memory theory of situated cognition. For example, Lave & Wenger (1991) conjecture that students are best able to learn vocabulary within the context of a conversation, rather than in classroom lessons.

In prior work (Harman et al., 2016), we found preliminary evidence to suggest that the context provided by a virtual world was able to assist participants in recalling information. We conjecture, however, that while the virtual world provided some benefit, it did not, wholly, afford the necessary context to the user. Specifically, we believe that the embodiment and immersion afforded to the user by virtual reality will better situate the user and assist them in recalling information more effectively. In addition to recall improvements, we believe that the added embodiment and immersion provided by the HMD will also result in certain behavioural changes between the two groups. This has led to the formation of our second hypothesis:

H2: Users provided with the HMD will be more exploratory in their approach than those provided with the desktop display.

To better explore this hypothesis, we will be examining two sub-hypotheses. These are:

- H2(a): Users provided with the HMD will traverse a larger portion of the environment than those provided with the desktop display.
- H2(b): Users provided with the HMD will adjust their view within the virtual world more often than those provided with the desktop display.

While we do not necessarily believe that virtual reality innately makes a participant more exploratory, we believe that the interaction mechanisms it affords makes it easier for a participant to operate in a way in which they are familiar and experienced. Carassa et al. (2005) discuss how the evidence for increases in emotions (Riva et al., 2007), telepresence and sexual presence (Renaud, 2002) when using a HMD, rather than a desktop display, are all examples of how users appear to behave more naturally within virtual reality. We are particularly interested in behaviour in this study as our prior work has found in-world behaviour to have interesting correlations with recall performance (Harman et. al., 2016). For example, we found that participants who completed the task faster tended to recall more information about the task.

ARTEFACT DESIGN

Virtual worlds are synthetic environments which provide users with an avatar through which they can interact with other users, or the world itself, to perform various tasks (Duncan, Miller & Jiang, 2012). In this work, we have constructed a virtual world which aims to be representative of a real-world airport. An issue which is often discussed when using virtual worlds for research, however, is that these environments may be difficult to use without extensive training (Virvou & Katsionis, 2008). With this in mind, we have developed our own virtual world specifically to explore this phenomenon. The virtual world has been developed using the Unity3D (Unity Technologies, 2016) game engine, with the airport environment containing a mix of both modelled and pre-built assets. Furthermore, we have also previously explored usability issues associated with this virtual world by conducting semi-structured interviews with 24 participants who used the tool (Harman et. al., 2015), and made modifications to improve usability where possible. Figure 1 shows two screen captures taken from the developed virtual world.

Unlike many other virtual worlds, the developed environment does not change in response to participant actions. For example, if a user walks up to a virtual avatar, the avatar will not initiate communication. Furthermore, if a user describes placing luggage on a conveyer-belt, the belt will not start moving. This has been done as including these interactions would greatly increase the time required to develop these environments and would limit the practical feasibility of the approach. Ideally, we believe these environments could be constructed quickly via 3D scanning, similarly to the environments generated by a Matterport (Matterport, 2016) depth camera, allowing for minimal overhead in recreating this setup.

To navigate the virtual environment, participants from the HMD and desktop display treatments were both given an *Xbox 360* game controller. This controller had four main functions:

- Left joystick controlled avatar movement
- Right joystick controlled avatar view
- ‘A’ button represented *accept*
- ‘B’ button represented *cancel*

We decided to use a game controller for this study as it allowed both treatments to have very similar mechanisms for movement and viewing. It should be noted, however, that as a core mechanism of a HMD is to assist with viewing, the HMD was also responsible for controlling view changes in this condition, and the vertical component of view change on the controller joystick was disabled.

