Law, AS, Trawley, SL, Brown, LA, Stephens, AN and Logie, RH

The impact of working memory load on task execution and online plan adjustment during multitasking in a virtual environment

http://researchonline.ljmu.ac.uk/1163/

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Law, AS, Trawley, SL, Brown, LA, Stephens, AN and Logie, RH (2013) The impact of working memory load on task execution and online plan adjustment during multitasking in a virtual environment. QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 66 (6). pp. 1241-1258. ISSN

The impact of working memory load on task execution and on-line plan adjustment during multitasking in a virtual environment

Anna S. Law[1], Steven L. Trawley[2], Louise A. Brown[3], Amanda N. Stephens[4] and Robert H. Logie[5]

[1]School of Natural Sciences and Psychology, Liverpool John Moores University, UK

[2]Centre for Gerontology and Rehabilitation, School of Medicine, University College Cork, Ireland

[3]Division of Psychology, Nottingham Trent University, UK

[4]School of Applied Psychology, University College Cork, Ireland

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

Short title: WM and planning during multitasking

Corresponding Author: Dr Anna Law

Liverpool John Moores University

Tom Reilly Building

Byrom Street

Liverpool

L3 3AF

Tel: +44 151 904 6227

E-mail: a.law@ljmu.ac.uk

Abstract

Three experiments investigated the impact of working memory load on on-line plan adjustment during a test of multitasking in young, non-expert, adult participants. Multitasking was assessed using the Edinburgh Virtual Errands Test (EVET - Logie, Trawley & Law, 2011). Participants were asked to memorise either good or poor plans for performing multiple errands, and were assessed both on task completion and the extent to which they modified their plans during EVET performance. EVET was performed twice, with and without a secondary task loading a component of working memory. In Experiment 1 articulatory suppression was used to load the phonological loop. In Experiment 2, oral random generation was used to load executive functions. In Experiment 3, spatial working memory was loaded with an auditory spatial localisation task. EVET performance for both good and poor planning groups was disrupted by random generation and sound localisation, but not by articulatory suppression. Additionally, people given a poor plan were able to overcome this initial disadvantage by modifying their plans on-line. It was concluded that, in addition to executive functions, multiple errands performance draws heavily on spatial, but not verbal, working memory resources but can be successfully completed on the basis of modifying plans on-line, despite a secondary task load.

Acknowledgements

We are grateful to Matthew Logie for undertaking the programming to create the EVET environment. We acknowledge Leverhulme Trust research grant number F/00 158/W which supported the research that we report here.

In a situation where there are several tasks to complete within a limited period of time, it is often necessary to interleave performance of those tasks, by switching back and forth between them. This type of activity is known as multitasking (Burgess, 2000), and it is a ubiquitous requirement of modern life in both domestic and workplace situations (e.g., Hambrick, Oswald, Darowski, Rench & Brou, 2010). Multitasking in non-expert individuals has thus far been studied mainly within the field of neuropsychology, in terms of the deficits in everyday and laboratory functioning demonstrated by individuals with cortical lesions (e.g., Alderman, Burgess, Knight & Henman, 2003; Burgess, Veitch, de Lacy Costello & Shallice, 2000). The aim of the present research was to investigate the cognitive demands of multitasking among healthy individuals, by studying the role of working memory resources in the process of plan execution and adjustment during multitasking. In contrast to previous correlational approaches (Logie, Trawley & Law, 2011; Trawley, Law & Logie, 2011), which examined the relationship between working memory capacity and multitasking performance, we used an experimental, dual-task methodology to examine the impact of concurrent task load on execution of pre-prepared plans in a virtual version of the Multiple Errands Test (Shallice & Burgess, 1991). The present study goes beyond a previous investigation using dual-task methodology (Law, Logie & Pearson, 2006), by manipulating the type of plans that people were given prior to the test, and by using more fine grained measures of multitasking performance.

