
# Event based prospective remembering in a virtual world


| | |
|---|---|
| Journal: | *Quarterly Journal of Experimental Psychology* |
| Manuscript ID: | QJE-STD 10-287.R1 |
| Manuscript Type: | Standard Article |
| Date Submitted by the Author: | 15-Feb-2011 |
| Complete List of Authors: | Trawley, Steven; University of Edinburgh, Psychology<br>Law, Anna; Liverpool John Moores University, Psychology<br>Logie, Robert; University of Edinburgh, Psychology |
| Keywords: | prospective memory, planning, virtual reality, working memory |
| | |

Title Page

# Event based prospective remembering in a virtual world

Steven L. Trawley[1], Anna S. Law[2] and Robert H. Logie[1]

[1] Human Cognitive Neuroscience, University of Edinburgh, UK

[2]Department of Psychology, Liverpool John Moores University, UK

Correspondence can be directed to any of the authors (e-mails: strawley@staffmail.ed.ac.uk, a.law@ljmu.ac.uk, rlogie@staffmail.ed.ac.uk).

Acknowledgements

We are grateful to Prof Jennifer Rusted for her helpful suggestions and comments on the manuscript and to Matthew Logie for undertaking the programming to create the EVET environment. We also are grateful to Fergus Craik for permission to use the Craik and Bialystok (2006) breakfast task in our research and to their programmer Perry Tohn for providing a copy of the programme and for help with its implementation in our laboratory. We acknowledge Leverhulme Trust research grant number F/00 158/W which supported the research that we report here.

Abstract

Most laboratory-based prospective memory (PM) paradigms pose problems that are very different from those encountered in the real world. Several PM studies have reported conflicting results when comparing laboratory with naturalistic based studies (e.g., Bailey, Henry, Rendell, Phillips & Kliegel, 2010). One key contrast is that for the former, how and when the PM cue is encountered typically is determined by the experimenter, whereas in the latter case, cue availability is determined by participant actions. However, participant-driven access to the cue has not been examined in laboratory studies focused on healthy young adults, and its relationship with planned intentions is poorly understood. Here we report a study of PM performance in a controlled, laboratory setting, but with participant-driven actions leading to the availability of the PM cue. This uses a novel PM methodology based upon analysis of participant movements as they attempted a series of errands in a large virtual building on the computer screen. A PM failure was identified as a situation in which a participant entered and exited the "cue" area outside an errand related room without performing the required errand whilst still successfully remembering that errand post-test. Additional individual difference measures assessed retrospective and working memory capacity, planning ability and PM. Multiple regression analysis showed that the independent measures of verbal working memory span, planning ability and PM were significant predictors of PM failure. Correlational analyses with measures of planning suggest that sticking with an original plan (good or bad) is related to better overall PM performance.

Introduction

Our ability to remember to perform an activity at a specific future time or place is known as prospective memory (PM). As such intentions can only be realized at a later time, and as subsequent tasks demand our attention, we typically encode these intentions in memory and then "forget" them until the appropriate situation arises. An individual must frequently recall an intention when there is no explicit reminder to prompt them. For example, imagine a man driving home who suddenly realizes, shortly before driving past a supermarket, that he had intended earlier that day to buy a comic for his daughter; appearing relieved, he quickly decreases his speed and pulls over. Laboratory based PM research, following the standard Einstein and McDaniel (1990; e.g. McDaniel and Scullin, 2010) paradigm, attempts to mimic such situations by requiring participants to perform both an ongoing task, such as lexical decision, and a concurrent 'background' PM task that requires the participant to make a specified response to a particular target embedded in the ongoing task (e.g., during a lexical decision task press the space bar if you see an example of a fruit).

Although very successful, this approach has never been used to explore performance in situations where either the timing or appearance of the PM cue can be influenced by the participant. Furthermore, with abstract stimuli divorced from any situational or social context, the anticipated prospective event is essentially something that the volunteers will never have encountered before. For instance, forming a PM to purchase a comic would require specific information about the retrieval cue (namely the supermarket) to indicate the moment at which the intention should be realized, such as its relative location (left/right roadside) and context (e.g., light/heavy traffic), both of which are affected by what route the driver actually travels. By taking a different route home the driver may encounter the supermarket from a different perspective than had been imagined when forming the intention. There is a body of work that has examined this discrepancy between initial cue encoding and what is perceived at retrieval. Several studies (e.g.Cook, Marsh & Hicks, 2005; Logie & Maylor, 2009; Maylor & Logie, 2010; Nowinski & Dismukes, 2005) have manipulated context via the initial instructions given to the participants. These papers agreed in their conclusion that the probability of successful cue detection is affected by how information is processed at encoding and subsequently interacts with the perceived PM cue at the point of retrieval. In situations where the disparity between encoding and retrieval was high, performance was always impaired in these studies (see also, Ellis & Milne 1996; McDaniel, Robinson-Riegler& Einstein, 1998). However, as far as we are aware, no laboratory PM study has explored how changes initiated by the participant could induce an encoding/retrieval PM disparity on cue appearance. All manipulations to date have been experimenter driven, with the timing and appearance of the PM cue insensitive to the actions of the participant as they perform the ongoing task.

Our primary aim, therefore, was to investigate the impact of participant-driven actions on PM performance and to do so by highlighting the relationship between planning and successful PM. This relationship was explored by Kliegel, Martin, McDaniel and Einstein (2000) who showed the importance of plan elaboration and plan following on successful PM performance. Participants could change cue presentation time by performing tasks in a different order than originally planned. However, the tasks used by Kliegel et al. (2000) were always located on a table in front of the participant. Clearly the order in which participants performed the tasks would not affect the relative location and appearance of the cue, whereas participant-driven task order is typical of many real life PM scenarios. The new approach that we adopt in the present study contrasts with conventional PM methods, in that context manipulations are generated by the participant, not by the experimenter. Differential predictions can be made depending upon whether or not one assumes that the pre-test plan generated by the participant creates a contextual relationship between each errand and its related information such as expected cue appearance from a given viewpoint based upon task order. If there is such a contextual relationship, we would expect participants who adhere closely to their plan will exhibit fewer PM errors than those who do not. If there is no such relationship, and participants have only a loose order planned, then spontaneous changes to actual completion of the task order in response to PM cues when they happen to appear should have little negative effect on PM errors. In this case, performance may even be better because the PM cue prompts enactment of an intention at the time the cue is encountered rather than the participant performing the tasks in the planned order regardless of when they encounter each cue.

A second aim was to explore the relationship between working memory and successful PM performance. The role of working memory in PM has typically been examined either by making the ongoing task harder or by giving participants an additional task to perform concurrently. For example, Marsh and Hicks (1998) conducted several experiments showing that only tasks that placed a demand on the central executive adversely affected PM. Moreover, several studies since have highlighted the relationship between individual differences of working memory capacity and PM (Brewer, Knight, Marsh & Unsworth, 2010; Einstein, McDaniel, Manzi, Cochran & Baker, 2000; Smith, 2003; Smith & Bayen, 2005; West & Craik, 2001). However, all of these studies have used verbal working memory tasks as their estimator of individual working memory capacity. To the authors' knowledge, this is the first time PM performance has been explored from a domain specific working memory process perspective (Baddeley & Logie, 1999; Logie & Baddeley, in press) by indexing both verbal and visuo-spatial working memory capacity as predictors in a regression model.