To describe each step in the task, participants would gaze at an associated item, or person, within the world. For example, if the user was trying to describe the task “*look at airport monitor for flight information*”, the participant would gaze at the monitor within the virtual world. This would generate a prompt and allow the participant to describe the task. In prior work, participants have entered these descriptions themselves with a keyboard. As participants using a HMD are unable to view their keyboard, however, this would make it difficult for them to enter this information if they were unfamiliar with the

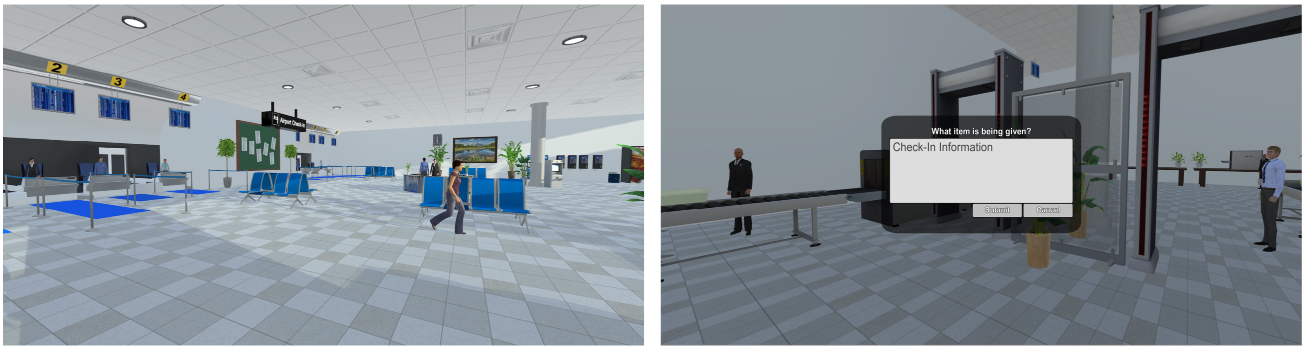

Figure 1: Screen captures of the developed virtual world.

keyboard layout. To describe this information in this study, participants instead dictated the desired descriptions and voice dictation software (Dragon NaturallySpeaking 14) would enter the corresponding text into the input field within the virtual environment. The participant could then choose to either accept the description entered, clear the textbox to change the description, or if necessary (e.g. the software was not capturing the text correctly), record their speech and have it manually transcribed after the experiment session. As both treatment conditions used the voice dictation software in this study, we believe the combination of random assignment and manual dictation if participants encountered issues sufficiently controlled for any difficulties participants may have had in dictating their descriptions.

The content in the virtual worlds given to those provided with the HMD and desktop display was identical. The only difference between the two treatments was the way in which they viewed the environment, and the way in which they modified the view of the avatar.

EXPERIMENT PROCEDURE

Participants who agreed to participate in the study were invited into a lab. They were then either presented with a 24" desktop monitor, or a HMD (Oculus Rift DK2) based on their random assignment. In addition to this, they were also provided with a game controller and microphone. A between-subject design was chosen as we believed that participation in one treatment may have meaningfully changed their descriptions in the subsequent treatment. They then underwent a short dictation session to calibrate the voice dictation software correctly and were given some basic training on how to navigate the environment and describe their actions. After this training, they were informed that they would be describing all of the steps involved when trying to board an airplane for a short domestic flight. Participants then used the virtual world tool to describe the task as best they could. After doing this, short semi-structured interviews were administered, followed by questionnaires measuring presence and usability.

Recall Measure

The recall measure is the main item of interest in this study. To measure recall, we will be comparing the number of tasks specified by the participants in the two groups. We have chosen to measure the number of tasks described as it allowed us to provide participants with some guidance during training as to what should constitute a task. During the training session, we had participants explain how they may go about submitting an application for leave to their supervisor. To help guide them, we gave them three examples of potential tasks: *I fill in my leave request*, *I e-mail the leave request to my supervisor* and *I read the leave request response*. This aimed to assist participants when considering what may be appropriate to describe.

We have not examined the individual tasks provided by the participants for correctness, but we have removed tasks which were immediately identified as erroneous (e.g. the same task being described multiple times in a row or a participant not providing a description for the task). As such, this paper is considering the quantity, rather than the correctness, of the information recalled. Finally, it should be noted that this task, while exploring recall, is technically measuring task *performance*. We believe that this is valid, however, as memory tests often require the participants to constrain the information they recall to the requirements of a task (such as a quiz) (e.g. Roediger & Karpicke, 2006; Ivanoiu et al., 2005).