The Multiple Errands Test is high in ecological validity, because it requires participants to complete a series of errands, while abiding by specified rules, for example in a shopping precinct (e.g., Alderman et al., 2003; Garden, Phillips, & McPherson, 2001), or hospital setting (Knight, Alderman & Burgess, 2002; Dawson et al., 2009). Another strength of the test is its sensitivity to neurological impairment. However, its drawbacks are that it is cumbersome and time-consuming to administer, and the experimenter relinquishes control over many aspects of the situation when using real environments. This has led to the development of computerised versions of the test that can be performed in virtual 3D environments (Law et al., 2006; Logie, Law, Trawley and Nissan, 2010; Logie et al., 2011; McGeorge et al., 2001; Rajendran et al., 2011; Rand, Basha-Abu Rukan, Weiss & Katz, 2009; Trawley et al., 2011). The present study used the Edinburgh Virtual Errands Test (EVET; Logie et al., 2011), in which participants had to complete a series of memorised errands in a virtual shopping and office building.

Burgess et al. (2000) suggested that planning may be one of the cognitive constructs that supports the ability to multitask, along with retrospective memory and intentionality (the ability to act on future intentions, or prospective memory). More recently, Logie et al. (2011) proposed that for errand-based multitasking such as required in real or virtual versions of the Multiple Errands Test, 'planning' can be fractionated into task-ordering and goal-setting processes that occur before task execution (pre-planning) and task re-ordering and goal adjustment processes that occur during the task (on-line plan adjustment). Law (2004) used an early version of a Virtual Errands Test (McGeorge et al., 2001), and found that participants took a more efficient route around the virtual environment when they had been given the chance to engage in pre-planning. However, this greater route efficiency did not impact on overall score, because participants had plenty of time to complete the task, even if they had taken an inefficient route. Moreover, performance was recorded by video of participant movement and actions, limiting precision and detail in the data. Therefore, in subsequent developments (Law et al., 2006; Logie et al., 2011; Trawley et al., 2011), the test was adapted to pose more of a challenge for healthy young adults. In EVET, participants have to memorise the list of errands rather than keeping a copy with them throughout the task, there is greater time pressure, and substantially more detailed information is recorded on-line about movements and actions of participants in the virtual environment. In the present study, the primary focus was the on-line aspect of planning during errand-based multitasking. Participants underwent a thorough process of memorising, in serial order, a list of errands presented to them in the form of a pre-planned sequence. One group received a plan that was in the optimally efficient order for errand completion (*good* plan), and the other group received a plan that was sub-optimal (*poor* plan). Although they had to memorise the plans in serial order, participants were permitted to re-order the errands while they actually performed the task.

It might be expected that participants who have learned a poor plan would have worse overall performance than those who learned a good plan. In contrast, if on-line plan adjustment during the test is sufficient for good performance, participants in the poor plan group should be able to overcome their disadvantage. For example, Phillips, Wynn, McPherson and Gilhooly (2001) showed that pre-planning of a complex single task (Tower of London) did not lead to better performance; on-line planning was an equally effective approach to the task and was also preferred by the majority of participants. In contrast, Logie et al. (2011) showed that, although people who started with a good (self-generated) plan performed better than people who started with a poor plan, participants who used their original plan (good or bad) tended to perform better than participants who changed their plans on-line. The interpretation offered for this last result was that, because plans referred to an ordered sequence of actions, changing the order part of the way through disrupted the performance of actions yet to be completed, and this cancelled out any possible advantage from changing a flawed plan. Multiple regression revealed that EVET performance was predicted by independent measures of on-line planning, retrospective memory and spatial working memory, but not by verbal working memory. Furthermore, Structural Equation Modelling identified a latent variable associated with on-line planning that was driven by, but was separate from, latent variables associated respectively with memory and with pre-planning. However, the cognitive resources required to support changes and adjustments to plans remained unclear.