Independent measures of retrospective memory, planning and PM were also used as predictors in the regression model.

As noted, in most laboratory paradigms for studying PM, cue presentation is predefined by the experimenter. A range of studies have used more naturalistic settings, many of which have focused on the age-prospective memory paradox in which older people appear to outperform younger people on PM tasks in the naturalistic setting but not in a laboratory setting (e.g. Bailey et al., 2010; Rendell & Craik, 2000). In these settings, the participant's actions do determine when and how a PM cue is encountered. However, genuine naturalistic settings are very complex and lack experimental control, so results may be driven by factors of which the experimenter is not aware or cannot influence. Realistic scenarios in the laboratory have been explored using video recordings of real world scenes (e.g. Farimond, Knight & Titov, 2006), or laboratory simulations (e.g. Rendell & Craik, 2000; Craik & Bialystok, 2006). However, the Farimond et al. (2006) simulated shopping task lacks an ongoing task and the authors acknowledge that limitation. Although a study by Kinsella, Ong and Tucker (2009) specifically addressed this limitation by asking participants to monitor the shopping video for “specials offers” while performing their virtual shop, all of these paradigms restrict when cues are encountered and/or the order in which participants perform actions. Therefore, a third aim was to introduce a novel PM methodology in a controlled, laboratory setting but where cue presentation is determined by the movement sequences chosen by the participant as they undertake a range of tasks. As such, the relationship between encoding and retrieval can be disrupted, virtually step by step, by the choices made by the participant in the intervening retention phase. Our approach is based upon analysis of the route the participant takes as they attempt a series of errands in a large virtual building using the Edinburgh Virtual Errands Task (EVET), (Logie, Trawley, & Law, 2010). This combines a simulation of a realistic setting with control of the environment, the range of cues that the participant will encounter and the range of actions that the participant may perform. In the EVET, each errand has a specific location within the virtual building, spread over thirty-eight rooms and four floors. Access to each floor is provided by two sets of stairs, one for travelling up and the other for travelling down. By allowing participants to roam freely in this virtual space we were able to examine the effect of cue encoding/retrieval disparity as a consequence of the participant's self-determined route. Participants could encounter PM cues (such as a room number or a stairwell) from a variety of directions, presenting several possible PM cue perspectives. Moreover, the context in which these cues are encountered is also variable, such as when they are encountered (early or late in the test) and what tasks are currently active (number of items carried). For example, one errand involved collecting a keycard on the left hand side of the second floor, but as part of a different errand the participant might be carrying a package to be delivered elsewhere in the building. Prior to starting the test, every participant indicated their optimum errand order, and therefore, by definition, their direction of travel to each errand. During the test, however, each participant has several possible navigational routes to the keycard, such as entering the left side of the second floor via the stairs or crossing the second floor concourse from the right. Furthermore, when they decide to perform this errand during the test, it may be when they happen to encounter a particular cue (e.g. a specific room number) and this may occur earlier than envisaged in the original plan, with several tasks already completed or left to do. This variation provides the basis for the encoding/retrieval discrepancy. In summary, the aim of this paper is to investigate how PM failures in the EVET were related to planning, participant-driven actions and independent measures of cognitive functioning, including tests of verbal and spatial working memory capacity.

# Method

## *Participants*

An initial total of 165 participants were recruited for the experiment. However, 12 participants were unable to finish the independent tests of PM and of planning because of technical problems, so their data were excluded from subsequent analysis. A final total of 153 participants (95 women and 58 men) were included in all subsequent analysis. We describe below the rationale for, and the procedure followed to generate scores.

## Tests and Procedure

All testing was conducted over a two hour session, which was split evenly between the EVET procedure in the first hour and the individual differences measures in the second hour. Except for the word recall test, all tasks were viewed on a 42cm colour monitor and run on a Dell XPS PC with an Intel Core Quad 2.33 Ghz processor and 1GB ATI graphics card. Viewing distance from monitor was approximately 50cm.

### *The Edinburgh Virtual Errands Task* (EVET)

The virtual environment was developed with the Valve Hammer Editor, a 3-D map creation programme freely available with the computer game Half Life 2™. The test building was rectangular in shape, with thirty eight rooms spread over four storeys. All rooms were on either side of the building, separated by a large open concourse on the ground floor and empty space through to the upper floors. Each floor was accessed by two sets of internal stairs located on either side of this space. Figure 1 shows a screen shot of the concourse on the ground floor (floor zero) and a birdseye view of the virtual building. Where appropriate, glass wall panels were used to facilitate learning of the building structure and to make navigation easier for participants.

Figure 1 about here

The participant explored the virtual environment using the keyboard and mouse. With this control method the keyboard was used for forward/lateral/backward movement (keys “a”, “d”, “s”, and “w”) and physical actions such as picking up objects (key “e”). The mouse provided control over visual pitch (up and down) and yaw (spin left and right) perspectives. Participant position and movement within the virtual building was automatically recorded as a series of XYZ spatial coordinates, at a sampling rate of approximately 10Hz. In addition, any actions made by the participants were recorded with a time stamp. Participants were given 8 minutes in which to complete a list of eight errands. Two different lists were used (half the participants completed one list, half completed the other), but both lists followed the same structure. These lists are shown in Table 1. Three of the errands had two stages, for example “Pick up brown package in T4 and take to G6”. One errand was an open-ended task which asked participants to sort as many red and blue file-binders as possible into separate boxes. Participants had to decide for themselves how long they could devote to this task without compromising their overall goal of completing all the errands. The remaining four errands were simple one-step tasks (e.g., turn off lift on ground floor) and two of these had to be completed at or before a particular time. These latter two tasks were removed from the analysis as they were time-based not event-based PM tasks. Participants who used List A started the task on the ground floor, while people who used List B started on the top floor.

The errands were listed in an inefficient order for completion, but participants had the opportunity to make a plan of their preferred order before they began the test.

Table 1 about here

Participants were first given the EVET instruction sheet which detailed the nature of the task, building layout and rules (which they were explicitly asked to follow throughout the entire test period). The building rules required participants only to use the left stairs for travelling down and the right stairs for travelling up, to avoid entering any non-task related rooms and to avoid picking up any non-task related objects. Next, participants completed the EVET practice session (approximately 5 minutes) which required each participant to follow a series of onscreen errand commands. The practice errands were; collect an object and deliver it, press a button on a wall within the environment, unlock the stairwell door with a key-code, and sort some red and blue folders into separate boxes. These practice errands were similar to, but not the same as those used in the main testing session.