In our previous work (Harman et al., 2016), we aimed to normalise knowledge of the task by providing the participants with a description of the tasks required beforehand. While this did reduce the variance in participant responses, it also had considerably less ecological validity when considering the overall goal of the approach. Rather than measure long-term memory recall performance, this previous study instead tested short-term recall ability, as participants were instead focused on recalling the description they were given prior to the experiment session. In this study, we have chosen to explore long-term memory recall performance by providing participants with no description of the individual tasks steps prior to the study. Knowledge of the task has instead been controlled for by asking participants to describe both the number of times they had been on a plane in the last five years and rate their subjective knowledge of the airplane boarding task.

Exploration

We have chosen to examine how the users traversed the environment as we believe that this may provide insight into the behavioural approach taken by the participants. This is important as we believe differences in behaviour may influence recall performance. To do this, we have split the virtual environment into 261 5m x 5m segments (approximately equivalent to real-world units). For this analysis, we will be comparing the average number of segments traversed by the two groups. In addition to this, we will also be examining movement heat maps to better explore any potential differences in how the environment was traversed by the two groups.

Change of View

We will be exploring change of view in order to better understand how the two treatment groups examined the virtual environment. This paper was primarily motivated by the concept that inserting a user into a situation with the appropriate context would assist them in recalling information related to that situation. For this to be an effective approach, however, the user must view and understand this context. For this reason, we conjecture that, to some degree, viewing more of the environment may result in participants recalling more information about the task.

In this experiment, participants given the desktop display adjusted their view with their game controller, while the participants given the HMD modified their view by adjusting their head. To account for noise due to slight, unintentional, head movements for those given the HMD, we have not considered any movements of less than two degrees in a single second. This cut-off value was chosen as it was able to eliminate the small, continuous, head movements for all participants, while only minimally affecting the less frequent, larger head movements.

In this study, we will be considering view change as the *average change of vertical viewing angle per minute*. We have chosen to use a measure which includes only the *vertical* component of this movement as participants, in addition to making view changes to examine the environment, also needed to modify their *horizontal* viewing angle in order to turn their avatar. This meant that horizontal view changes did not necessarily indicate that a user was actively attempting to adjust their view to in order examine the environment, but instead that they may have simply needed to turn their avatar to get around corners or avoid other items in the virtual world which may have blocked their path. In contrast, all vertical view changes were performed solely to explore the environment, and were not tied to movement in any way (participants had no way to move their avatar up or down and remained on the floor at all times). As vertical changes were exclusively performed in order to explore the environment, we believe that vertical view change is a more accurate measure than total view change for this study.

Time Taken

We have chosen to examine the time taken by both groups as it has been found to yield interesting results in our prior work (Harman et. al., 2016). This was largely consistent with prior elicitation work within the domain of business process modelling, which found that people who completed the task faster tended to also perform better (Claes et al., 2012).

Presence

Presence, as related to immersive virtual reality, is considered to be the concept of transportation. Schumie et al. (2001), explain that people are considered present when they report a sensation of being, to some degree, in the virtual world (e.g. "you are there"). We have chosen to measure presence in this study as Carassa et al. (2005) discuss the possibility of a link between presence and recall performance. In this study, presence has been measured by administering the Witmer and Singer (1998) presence questionnaire. This questionnaire was chosen despite some criticism (Schubert, Friedmann & Regenbrecht, 2001), as it is still the most widely administered survey for measuring presence.

Usability

We have considered usability as the HMDs have been known to suffer from challenges with usability among novice users (Sutcliffe & Kaur, 2000). If a user has difficulty using the tool, this may have ramifications on recall performance and other aspects of the session. Usability has been measured by administering the IBM usability satisfaction survey (Lewis, 1995).

RESULTS

In this study, participants were randomly assigned to either the HMD or desktop display conditions, with each condition having 32 participants. When looking to compare measures between the HMD and desktop display conditions, we will be using two-tail Student's *t* tests when the associated data is both continuous and passes the Shapiro-Wilk W normality test ($p < 0.05$). If either of these conditions is not met, a two-tail Mann-Whitney U test will be used instead. When examining the results for possible correlations, we will be using Pearson product-moment correlation coefficients.