If a participant does engage in on-line plan adjustment while multitasking, it might be expected that he or she would use the resources of working memory to assist in the process of re-ordering the errands and keeping track of current and future goals. Therefore, we might expect on-line plan adjustment to be more difficult when combined with secondary tasks that load components of working memory. This might be particularly evident when participants start with a poor plan, because there would be a greater need for on-line plan adjustment to maximise performance. In Experiment 1, we used the secondary task of articulatory suppression, which is widely considered to disrupt the operation of verbal working memory. Indeed, sub-vocal rehearsal has previously been shown to support performance in task switching paradigms (Emerson & Miyake, 2003; Miyake, Emerson, Padilla & Ahn, 2004) and concurrent articulatory suppression was shown to impair overall score in an earlier incarnation of the Virtual Errands Test (Law et al., 2006). In contrast, Logie et al. (2011) found that individual differences in spatial working memory correlated with EVET performance but verbal working memory did not. So, the previous results based on individual differences suggested that having a high verbal working capacity was of no particular benefit in this form of multitasking. However, even if variance in maximum capacity of a cognitive resource across individuals does not correlate with task performance, the resource could still be required for that task at less than its maximum capacity (Logie, 2011). Therefore we cannot conclude from our previous results that verbal working memory resources are not required for multitasking performance as measured by EVET, for example to rehearse and/or re-order an errand list. In Experiment 1, we explored experimentally the involvement of verbal working memory resources in EVET performance by preventing the use of subvocal rehearsal through use of articulatory suppression. Specifically, we investigated whether participants would be less successful at on-line plan adjustment (i.e., more likely to stick to the pre-learned plan) when they had to perform articulatory suppression at the same time as attempting EVET, than when they completed EVET under single-task conditions.

Experiment 1

Method

Participants

Forty undergraduates (29 female, 11 male) at the University of Edinburgh participated in exchange for course credit. Their mean age was 18.5 years (SD = 0.8 years). None of these participants had taken part in the previously published studies with EVET (Logie et al., 2011; Trawley et al., 2011).

Design

The experiment involved a 2x2 mixed design with two levels of the between-participants factor Plan (good vs. poor) and two levels of the within-participants factor Demand (single vs. dual task). Plan type, demand and task list (set A or B) were fully counterbalanced across participants.

Tasks

EVET was performed on a Dell XPS PC with an Intel Core Quad 2.33Ghz processor and 1GB ATI Graphics Card, with 42 cm colour monitor. It was created using the Valve Hammer Editor map creation program supplied with the PC game Half Life 2^{TM}. Details of the environment are given in Logie et al. (2011). In summary, the 3D environment was a four-storey building comprising an open rectangular concourse from front to back in the centre of the building, and with shops or offices along the left and right hand sides (from the perspective of the building entrance). Two stairwells connected the floors, one on the left side of the building and one on the right (see figure 1A). The building contained a total of 38 rooms, with eight on the ground floor and ten each on the first, second, and third floors. Room numbers incorporated the relevant floor, for example G10 referred to room 10 on the ground floor. Room numbers were displayed on notices outside the doors. The current task time, measured in minutes and seconds, was continuously displayed at the top of the screen. Any objects collected by the participant during the test were displayed on the left (see Figure 1B).

[Figure 1 around here please]

Participants controlled their movement around the building with the mouse and keyboard. Forward and backward movement was achieved with the “w” and “s” keys, respectively. Sideways movement required the “a” and “d” keys, and participants could look up or down using the mouse. Objects were manipulated (e.g., picked up/dropped off) using the “e” key, once centred in the cross-hairs in the middle of the screen. All such actions were recorded into a text file, along with a time stamp, and the participant’s location was recorded approximately every 100 milliseconds as a set of XYZ co-ordinates.

Lists A and B (table 1) were used for the single and dual-task conditions of the experiment. The lists were organised into optimal (A1, B1) and non-optimal (A2, B2) order for completion and these were given to participants in the good and poor planning groups, respectively. The optimal ordering of the errand lists was derived from data from 165 young adults who completed EVET during a previous study (Logie et al., 2011). The routes taken by the five highest scorers from that previous study were examined, and errands were numbered in the order that they had been completed. For each errand, these five scores were averaged. The average scores were placed in rank order to yield an optimum plan. The non-optimal plans, while feasible, were constructed so as to increase the travel time between errand rooms.

Performance on the secondary task was recorded during practice and at test using a handheld digital voice recorder. This recording was later transcribed for subsequent analysis (see below).