Next, participants studied their allocated errand list (set A or B) for two minutes after which they were given a free recall test of the list, and the number of errands correctly recalled was recorded. This was followed by five minutes of further study then a test of cued recall, and again, each participant was scored on the number of errands correctly recalled. After these measures of list recall were taken, participants were provided with a schematic building map and a copy of the errand list. They were asked to indicate the order in which they planned to perform the errands to achieve maximum efficiency, but they were also told that they could change their plan during the actual test. Upon completion of their plan, which took each participant approximately five minutes, the task list was removed along with their written plan, and they were asked again to verbally recall the errand list and building rules. Any mistakes were corrected, and this process was repeated until recall of the list was at 100%

(this required approximately a further two minutes of study time). This minimised the risk that participants would fail to complete errands simply because they could not recall them. Any participants who failed to recall all of the errands after all of these procedures had been followed were asked to perform the EVET anyway, but their data were not included in subsequent analysis (this happened very rarely and these data were not part of the original sample of 165). Including the initial learning phase, plan creating and final checking each participant spent approximately fourteen minutes working with the errand list before starting the EVET. The EVET test lasted for 8 minutes (neither task list nor plan were present during the test). Afterwards they were scored on their free recall of all of the errands regardless of whether all had been completed.

*Independent tests of Cognitive Resources*

*The Word Recall Task* was based on the Capitani, Della Sala, Logie, and Spinnler (1992) general procedure and was used as an independent measure of retrospective memory. It consisted of five lists of twelve words that were read out by the experimenter at a rate of one per second. At the end of each list participants were prompted to recall the words in any order. The dependent variable was total score out of a maximum of 72.

*Working Memory Verbal Span* required participants to verify a series of unconnected sentences while memorizing the last word of each sentence based on Baddeley, Logie, Nimmo-Smith and Brereton, (1985; Duff & Logie, 2001). All sentences were presented in sets, starting with a set of two and finishing with a maximum set size of seven. Regardless of participant performance each set was repeated three times. All sentences were presented for three seconds, and preceded by a fixation cross for one second. Total correct recall of the sentence-final words was calculated as a proportion of maximum possible recall score (81 max). Sentence presentation was controlled by E-Prime 2 (Psychology Software Tools).

*Working Memory Spatial Span* was based on a task devised by Shah and Miyake (1996). Participants were shown a series of block capital letters that appeared consecutively on a computer monitor. They had to judge whether the letter was shown in its normal configuration or as a mirror image. Additionally the letters were shown in different orientations within a circular area, and participants had to memorize these orientations and recall them at the end of the set. The task began with a set-size of two letters, and increased by one letter each time to a maximum of five. All participants completed three repetitions at each set size regardless of whether they had performed previous trials successfully. Letters remained on the screen for three seconds (preceded by a one second fixation cross). Total correct recall was calculated as a proportion of the maximum possible score (70). Presentation was controlled by E-Prime 2 (Psychology Software Tools).

*The Travelling Salesperson Task* (TST) required participants to imagine they were a salesperson who had to visit several target locations in the shortest distance possible. As this task involved the planning of routes between specific locations we used this as an index of planning ability. In our version cities were represented by a 5-by-5 array of coloured shapes (created using Matlab 7.1). At the bottom of each array was an information bar that contained nine coloured shapes, with the first labelled “Start/End” and the rest “Target Locations”. Participants were asked to plan the shortest route that connected all the destinations (assuming straight line distances), and use the mouse to click on each of these target locations in turn. When participants clicked on a location it disappeared from the information bar at the bottom of the screen, leaving only those that had yet to be visited. Participants completed two practice arrays before the main test; the first containing only targets (no distracters) and the second with the full array. They were then given ten test arrays, each of which only had one optimum solution for the set of target locations – this was calculated using an algorithm for travelling salesman problems (Kirk, 2007). Performance was scored as the proportion of distance longer than the optimum, averaged across the ten arrays.

*The Breakfast Task* was devised by Craik and Bialystok (2006) who kindly provided a copy of the computer programme. It was a simulation of the task of cooking breakfast, with different screens showing different foods and a main screen where participants had to set a table by using the mouse to drag and drop items of cutlery into place settings. Each food required to be cooked for a particular length of time (2 minutes to 5.30 minutes), and it was the participant's task to make sure they were all ready at the same time. Therefore, they first had to click on an icon of the food with the longest cooking time (i.e. 5:30 minutes) as shown beside the virtual table. This took them to the screen showing the food along with a timer bar. They clicked on the food icon to start the timer which showed the progression of cooking. They then had to return to the main screen and continue to move cutlery to the virtual place settings until it was time to start the food with the next longest cooking time. This continued until the time at which all the foods should be ready. Participants then had to visit each screen to stop the cooking of each food. Prior to the actual test, participants were given a simple practice scenario involving only two breakfast foods. The outcome measure was the average deviation between the actual start time for each food, and the time that it should have been started. As the task primarily involved prospective memory (for starting each of the foods at the correct time while engaged in another task (table setting), it was taken as a measure of PM ability that was independent from the EVET.

## Results

Results for overall performance on errand completion are reported elsewhere (Logie et al., 2010). Here, we focus on prospective memory data that were not included in that previous report. We describe below the rationale for, and the procedure followed to generate the four main outcome measures.

*PM error Scoring:* The PM error measure relied on the common EVET situation of participants walking past a room that they should have entered to complete an errand. If at the cued recall at the end of the session the participant could still successfully recall that intention then this was marked as a PM error. Although this approach to PM assessment appears very different from that used in the typical Einstein-McDaniel paradigm, the two are equivalent in all important respects. In both cases, the participant has been asked to form an intention, with a specific action to perform upon encountering a specific cue. During this retention period the participant is engaged in an ongoing task (navigation) that demands attention. Furthermore, all participants were checked for failures of retrospective memory for the tasks they were asked to perform. However, by allowing the participant free movement we are attempting to create realistic PM scenarios, in contrast to the more common practice of the experimenter prescribing the exact cue context from the start. PM error score was calculated as the number of errors divided by the number of cues encountered.

*EVET travel time:* This indicated the total amount of time each participant spent travelling in the EVET building. Time spent in a room was excluded (i.e., completing a specific errand), so it was predicted that this measure would directly index each participant's ability to efficiently navigate their path through the virtual building.

*Errand follow score:* This score was designed to highlight the overlap between planned and actual errand performance for each participant. Furthermore, it indexes the relationship between encoding and retrieval that is a function of the choices made by the participant during the test. The correspondence between these errand orders was based on allocating one point for each errand that was conducted in the same position or sequence as planned. The follow score was calculated by dividing total overlap points by number of tasks completed.

*Plan Efficiency:* We identified the optimum plan by calculating the minimum distance required to complete all eight errands, while following the building rules and working within the time constraints imposed by the two time based errands. This calculated optimum plan was validated by finding it matched with the average task rank order of the five highest performing subjects (see Logie et al., 2010).