The average age of this cohort was quite young, at 22.44 (SD = 4.22). No significant difference was found between the age of those given the HMD (M = 21.88, SD = 3.21) and the desktop display (M = 23.00, SD = 5.03, U = 473, z = -0.52, $p > 0.05$). 23 males and 9 females were assigned to the HMD group and 25 males and 7 females were assigned to the desktop display group. Perceived understanding of the boarding process was quite high, with an average response of 5.61 (SD = 1.41) on a 7-point Likert scale. No statistical difference in perceived understanding of the boarding task was found between those given the HMD (M = 5.89, SD = 1.01) and those given the desktop display (M = 5.81, SD = 1.26, U = 492, z = 0.26, $p > 0.05$).

Given the exploratory nature of the research and relatively small sample size we did not apply a Type I error correction (e.g., Bonferonni) to our analysis. Instead we have elected to provide effect size calculations (Cohen's d) for all findings to provide the reader a sense of the magnitude of all findings. Following Rosenthal (1994), we have converted the r values obtained from the Mann-Whitney U tests into Cohen's d effect sizes to provide a consistent effect size estimate across the two statistical tests. Following Cohen (1992), we treat effect sizes of between 0.2 and 0.5 as *small*, 0.5 and 0.8 as *medium* and greater than 0.8 as *large*. Due to the increased possibility of a Type I error, the findings reported in this study should be interpreted with a greater degree of caution.

Recall

In this study, we were primarily interested in investigating whether the added immersion provided by viewing the environment on a HMD, rather than a desktop display, may improve memory recall ability of participants. To evaluate H1, we compared the total average of tasks specified by both treatment groups. Results found that participants given the HMD were able to recall a larger number of steps (M = 18.06, SD = 7.02) than those given the desktop display (M = 13.91, SD = 5.91, U = 345, z = 2.24, $p < 0.05$, $d = 0.64$).

Exploration

To evaluate H2(a), we have compared the mean amount of the environment traversed by the two groups. Results from this test found that participants given the HMD traversed a larger number of segments (M = 86.98, SD = 21.44) than participants given the desktop display (M = 71.25, SD = 9.50, $t(62) = 3.79$, $p < 0.001$, $d = 0.95$).

To better understand how the users traversed the environment, and the sections of interest to participants, we have also examined heat maps containing aggregated user movements of the two groups. Figure 2 contains heat maps indicating a portion of the environment with a noteworthy difference in traversal. The baggage collection area (an area not necessary for the boarding task the participants were asked to describe) was traversed by nineteen participants from the HMD condition, while only four participants given the desktop display chose to do so. This suggests that there was indeed a noteworthy difference in the way the participants from both conditions chose to explore the world.

Change of View

To evaluate Hypothesis 2(b), we have measured the average amount participants adjusted the vertical view of their avatar. This found that participants given the HMD adjusted their view *vertically* more often per minute (M = 62.41, SD = 19.80) than those given the desktop display (M = 16.09, SD = 7.33, $t(62) = 12.41$, $p < 0.0001$, $d = 3.10$).

Presence

Participants reported *lower* presence in the HMD condition (M = 4.27, SD = 0.69) than those in the desktop display condition (M = 4.70, SD = 0.40, U = 299, z = -2.85, $p < 0.01$, $d = 0.76$).

Usability

Participants reported lower usability in the HMD condition (M = 5.22, SD = 0.41) than those in the desktop display condition (M = 5.93, SD = 0.57, U = 133.5, z = -5.08, $p < 0.001$, $d = 1.64$).

Time

Participants took longer to complete the task in the HMD condition (M = 1073.16, SD = 634.89) than in the desktop display condition (M = 635.81, SD = 341.14, U = 286, z = 3.03, $p < 0.01$, $d = 0.81$). A moderate positive correlation was found between time taken and the number of tasks participants were able to recall (Pearson's $r = 0.65$, $p < 0.0001$).