Insert Table 2 about here

Descriptive statistics and intercorrelations among all these measures are shown in Table 2. Planning efficiency correlated with better overall PM performance – i.e., participants with efficient plans tended to have fewer PM errors. However, this relationship did not hold when controlling for whether participants actually followed their plan ($r$ = -.09, $p$=.29). In contrast, the partial correlation between the plan following measure and PM performance (when controlling for plan efficiency) was significant ($r$ = -.33, $p$<.001). This relationship suggests that participants who stuck with their original plan (good or bad) tended to have fewer PM errors than participants who changed their plan on-line, even if the change resulted in a plan that was closer to the optimum. The role of spatial working memory is highlighted through a significant relationship with EVET travel time ($r$ = -.19, $p$<.02), whereas no significant relationship was found between EVET travel time and verbal working memory capacity. This is consistent with domain-specific spatial working memory resources linked with navigation around the building.

Insert Table 3 about here

Additional analyses focused on examining which of the five independent measures contributed unique variance to the prediction of PM errors. This was carried out using multiple regression techniques, and the results of multiple linear regression with backwards stepwise elimination measures are shown in Table 3. The regression model showed that independent measures of planning ability (TST), PM (breakfast task) and verbal working memory span were reliable predictors, while neither spatial working memory nor the word recall task had any unique relationship with number of PM errors. The failure of spatial working memory performance to act as a reliable predictor argues for domain-specific working memory processes, and highlights a role for verbal working memory capacity in successful prospective memory as assessed by the multiple errands methodology.

## Discussion

The aim of this paper was threefold: First, to explore the role of planning in successful PM performance, second, to investigate whether domain-specific or domain-general working memory processes are at play in PM, and finally, to validate a novel approach to PM assessment using cost effective virtual reality software. With regard to the first aim, the importance of the planning task in the regression model is in line with Kliegel et al. (2000) who highlighted the role that planning has in successful PM. One novel finding here is the correlation between plan following and PM performance in the new paradigm. It would appear that following a plan, rather than changing the plan on-line, provides some PM retrieval support. Although previous research has demonstrated this relationship (Kleigel et al., 2001), our study is the first to show this effect of planning on PM performance in participants whose choices in the environment affect the match or the discrepancy between the context for encoding and the context for retrieval.

With regard to our second aim, the significance of verbal working memory highlighted in our regression model is in accordance with previous work that has demonstrated a link between working memory span and higher PM performance (e.g., Brewer et al., 2010; Smith et al., 2003; Smith & Bayen, 2005; West & Craik, 2001; Einstein et al., 2000). Meilinger, Knauff and Bülthoff (2008) reported a study where participants learned specific routes through a virtual city while either performing a verbal, visual or spatial concurrent task. During a subsequent test phase, participants who had performed a verbal or spatial task during the learning phase were more likely to get lost. This finding is in line with our results, which demonstrated a relationship between lower PM errors and higher verbal working memory capacity. The nonsignificance of spatial working memory capacity as a predictor could be interpreted as evidence that PM is primarily a cognitive process that is represented in the verbal domain, and no spatial representations are required for successful performance. Alternative explanations are possible; the first relates to the PM error measure itself. By only examining behaviour around the PM errand location itself we are, in effect, ignoring the navigational effort it took to get there. This interpretation is supported by the significant relationship that spatial, but not verbal, working memory capacity had with our index of movement efficiency (EVET travel time). The removal of navigational effort from our measure of PM performance addresses the disparity with Marsh and Hicks (1998) finding of interference from both spatial and verbal concurrent tasks on PM performance. However, it is important to clarify the distinction between the Marsh and Hicks study and our approach. In addition to our PM error measure not indexing spatial ability, the absence of spatial working memory as a predictor of PM performance does not indicate there is no functional relationship. Rather, it states that spatial working memory capacity cannot predict PM performance in this version of the EVET task. However, it may be that only a minimal level of spatial working memory is required for the task. This would make a measure of the maximum spatial working ability of each participant insensitive to variations in EVET performance (see Logie & Baddeley, in press for a discussion). A different virtual environment, for example one familiar to the participant, could result in spatial working memory being a better predictor, if planned errand order is based on a route rather than solely a list of errands. An everyday example here might be planning for shopping in a familiar supermarket where the locations of specific goods are known. Also of note is that we have used a two dimensional spatial task as an independent measure of working memory ability. Considering the three dimensional nature of the EVET a spatial wayfinding task might be more suitable for use in future studies (for example see Wolbers & Buchel, 2005).

Similarly, based on the above premise, the failure of the retrospective memory measure (the recall task) as a significant predictor was not unexpected, given that our index of PM performance only considered tasks that the participant could successfully recall after the EVET. Specifically, we actively attempted to separate prospective from retrospective failures. This is in line with the standard prospective memory research methodology in which participants are asked post-test to recall their instructions. Therefore, since PM failures cannot be attributed to retrospective failures it is not surprising that the retrospective memory was not a significant predictor of PM performance. A further manipulation could incorporate concurrent task methodology that may highlight the resources required. For example, an interesting question for a future study is whether concurrent performance of a verbal or spatial orientated task, would selectively interfere with PM performance during the EVET.

The final aim of this paper was the development and validation of a novel methodology, which has been demonstrated, in part, by the planning effects reported above. By allowing free movement we are creating a larger and more complete picture of the factors contributing to PM performance. It is hard to envision how the standard laboratory paradigm could address the relationship between encoding and retrieval as conceptualised in this paper. EVET incorporates advantages of a naturalistic PM paradigm with experimental control of the environment. It also allows for very much shorter testing time than is possible with naturalistic paradigms that may take several hours (e.g. Shallice & Burgess, 1991), or several days (e.g. Rendell & Craik, 2000). Like the typical laboratory PM task, EVET has an ongoing task of navigation around the virtual building. However, our experimental platform is sufficiently flexible that, in future studies, it could readily be used to investigate other research questions such as the impact on PM of different additional ongoing embedded tasks (e.g. Scullin, McDaniel & Einstein, 2010; Smith, Hunt, McVay, & McConnell, 2007). A potential caveat might be whether this novel multiple task approach to the study of PM can be compared with results from studies that measure PM using more traditional single task PM methodologies. As noted, one of our aims was to introduce a new kind of paradigm that can address questions about PM performance that cannot readily be addressed by traditional PM laboratory paradigms. A further aim was to incorporate the experimental control that is missing from naturalistic PM paradigms. It is worth noting that Burgess (2000) accounted for multitasking impairments that are sequellae to frontal lobe damage, in part by partitioning specific measures of PM contributions to multitasking performance. This work shows that not only is PM a key component of successful multitasking, but that it can be indexed separately from other cognitive processes. Similarly, Kliegel et al. (2000) indexed PM performance on their complex PM task which required participants to perform multiple tasks. Therefore we see our results as being complementary, but adding to those obtained from typical laboratory paradigms.

A potential implementation of this methodology would be to create virtual analogues of real-world locations and explore the effect of location familiarity on PM. Titov and Knight (2001) have shown that familiarity with an environment improves prospective memory performance. These authors developed a video paradigm that attempted to replicate an everyday shopping experience which manipulated context by using two films; both show very similar shopping streets, one familiar and the other unfamiliar. The familiar location produced significantly more successful PM responses than the unfamiliar. They argued that, although the two videos were in principle identical, location familiarity (and consequent availability of contextual cues) enhanced planning and organisation of the PM tasks. However, using video material of actual locations results in several methodological issues. In addition to the difficulties involved with identifying suitable intentions, cues and responses from video material, the suggestion of “movement” is dictated by the serial order of video clip presentation. With such passive video presentation, variability in navigational strategies between individuals cannot be assessed. Moreover, research has shown that navigator movement strategies (Hölscher, Meilinger,Vrachliotis, Brösamle, & Knauff, 2006; Hölscher, Büchner, Brösamle, Meilinger, &Strube,2007) and target orientation (Frankenstein, Meilinger, Mohler & Bülthoff, 2009) are heavily influenced by their degree of familiarity with the environment. This new approach may allow researchers to explore how and when a-priori knowledge of the structural and functional aspects of a location is used in PM. The important role of “cue specificity” in PM, as highlighted by Ellis and Milne (1996), provides a theoretical framework for future studies into the role of location familiarity and PM performance.

The role of planning in successful performance was highlighted by the significance of planning (TSP) in the regression model and the importance of plan following for PM error behaviour. We have not yet addressed the question of what processes are involved when people form a plan for a future activity (in contrast to the plan following measure discussed above). From our current data set we can see that among the independent measures of cognition only word recall had a significant correlation with planning efficiency. The absence of a correlation with either the working memory or planning tasks is unexpected. One explanation centres around the difficulty of creating an EVET plan, and opportunities for elaborating plans (see Kliegel et al., 2000). In future studies with EVET, planning difficulty could be manipulated by allowing the participant to determine their preferred level of plan elaboration.

In conclusion, by using a novel methodology for examining PM in a healthy young adult population, the data demonstrate how participant-driven plans are implemented and the how their implementation affects PM performance. Furthermore, our results are consistent with domain specific cognitive resources, not a global attentional resource, for successful PM performance.

References

Baddeley, A. D. & Logie, R. H. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (eds.) *Models of Working Memory*, pp. 28-61. New York: Cambridge University Press.

Baddeley, A. D., Logie, R. H., Nimmo-Smith, I and Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language 24*, 119-131.

Bailey, P. E., Henry, J. D., Rendell, P. G., Phillips, L. H., & Kliegel, M. (2010). Dismantling the "age-prospective memory paradox": The classic laboratory paradigm simulated in a naturalistic setting. *Quarterly Journal of Experimental Psychology*, 63, 646-652.

Brewer, G. A., Knight, J. B., Marsh, R. L., & Unsworth, N. (2010). Individual differences in event-based prospective memory: Evidence for multiple processes supporting cue detection. *Memory and Cognition 38* (3), 304-311.

Burgess, P. W., Veitch, E., de Lacy Costello, A. & Shallice, T. (2000). The cognitive and neuroanatomical correlates of multitasking. *Neuropsychologia, 38,* 848–863.

Capitani, E., Della Sala, S., Logie, R., & Spinnler, H. (1992). Recency, primacy and memory: Reappraising and standardising the serial position curve. *Cortex, 28,* 315-342.

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2005). Associating a time-based prospective memory task with an expected context can improve or impair intention completion. *Applied Cognitive Psychology, 19*, 345-360.

Craik, F. I. M., & Bialystok, E. (2006). Planning and task management in older adults: cooking breakfast. *Memory & Cognition, 34,* 1236-1249.

Duff, S. C. & Logie, R. H. (2001). Processing and storage in working memory span. *Quarterly Journal of Experimental Psychology*, *54A*, 31-48.

Einstein, G. O., & McDaniel, M. A. (1990). Normal Aging and Prospective Memory. *Journal of Experimental Psychology-Learning Memory and Cognition, 16*(4), 717-726.

Einstein, G. O., McDaniel, M. A., Manzi, M., Cochran, B., & Baker, M. (2000). Prospective memory and aging: Forgetting intentions over short delays. *Psychology and Aging, 15*(4), 671-683.

Ericsson, K., & Simon, H. (1993). Protocol Analysis: Verbal Reports as Data (2nd ed.). Boston: MIT Press.

Ellis, J., & Milne, A. (1996). Retrieval cue specificity and the realization of delayed intentions. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology, 49*(4), 862-887.

Farrimond, S., Knight, R. G. &Titov, N. (2006). The effects of aging on remembering intentions: Performance on a simulated shopping task. *Applied Cognitive Psychology*, 20, 533–555.

Frankenstein, J., Meilinger, T., Mohler, B. J., & Bülthoff, H. H. (2009). Spatial Memory for Highly Familiar Environments. In Taatgen, N., & Van Rijn, H. (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society (pp.* 2650-2655). Curran, Red Hook, NY, USA.

Phillips, L. H., & Crawford, J. R (2004).A metaanalytic review of prospective memory and aging. *Psychology and Aging, 19*(1), 27-39.

Hölscher, C., Büchner, S., Brösamle, M., Meilinger, T., & Strube, G. (2007). Signs and Maps – Cognitive Economy in the Use of External Aids for Indoor Navigation. In McNamara, D. S. & Trafton, J. G. (Eds.), *Proceedings of CogSci 2007,* (pp. 377-382). Austin, TX: Cognitive Science Society.

Hölscher, C., Meilinger, T., Vrachliotis, G., Brösamle, M., &Knauff, M. (2006). Up the Down Staircase: Wayfinding Strategies and Multi-Level Buildings. *Journal of Environmental Psychology 26*(4)*,* 284-299.

Kinsella, G. J., Ong, B., & J., Tucker. (2009). Traumatic Brain Injury and Prospective Memory in a Virtual Shopping Trip Task: Does It Matter Who Generates the Prospective Memory Target? *Brain Impairment, 10*(1), 45-51.

Kirk, J (2007). Traveling Salesman Problem - Genetic Algorithm. Retrieved from Matlab File Exchange website: http://www.mathworks.com/matlabcentral/fileexchange/13680-travelingsalesman-problem-genetic-algorithm.

Kliegel, M., McDaniel, M. A., & Einstein, G. O. (2000). Plan formation, retention, and execution in prospective memory: A new approach and age-related effects. *Memory & Cognition*, *28,* 1041-1049.

Logie, R.H. & Baddeley, A.D. (in press). Working Memory: The State of the Science. Current Directions in Psychological Science.

Logie, R.H. & Maylor, E.A. (2009). An internet study of prospective memory across adulthood. *Psychology and Aging,* 24, 767–774.

Logie, R.H., Trawley, S. L., & Law, A.S (2010). Multitasking: Multiple, Domain-Specific Cognitive Functions in a Virtual Environment. Manuscript submitted for publication.

Marsh, R. L., & Hicks, J. L (1998). Event-based prospective memory and executive control of working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(2), 336-349.

Maylor, E.A. & Logie, R.H. (2010). A Large-Scale Comparison of Prospective and Retrospective Memory Development from Childhood to Middle-Age. *Quarterly Journal of Experimental Psychology*, *63*, 442-451.