Interview Responses

In addition to the quantitative tests performed above, we also conducted interviews with participants taking between 5 and 15 minutes each. A semi-structured interview approach was followed whereby the interviewer drew on pre-determined questions but followed up on points of interest as appropriate. These interviews were optional, with 38 participants taking part in total (17 HMD, 21 desktop display). The interview questions investigated the topics of performance self-evaluation, user satisfaction, subjective presence and tool usability. The interviewer then subsequently grouped similar question responses together, with common responses then being developed into emerging themes. While several themes emerged from this analysis, in this paper we will only be exploring the themes related to subjective memory performance and environment believability. While many of the other themes identified, such as those related to virtual reality reception and tool usability, are interesting for future work, they are not central to the core discussion of memory recall performance presented in this paper.

The first theme identified was the theme of *convincing environment*. This theme was identified by 11 participants given the HMD and 5 participants given the desktop display. The following excerpts are the responses given when participants were asked about how they felt about the environment they were put into:

> *Participant 30 (HMD): "The environment felt very real. There were a couple times I forgot I wasn't really there. I even tried to push open a door once… It makes you feel silly because you're really just sitting in a room."*

> *Participant 61 (HMD): "It was really cool. The headset makes it all seem so real… getting to look around and move your head. I'm not used to that."*

> *Participant 20 (Desktop Display): "I felt like I was really there. I even had to wait in line for a bit. Because that's what I do every time I go to the airport. I did the exact same as I would do at an airport."*

The higher prevalence of this theme in the HMD condition is especially interesting, as the environment presented to the two groups was identical. They viewed the same people and objects, and then described the same task. The only difference was the way in which the two groups viewed the environment.

The second theme we identified was the antithetical response to the theme of environment authenticity discussed above. Specifically, this was the theme of *unconvincing environment*. This theme was identified by 2 participants given the HMD, but by 11 participants given the desktop display). The following are also excerpts from interviews conducted when participants were asked how they felt about the environment they were put into:

> *Participant 5 (Desktop Display): "You don't really get a response from any of the people in there [the virtual environment]. You have to do everything yourself. At security they would talk to me first."*

Figure 2: Heat maps of avatar movements for those given the HMD (left) and desktop display (right). Participants given the HMD appear to have traversed the baggage collection area more often than those given the desktop display.

Participant 12 (Desktop Display): "The world felt dead. Whenever I did something, nothing happened. There are usually a lot more people in an airport too."

Participant 50 (HMD): "It [the virtual world] was strange. It felt empty. My hands are usually full with my bags and stuff. It… didn't have that."

The third theme identified was the theme of *positive self-evaluation*. Specifically, 28 participants (11 HMD, 17 desktop display) responded positively when they were asked to comment on how well they believed they performed the given task. The following excerpts are the responses given when the participants were asked to describe how well they believed they did:

Participant 4 (Desktop Display): "I think I did well. I knew what to do, I interacted with all the people like I normally would."

Participant 11 (HMD): "Yes… I think I did a good job of explaining everything. I've done it so much… the number of times I've done it made it so easy."

Participant 22 (Desktop Display): "I think I did it [the task] really well. I would be surprised if I forgot something."

From these 28 participants, we found that 20 (8 HMD, 12 desktop display) reporting a positive self-evaluation had an objective recall score below the median. This is interesting, as it suggests that participants were unable to subjectively evaluate their performance correctly. This is consistent with the findings of Dunning and Kruger (1999), which found that relatively unskilled people subjectively rated their performance ability much higher than it was objectively.

DISCUSSION

The primary aim of this study was to determine whether improvements to interface immersion could directly translate into better episodic memory recall. An interface using the HMD was chosen over other immersive improvements, such as scene level of detail, viewing perspective or the inclusion of audio, as we believed it was likely to show the largest effect. Furthermore, as virtual reality has generally been associated with higher levels of presence and Carassa et al. (2004) provided a theoretical basis for the importance of presence for memory recall within virtual worlds, we decided virtual reality would be an appropriate area to explore.

This work has explored how we may better improve memory recall in order to assist with eliciting knowledge from experts. This is important as accurately eliciting expert knowledge is a requirement of many domains (Davis et al., 2006). In our qualitative analysis, however, our third theme shows some preliminary indicators that participants may not only omit tasks during elicitation, but also fail to adequately rate their performance. If a participant is unable to effectively determine the quality of what they have described, it is unlikely that others who are less familiar with the content will be likely to immediately notice errors or omissions immediately either. For example, in a business setting, this may lead to a product only being discovered defective after it has been developed, due to inadequate specifications being provided during requirements elicitation. This highlights the benefit of exploring the intricacies of eliciting information. If people are unable to articulate their knowledge correctly, exploring an elicitation approach which best assists users where possible will help to limit future complications.