McDaniel, M. A., Robinson-Riegler, B., & Einstein, G. O. (1998). Prospective remembering: perceptually driven or conceptually driven processes? *Memory and Cognition , 26*(1), 121-134.

McDaniel, M.A. & Scullin, M.K. (2010). Implementation intention encoding does not automatize prospective memory responding. *Memory and Cognition, 38*, 221-232.

Meilinger, T., Knauff, M., &Bülhoff,. H. H. (2008). Working memory in way finding – A dual task experiment in a virtual city. *Cognitive Science 32,* 755-770.

Nowinski, J. L, Dismukes, K (2005). Effects of ongoing task context and target typicality on prospective memory performance: The importance of associative cueing. *Memory, 13,* (6), 649-657.

Paraskevaides, T., Morgan, C. J. A., Leitza, J.R., Bisby, J. A., Rendell P. G., &Curran, H. V. (2010). Drinking and future thinking: Acute effects of alcohol on prospective memory and future simulation, *Psychopharmacology, 208,* 301-308.

Rendell, P. G. & Craik, F. I. M. (2000). Virtual week and actual week: Age-related differences in prospective memory. *Applied Cognitive Psychology*, *14*, 43-62.

Sellen, A. J., Louie, G., Harris, J. E., & Wilkins, A. J. (1997). What brings intentions to mind? An in situ study of prospective memory. *Memory, 5(4),* 483-507.

Shah, P., Miyake, A. (1996). The Separability of Working Memory Resources for Spatial Thinking and Language Processing: An Individual Differences Approach. *Journal of Experimental Psychology: General, 125*, 4-27.

Scullin, M. K., McDaniel, M. A. & Einstein, G. O. (2010). Control of Cost in Prospective Memory: Evidence for Spontaneous Retrieval Processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 190–203.

Shallice, T., & Burgess, P. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114,* 727-741.

Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 347-361.

Smith, R. E. & Bayen, U. J. (2005). The effects of working memory resource availability on prospective memory: A formal modeling approach. *Experimental Psychology,* 52, 243–256.

Smith, R. E., Hunt, R. R., McVay, J. C., & McConnell, M. D. (2007). The cost of event-based prospective memory: Salient target events. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 33, 734–746.

West, R., & Craik, F. I. M. (2001).Influences on the efficiency of prospective memory in younger and older adults. *Psychology and Aging, 16,* 682-696.

Wolbers, T., & Buchel, C. (2005).Dissociable retrosplenial and hippocampal contributions to successful formation of survey representations. *The Journal of Neuroscience, 25*, 3333–3340.

Table 1. EVET errand lists (A & B).

| Errand List | A | B |
|---|---|---|
| 1 | Pickup Brown Package in T4 and take to G6 | Pickup Computer in G4 and take to T7 |
| 2 | Pickup Newspaper in G3 and take to Desk in S4 | Pickup Milk Carton in T3 and take to Desk in F4 |
| 3 | Get Keycard in F9 and unlock G6 (via G5) | Get Keycard in S9 and unlock T7 (via T6) |
| 4 | Meet person S10 before 3:00 minutes | Meet person F10 before 3:00 minutes |
| 5 | Get stair-code from notice board in G8 and unlock stairwell | Get stair-code from notice board in T10 and unlock stairwell |
| 6 | Turn on Cinema S7 at 5:30 minutes | Turn on Cinema F7 at 5:30 minutes |
| 7 | Turn off Lift G Floor | Turn off Lift T Floor |
| 8 | Sort red and blue binders in room S2. Sort as many binders as you can. | Sort red and blue binders in room F2. Sort as many binders as you can. |

Table 2. Descriptive Statistics and Correlation matrix of PM performance and predictive measures

| | Mean | Standard Deviation | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 PM Error Score | 13.84 | 14.34 | -.43** | -.16* | -.35** | -.23** | -.14 | .30** | -.13 | -.17* |
| 2 EVET travel time | 305.10 | 42.08 | | -.02 | -.37** | -.14 | -.19* | .29** | .01 | -.11 |
| 3 Plan Efficiency | 46.36 | 18.59 | | | .23** | .02 | .001 | -.08 | .002 | .17* |
| 4 Plan Follow | 62.30 | 24.09 | | | | -.06 | -.06 | -.30** | .04 | .09 |
| 5. Verbal Working Memory | 82.85 | 16.08 | | | | | .29** | -.18* | .09 | .44** |
| 6 Spatial Working Memory | 73.18 | 24.23 | | | | | | -.09 | .02 | .13 |
| 7The Travelling Salesperson Task | 10.07 | 6.42 | | | | | | | .06 | -.02 |
| 8 Breakfast Task | 16.12 | 14.75 | | | | | | | | .087 |
| 9WordRecallTask | 29.19 | 4.60 | | | | | | | | |

**Correlation is significant at the 0.01 level (2-tailed).

*Correlation is significant at the 0.05 level (2-tailed).

n = 153

Measures 1,3-6 scored as percentage of maximum; Measures 2 & 7 scored as time (seconds).

All measures of z-score skewness and kurtosis below 2 except for PM error measure (skewness, 5.88; kurtosis, 2.33), The Travelling Salesperson Task(skewness, 6.02; kurtosis, 2.47) and the Breakfast Task (skewness, 10.70; kurtosis, 16.55).

Table 3.Results of multiple regression with backwards stepwise elimination to assess the contribution to common variance between Prospective Memory Errors and scores on five different measures of mental ability as described in the text.

| Variable | *B* | *SE B* | *Beta* | *t* | *p* |
|---|---|---|---|---|---|
| TST | 62.05 | 17.30 | .28 | 3.59 | <.0001 |
| Verbal Working Memory | -.19 | .09 | -.17 | -2.20 | .001 |
| Breakfast Task | -.13 | .07 | -.13 | -1.73 | .09 |

$F$=8.14, $df$=3, 152, $p$<.001 $R^2$=.14

Model selection procedure: backwards stepwise elimination.

Excluded (Not Significant): Spatial working memory and word recall tasks.

Figure Captions

Figure 1. Screen shot of EVET concourse area on the ground floor (left) and birdseye view of the building (right) showing details of the top floor.

Figure 1

Dear Dr Radvansky

Thank you for your letter of 20th November 2010, and the comments of reviewers regarding the article "Event based prospective remembering in a virtual world" by myself, Robert Logie and Anna Law. We found the comments extremely helpful when revising the paper, and note below the changes made as well as our responses to the comments of each reviewer. All changes in the paper have been coloured red. We thank you for the opportunity to revise and resubmit our paper, and hope that the manuscript is now appropriate for publication in the Quarterly Journal of Experimental Psychology.

We look forward to hearing from you.

Yours sincerely,

Dr Steven Trawley
For and on behalf of all authors

Editor
**1. Would it be possible to provide a map of the test space to give the reader a better idea of the task.**
R – We have included an overhead picture of the virtual building in figure 1.