The results of this experiment supported H1 and identified that participants given the HMD had a statistically significant improvement (with a medium effect size) in episodic memory recall over participants given the desktop display. While this finding does suggest that immersive systems may be better able to assist users with episodic memory recall, we have also examined several other items of interest which aim to better explain the observed result.

In addition to looking at memory performance, our second research question aimed to explore potential behavioral differences between the two groups. To do this, we looked at both the way the two groups traversed the environment, and the way the two groups viewed the environment.

When examining how the two groups traversed the environment, we found that participants given the HMD explored more of the environment than those given the desktop display (supported by a large effect size). This result supports H2(a) and provides some evidence that the participants given the HMD were more exploratory in their approach. To better understand this, we examined movement heat maps of the two groups. This revealed that participants given the HMD traversed the baggage collection area (an area not required for the task) much more often than participants given the desktop display. While this is interesting, with the data available it is not possible to adequately determine why participants chose to explore this area. Further research is required to better explore this phenomenon and how it may have affected the observed results.

When examining how the two groups viewed the environment, we found that participants given the HMD also adjusted their view more often than those given the desktop display. This finding supports H2(b) and provides further indication that the HMD participants were more exploratory in their approach and were potentially exposed to greater degree of context than the participants viewing the environment on the desktop display. While we cannot say, with any degree of certainty, that these differences in behavior resulted in better recall performance, these findings are consistent with existing literature which suggests that providing participants with adequate levels of context is important when attempting to achieve a situated cognition effect (Brown et al., 1989).

In this study the participants given the HMD reported lower subjective presence scores than those using given the desktop display. This is an interesting finding as it is *not* consistent with most prior studies examining presence within virtual reality (e.g., Schuemie, 2001; Sanchez-Vives & Slater, 2005; Lorenz et al., 2015). Furthermore, this result runs contrary to the theoretical understanding of presence, situated cognition and virtual worlds proposed by Cassara et al. (2005). When examining the qualitative results of this study, however, many participants given the HMD described the environment as being quite convincing. As the believability of the environment is considered to be a core component of presence (Witmer & Singer, 1994), we believe this provides some indication that those given the HMD were experiencing some degree of presence. Despite this, the subjective scores for presence reported by participants were quite low. This suggests the possibility that the questionnaire we chose to subjectively evaluate presence was inappropriate for the task given.

The Witmer & Singer (1994) presence questionnaire was chosen for this study despite some criticism (Schubert, Friedmann & Regenbrecht, 2001) as it is, by far, the most commonly used presence questionnaire. Upon consideration of results, however, the role-play task the participants were asked to complete may have been too different from the original intent of the survey. Specifically, the limited implementation of certain aspects of the virtual environment, such as the inclusion of no environmental interactions or object movements may have negatively impacted our presence results in certain areas of the survey (e.g. *How natural did your interactions with the environment seem?* and *How compelling was your sense of objects moving through space?*). We believe that these differences in virtual environment design and task content may have contributed to the unexpected values of subjective presence reported by those given the HMD. Finally, existing research also suggests that HMD characteristics may also meaningfully affect performance (Lin et. al., 2002). If a more recent headset, such as the Oculus Rift CV1, was used instead, the observed presence scores may have been higher for the HMD participants. Due to the conflicting responses provided by both the questionnaires and interviews, we are unable to say with certainty what sort of role presence may have played in the results of this study. A more thorough exploration of presence and how it may assist in the memory recall process will need to be conducted in future work.

This study also found that participants given the HMD had a much more difficult time operating the virtual world (supported by a large effect size) despite all aspects of the environment being identical for both treatments with the exception of the viewing interface. As many of the participants were using the HMD for the first time, it appears likely that the use of a HMD may have overwhelmed users, as they were not able to look down at their controller to check for button positions. While it is not clear how this result may have affected the other results discussed in this paper, it is also possible that the poor subjective presence scores may have been partially due to poor usability scores. For example, the presence questionnaire item "*How much did the control devices interfere with the performance of assigned tasks or with other activities?*" received a poor score, potentially because of difficulties experienced with tool usability.