**2. How large was the screen on which the virtual environments were presented and how close were people to this screen? This will give some rough idea of how immersive the experience was.**
R – We have now included this information in the methodology section (page 9 – line 10).

Reviewer 1
**1. The primary aim of the paper was well developed and executed. However, one lingering issue that I had with the manuscript was whether or not the EVET is primarily tapping prospective memory abilities or multitasking. Although it is admirable to strive for greater validity with our tasks, there exists a tradeoff with our understanding of the underlying mechanisms of the constructs we are studying. For example, the recall tasks may not have predicted prospective memory failures in the EVET because the need for memory is severely diminished by the multitasking nature of the task and the overlearning of the materials. If so, would these results still translate to prospective memory in everyday life? Perhaps follow up work could collect data on more traditional prospective memory tasks used by Einstein, McDaniel, Smith, Marsh, and others to address the relation between the EVET and standard laboratory tasks.**

R- We thank the reviewer for this suggestion and would plan to follow up this suggested experiment in our future work. We also plan in future studies to compare EVET with a similar task in a genuine real world setting to address the question as to whether EVET is mimicking PM in everyday life. However, these were not questions that we were trying to address here. We have made clearer in the text, the novel contributions to the understanding of prospective memory in a participant-driven paradigm within the study that we report. We also note our view that this new work complements the more traditional laboratory PM tasks which are less well suited to asking the major questions that we pose regarding participant-driven actions. In response to the specific comment, the discussion now expands on the explanation for the non-significance of the recall task (19 – line 17), suggests future versions of the EVET which incorporate a traditional ongoing task (starts page 20 – line 6)., and argues for the validity of using a multitasking test to study prospective memory (ends page 21 – line 2).

**2. The secondary aim of the article should be unpacked. Exploring how participants plan their actions for the future is an important endeavor. The important point is that it would be nice for the authors to speculate more on exactly what people are doing when they are planning (i.e., what information is being used to plan, what is the role of WMC in planning, what factors influence planning).**

R- We agree that our original text was insufficiently clear on this aim, and feel that this arose, in part from the initial focus on the novel methodology rather than the main research questions that the new methodology was devised and used to address. In the revised paper, we now make clear that the impact of planning and participant-driven actions on PM performance is our primary theoretical aim (page 17 – line 15). We have added text in the introduction to make the rationale and expectations clearer. We have also included new text near the end of the discussion (page 22 – line 1) where we discuss the correlation between free recall and planning efficiency, and added some speculation as to its interpretation, as requested by the reviewer. We also indicate how the planning requirement could be investigated further in future studies.

**3. I had a particularly tough time reconciling the third aim of the manuscript with previous research in both the working memory and prospective memory literatures. For example, Marsh and Hicks (1998) found that interference in both spatial and verbal domains (domain specific) created no differences in prospective memory cue detection. When demands were placed on the central executive (domain general) effects on cue detection were observed. Moreover, Brewer et al. (2010 M&C) found that a working memory composite score predicted nonfocal but not focal cue detection. That is to say, WMC measures do not predict prospective memory under all conditions and the true boundary conditions have yet to be fully delineated. The results in the current manuscript suggest that the domain general results reported by Marsh and Hicks may not hold in the EVET or even in more ecologically valid settings. However, given the unusual relation between WMC and PM I think that the authors should not make very strong claims that PM is reliant only on verbal WMC (e.g., the last sentence of the current version of the manuscript). Also, without knowing how reliable the EVET is then perhaps the correlation with spatial WMC is attenuated.**

R- We have modified and extended our description of the findings both for spatial working memory in relation to travel time and navigation in the environment, and for verbal working memory in relation to the PM error measure (page 18 – line 20) in response to this comment.

**Minor points**

**Perhaps the authors should provide a more thorough review of the multitasking literature and relate it to PM in the current study. I'm still not sure that I would necessarily consider the EVET a prospective memory task.**

R –We have included additional arguments about the validity of the EVET as a paradigm within which to study prospective memory, and note that our aim is to examine PM in a more realistic scenario (page 20 – line 6). We agree that EVET is not like traditional laboratory paradigms for studying PM, but argue that it is closer to the real-world type PM paradigms that have been used in studies of the age-prospective memory paradox than are the standard laboratory paradigms. In common with the real-world PM paradigms, EVET involves participant-driven actions in the environment that determine when and how the PM cue is encountered. A major difference between real-world PM paradigms and EVET is that our paradigm combines key features of a real world setting with considerably more experimental control than can be achieved in actual everyday life, which is a great deal more complex than is EVET. We are not addressing the age-prospective memory paradox here, although plan to do so in future studies. However, we are using a laboratory controlled version of more realistic settings that have been used on the paradox studies to understand how pre-planning and participant-driven actions affect PM performance in healthy young adults. To achieve these aims requires the development of novel methodologies rather than traditional laboratory paradigms. There is a limited previous literature on this form of multitasking that has been used to explore PM impairments following frontal lobe damage, and we now refer to some of that previous work as requested by the reviewer. However, we feel that a more detailed review of the multitasking literature would detract from the main foci of the paper. Also, we make clearer why we feel that our use of this kind of paradigm to study PM in a realistic setting with healthy young adults is novel, and has yielded novel insights.

**How long on average did it take for participants to achieve 100% recall of their intentions in the EVET? This is somewhat important information because of the unfettered nature of participant-guided encoding.**
R - We have included this information on page 12 (line 6) – total study time (inc learning, planning and final check) was approximately 14 minutes.

**In the descriptions of some of the tasks there is no explicit mention of the dependent variable contributing to the analyses. The descriptions should be more consistent in this regard.**
R - We have included this information for the word span task and made the descriptions consistent throughout the manuscript.

**In the EVET, what happened to the error score if a participant never crossed an intention-related room? Was this scored as an error and did the numerator differ across participants? If so, does the regression analysis depend on this scoring method?**
R - Both the numerator AND denominator did differ across participants. Variation in the denominator was inevitable due to participant movement - some participants would not access all errand related parts of the building. Therefore, the score reflects the number of PM failures as a function of the number of opportunities to implement a PM intention.

**It would be helpful to explain the logic behind using the highest scoring participants' plans to define plan efficiency. Was there no other way to define a plan efficiency measure that was not dependent on participants' performance?**

R - The efficiency score created from the best performers did in fact match the optimum errand order that we created prior to testing. We have now included new text (page 16 - line 3) which now uses the top performers to validate our calculated optimum score.

**Perhaps the authors could use a composite of WMC scores (verbal + spatial) or derive a factor of the two scores to investigate domain general processes and PM.**
R –We did investigate this and although a composite WM score (spatial + verbal) was a significant predictor in the regression model for PM performance, it did not add a significant change to the model R-squared than did the verbal WM score alone ($R^2$ = .142 compared to $R^2$ = .141).

**The end of the first paragraph of the discussion seemed to ramble a bit.**
R - We thank the reviewer for this helpful comment and have deleted this text.

**It would be nice for the authors to speculate on how 'specificity' may or may not be translated into the EVET. Also, some of the intentions seemed fairly specific.**
R –We have now included a reference to "specificity' in terms of errand familiarity and how this might be implemented in future studies (page 21 - line 22)

**I found the discussion of implementation intentions to be slightly tangential.**
R - On reflection, we agree and so have removed this text from the discussion

**Table 2 should include measures of skew and reliability where appropriate.**
R –Measures of normality have been noted (page 30 – line 22).

Reviewer: 2
**I am a bit confused on the authors' discussion of why spatial working memory was not correlated with PM performance. They mention that spatial working memory is correlated with travel time, showing that those with higher spatial working memory traveled the building faster. I believe that they are arguing that travel time is an indirect measure of PM; that is, if you finish your plan/route quickly, then you are likely remembering what you have to do. Is that correct? If so, the authors may want to spend a little more time addressing this point. Otherwise, it is unclear why spatial working memory wouldn't correlate with PM performance. What is the authors stance on the processes participants use when coming up with their initial plan? Are they using spatial representations, and do they access those during navigation?**
In response to this and a similar comment from reviewer 1 we have expanded upon our explanations for this finding (page 18 – line 20).

**Table 3 presents the results from their regression models. Their description of the analysis is a bit unclear; are these three predictors in Table 2 all part of the same model, or do they comprise their own models?**
R – One regression model was run with the PM error measure as a criterion and the five independent measures of cognition as predictors. The three variables in table 3 are the only variables that made a significant contribution to the model.

**As a minor issue, I believe the authors mean the same thing when they say "Plan Quality" and "Plan Efficiency." Is that correct? If so, it would be good to make sure term use is consistent.**
R –We have corrected table 2 – now reads "Plan Efficiency" instead of "Plan Quality".

**My only major concern is that it is questionable whether the results add much to the existing literature. For example, there has been some good work regarding naturalistic experience and PM. Perhaps with revision, the authors can more strongly highlight the novelty of their results. I hope that my comments have been useful for the authors as they continue to explore this interesting line of work.**

R - We note in response to Reviewer 1 above the major point that we have examined PM performance in a paradigm that is close to naturalistic experience, but with a great deal more experimental control and measurement precision than is possible outside of the laboratory. So, this form of 'paradigm shift' is novel. A major novel finding is that PM performance is better when participants stick with their preplanning for the task order, rather than if they try to change the planned order on-line. The finding that verbal, but not spatial working memory contributes to PM performance adds to findings in the area, and suggests reasons for contrasting results in previous literature. We hope that these novel contributions are clearer in the revised paper.

**Reviewer: 3**

**As noted the other important contribution is the introduction of the novel methodology (EVET) and it is appropriate to have as a primary aim: to introduce this methodology. Indeed, I think this aspect could be capitalized more with further comparisons to similar tasks and clarification of the unique contribution of EVET.**

R - We have expanded on the novelty of the paradigm in the introduction, and as noted in our responses above to comments from reviewers 1 and 2. Also as noted, we have made clear that our primary aim is to investigate PM failure in relation to planning and participant-driven actions, developing the EVET paradigm to do so. We contrast the EVET with naturalistic PM paradigms that involve a great deal less control and a great deal more complexity.

**The authors have appropriately indicated two key areas where the task has useful variations to the classic McDaniel and Einstein paradigm. Firstly, stimuli and a context that closely represents situation in daily life, rather than abstract stimuli separate from daily life. Secondly, the timing and/or appearance of the PM cues can be influenced by participants. However there is one possible advantage of M and E paradigm is that it has the capacity to measure ongoing task performance. Authors might comment on whether this might be possible in the future with EVET.**

R - We thank the reviewer for this excellent point. We have included additional text (page 20 – line 9) noting that ongoing tasks could readily be incorporated in EVET which is a very flexible experimental paradigm. One suggestion for the EVET would be to present two types of errands; location unspecified and location specified. For example, participants would be presented with the standard EVET errand list but with additional instruction to press a button if they see any wall pictures. With this approach, location specified tasks would function as the ongoing task and location unspecified tasks as the PM task. A within-subjects manipulation would allow a comparison between performance with and without the additional location unspecified PM tasks.

**The feature of closely representing tasks in daily live with a virtual experience is shared by each of Virtual Week (Rendell &Craik, 2000), Shopping Street Task (Titov& Knight, 2001; Farrimond, Knight, &Titov, 2006)) and Breakfast task (Craik& Bialystok, 2006). If space, I recommend some more highlighting of the additional contribution made by EVET in comparison to these tasks. For example, Virtual Week like the M & E paradigm unfolds according to a predetermined schedule and thus may not be as appropriate for a planning oriented study. Although, Paraskevaides et al. (2010) (cited in this manuscript) used Virtual Week to manipulate encoding conditions. The**

**breakfast task is closer to the EVET, as in the breakfast task an individual has considerable control over the scheduling and time allocation given to ongoing tasks and therefore provides an opportunity to test the role planning. This task was appropriately included as a second measure of PM. In the introduction the authors suggest some fine distinctions between the tasks. The paper could benefit from some exploration in the discussion of the merits of these two methodologies, in light of the results.**
R -We acknowledge that this is an important point and now include comparisons between EVET and the laboratory simulations mentioned by the reviewer (page 7 – line 4).

**Finally, the shopping task at first impressions seems very similar involving a video paradigm where participants taken on a tour of shops. The authors appropriately observe that the participant is much more active in EVET. They note that shopping task involves a passive video presentation, where participants observe a tour of shops and have to complete specific actions in relation to specific events. I think this discussion should go further to cover how the EVET makes a substantial advance on the Shopping task by having an ongoing task. The Shopping task lacks the critical feature of PM tasks- an ongoing task. Farrimond, Knight, and Titov (2006) acknowledge this as a limitation and concede the shopping task lacks a significant feature of a PM task, an ongoing task that has to be interrupted (see p. 535 and p. 551) and acknowledge their task lies outside the domain of PM research and may be best described as task requiring vigilance and recognition. In shopping task participants passively travel by video presentation from shop to shop and the only tasks are the list of intentions to complete. Street sounds and voices provide some distraction, but do not constitute an ongoing task. I think the EVET is a substantial advance on the Shopping task because of the passive video presentation compared to the participant controlled exploring of the 3 D virtual. There should be some comment about the lack of ongoing task in Shopping street and how this is addressed in EVET. I think it is addressed with the tasks of maneuvering around the complex 3 D buildings, following maps and following rules (eg. left stairs for down and right stairs for up).**
R-We have added some discussion of these points in the introduction and in the discussion. In particular, we note in the discussion that there is an ongoing navigational task requirement, but note also that in future studies we could explore the different research question of the impact on PM of having or not having an additional embedded task.

**Some other minor matters**
**(1) One additional advantage of the EVET that could be highlighted is the relatively short length of testing time. The times for components were noted but the total time could be made clearer.**
R-We have highlighted the modal testing times in the manuscript and note that this is an advantage for EVET over some other naturalistic PM paradigms.

**(2) Correlation table, could delete 1 for correlation with same variable**
R-These have been deleted from Table 2.