Finally, time taken was considered as an item of interest this study. This showed that participants given the HMD took more time to describe the task than those given the desktop display (supported by a large effect size). Furthermore, a medium positive correlation was found between time taken to describe the task and the number of task steps specified. The size of this correlation meant it was not possible for us to control for time taken and still expect to see meaningful variation in recall. This is somewhat problematic, as to our knowledge, no prior studies have ever investigated the effect of viewing time on a situated recall performance. It is possible that by spending more time in the environment, participants given the HMDs were able to subsequently recall more information about the task over time. With the data available in this study, it is not possible for us to determine, with certainty, how much this has affected recall performance scores. There are other factors, however, which may account for this difference in time taken between the two groups. As participants were describing tasks while traversing the environment, it would be expected that a person describing more information about the task would take longer to finish. As a key component of situated cognition is reliant on in situ stimulus, however, we were unable to separate the airport viewing and recall phases of the experiment. Furthermore, as participants given the HMD reported lower usability scores, it may suggest that they would have required longer to become familiar with the tool and complete the task. For these reasons, we believe that despite the difference in time taken between the two groups, it may not necessarily be indicative of a causative link between time taken and recall performance.

CONCLUSION AND FUTURE WORK

In this study, we have discussed the potential of using a virtual world role-play approach in order to assist in eliciting knowledge. This approach was motivated by the theory of situated cognition, which postures that by situating a user within a specific context, it becomes easier for them to recall related information. Specifically, this study has explored whether the added immersion and embodiment afforded by a HMD may further improve the recall performance of users.

The results of this study indicated that participants using the HMD described a noteworthy amount more tasks (supported by a medium effect size) than those using the desktop display. When comparing behavior, we found that participants given the HMD also tended to traverse more of the environment and modify their view more often (supported by large effect sizes). These findings supported each of our initial hypotheses and provide some evidence as to the efficacy of using a HMD for elicitation, rather than a desktop display.

Contrary to many prior studies, we found that participants reported lower presence within the HMD. In contrast, however, our qualitative analysis identified themes which we believe to be representative of presence. For this reason, we cannot say with certainty what levels of presence were experienced by participants. Future work will need to be conducted in order to better explore the relationship between presence and memory recall performance within this context.

REFERENCES

Brown, J., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational researcher*, *18*(1), 32-42.

Carassa, A., Morganti, F. and Tirassa, M. (2005). A Situated Cognition Perspective on Presence. In *Proc. CogSci 2005*, Sheridan Printing, pp. 384–389.

Carraher, T., Carraher, D., & Schliemann, A. (1985). Mathematics in the streets and in schools. *British journal of developmental psychology*, *3*(1), 21-29.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.

Claes, J., Vanderfeesten, I., Reijers, H. A., Pinggera, J., Weidlich, M., Zugal, S., ... & Poels, G. (2012). Tying process model quality to the modeling process: the impact of structuring, movement, and speed. In *BPM 2012* (pp. 33-48). Springer Berlin Heidelberg.

Costain, G., & McKenna, B. (2011). Experiencing the elicitation of user requirements and recording them in use case diagrams through role-play. *Journal of Information Systems Education*, *22*(4), 367.

Davis, A., Dieste, O., Hickey, A., Juristo, N., & Moreno, A. M. (2006, September). Effectiveness of requirements elicitation techniques: Empirical results derived from a systematic review. In *Proc. RE 2006* (pp. 179-188). IEEE.

Dede, C., Nelson, B., Ketelhut, D. J., Clarke, J., & Bowman, C. (2004, June). Design-based research strategies for studying situated learning in a multi-user virtual environment. In *Proc. ICLS 2004* (pp. 158-165). International Society of the learning Sciences.

Dickey, M. D. (2003). Teaching in 3D: Pedagogical affordances and constraints of 3D virtual worlds for synchronous distance learning. *Distance education*, *24*(1), 105-121.

Dickey, M. D. (2005). Three-dimensional virtual worlds and distance learning: two case studies of Active Worlds as a medium for distance education. *British journal of educational technology*, *36*(3), 439-451.

Duncan, I., Miller, A., & Jiang, S. (2012). A taxonomy of virtual worlds usage in education. *British Journal of Educational Technology*, *43*(6), 949-964.

Harman, J., Brown, R., Johnson, D., Rinderle-Ma, S., & Kannengiesser, U. (2016). Augmenting process elicitation with visual priming: An empirical exploration of user behaviour and modelling outcomes. *Information Systems*.

Harman, J., Brown, R., Kannengiesser, U., Meyer, N., & Rothschädl, T. (2015). Model as you do: engaging an S-BPM vendor on process modelling in 3D virtual worlds. In *S-BPM in the Wild* (pp. 113-133). Springer International Publishing.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, *77*(6), 1121.

Ivanoiu, A., Adam, S., Van der Linden, M., Salmon, E., Juillerat, A. C., Mulligan, R., & Seron, X. (2005). Memory evaluation with a new cued recall test in patients with mild cognitive impairment and Alzheimer's disease. *Journal of neurology*, *252*(1), 47-55.

Joyner, B., & Young, L. (2006). Teaching medical students using role play: twelve tips for successful role plays. *Medical teacher, 28*(3), 225-229.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, *7*(1), 57-78.

Lin, J. W., Duh, H. B. L., Parker, D. E., Abi-Rached, H., & Furness, T. A. (2002). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Virtual Reality*, 2002. Proceedings. IEEE (pp. 164-171). IEEE.

Lorenz, M., Busch, M., Rentzos, L., Tscheligi, M., Klimant, P., & Fröhlich, P. (2015, March). I'm There! The influence of virtual reality and mixed reality environments combined with two different navigation methods on presence. In *Proc. VR 2015* (pp. 223-224). IEEE.

Mace, J. H., Clevinger, A. M., & Martin, C. (2010). Involuntary memory chaining versus event cueing: Which is a better indicator of autobiographical memory organisation?. *Memory*, *18*(8), 845-854.

Mania, K., & Chalmers, A. (2001). The effects of levels of immersion on memory and presence in virtual environments: A reality centered approach. *CyberPsychology & Behavior*, *4*(2), 247-264.

Matterport. (2016). *Immersive Spaces for Real-world Applications - Matterport*. [online] Available at: http://matterport.com [Accessed 23 Jun. 2016].

Renaud, P., Rouleau, J. L., Granger, L., Barsetti, I., & Bouchard, S. (2002). Measuring sexual preferences in virtual reality: A pilot study. *CyberPsychology & Behavior*, *5*(1), 1-9.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning taking memory tests improves long-term retention. *Psychological science*, *17*(3), 249-255.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231-244). New York: Russell Sage Foundation.

Unity Technologies. (2016). *Unity - Game Engine*. [online] Available at: http://unity3d.com [Accessed 13 Jun. 2016].

Virvou, M., & Katsionis, G. (2008). On the usability and likeability of virtual reality games for education: The case of VR-ENGAGE. *Computers & Education*, *50*(1), 154-178.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and virtual environments*, *7*(3), 225-240.

Sanchez-Vives, M., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, *6*(4), 332-339.

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence*, *10*(3), 266-281.

Schuemie, M. J., Van Der Straaten, P., Krijn, M., & Van Der Mast, C. A. (2001). Research on presence in virtual reality: A survey. *CyberPsychology & Behavior*, *4*(2), 183-201.

Slater, M., McCarthy, J., & Maringelli, F. (1998). The influence of body movement on subjective presence in virtual environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *40*(3), 469-477.

Smith, E. A. (2001). The role of tacit and explicit knowledge in the workplace. *Journal of knowledge Management*, *5*(4), 311-321.

Sutcliffe, A. G., & Kaur, K. D. (2000). Evaluating the usability of virtual reality user interfaces. *Behaviour & Information Technology*, *19*(6), 415-426.

Zowghi, D., & Coulin, C. (2005). Requirements elicitation: A survey of techniques, approaches, and tools. In *Proc. EMSR 2005* (pp. 19-46). Springer Berlin Heidelberg.