[Table 1 about here please]

Procedure

Participants were informed that they would have eight minutes to complete a list of errands in a virtual building, and that they would have to memorise this list in a specific order. It was explained that they could “hold” multiple objects at the same time in the environment and that, although they had to memorise the errands in a particular order, they were free to vary this order once they started the task. They were also told that they would perform the task twice, once alone and once with a concurrent verbal task. They were introduced to the rules for navigating the virtual building and given training in using the controls to move around and perform tasks. The rules were that they should not enter rooms (or pick up objects) that were not on their errand list, and that each staircase was only to be travelled in a particular direction. The training session (approximately five minutes) involved performing a number of actions while guided by on-screen commands – finding a room, picking up an object, delivering it to another room, unlocking the door to the upper floors with a keycode, pressing a button located on a wall within the environment and sorting coloured folders into boxes. Training errands involved different rooms and objects from the main experimental errand lists.

After training participants were given the errand list for the condition they were to attempt first. They were given two minutes to study this list before it was removed and they were asked to recall as much as they could, in the order that it had been presented. This serial recall performance was recorded and they were then given a further five minutes to study the list. After this the list was removed and they were asked to recall the errands again. Participants were not allowed to start the test until they reached 100% recall of the errand list in the planned order (two participants failed this requirement and were replaced).

Before participants began the dual-task condition of EVET, they were given an explanation of the concurrent task (articulatory suppression) to be performed throughout the whole eight minutes of EVET. Following Law et al. (2006), they were asked to say out loud the word “December” at the rate of once per second, and were given practice at doing so. In both the practice and in the main experiment, the inter-utterance intervals were recorded via a voice key and computer to allow assessment of articulation rate. Upon completion of both single and dual-task versions of EVET, participants were again asked to recall the errands.

Results

EVET Score

Following Logie et al. (2011) and Burgess et al. (2000), a weighted scoring procedure that rewarded task completions and penalised rule breaks was adopted. Extra bonus points were awarded for the three errands in which performance could vary (namely the two errands with a time component and the open-ended folder sorting task – see Tables 1 and 2). For the eight other tasks, a point was awarded on completion. Participants could also incur additional penalty points for breaking any of the building rules, such as entering incorrect rooms or ignoring the stair rule (see Table 2).The numbers of bonuses and deductions was based on the frequency distributions of these task completions and errors in a previous sample of 165 healthy adult participants. The rationale for the allocation of bonuses and penalties is fully described in Logie et al. (2011).

[Table 2 about here please]

The average scores for the good plan and poor plan groups, in both single and dual-task conditions, are shown in Table 3. It is clear that these average scores are all very similar. A 2x2 mixed ANOVA, with two levels of the within-subjects factor Demand (single vs. dual task) and two levels of the between-participants factor Plan (good vs. poor) confirmed that there were no significant main effects of Demand or Plan in the EVET score data, *F*s < 1, and no interaction , *F*< 1.

[Table 3 about here please]

Task completions, bonuses and penalties

As there were no effects on the overall EVET score, each component was examined independently to look for more subtle effects of articulatory suppression on performance. The mean number of task completions, bonus points awarded and penalties applied for each group in each condition is shown in Table 4. These data were unsuitable for parametric analyses due to violations of normality in nearly all cases. Wilcoxon tests confirmed that there was no significant difference between single and dual task conditions for any of the 3 measures (all ps > 0.7). Collapsing across demand, Mann-Whitney tests showed that there were no significant differences between the good plan and poor plan groups on the total number of tasks completed, bonus points or penalty points (all ps > 0.7).

[Table 4 about here please]

Plan Following

For each participant in each condition, the rank order correlation between the memorised sequence (either good or poor) and the executed sequence was calculated using Kendall's Tau, and used as an index of plan following. Any errands not completed were removed before this value was calculated. Kendall's Tau calculates how many pairs of ranks would need to be inverted in order to change one sequence into another. The average Tau values for each group in each condition are shown in Table 3. Values closer to one indicate closer adherence to the memorised sequence, while values closer to zero indicate greater deviations from the memorised sequence and more on-line plan adjustment. It is clear from the data in Table 3 that the good plan group have higher plan following scores than the poor plan group. A 2x2 ANOVA confirmed that this main effect was highly significant, $F(1,38) = 65.007$, $p < 0.001$, partial $\eta^2= 0.631$, but that there was no significant main effect of demand, $F < 1$, or interaction, $F < 1$. Therefore, participants who were given a poor plan were more likely to deviate from it during EVET performance, regardless of whether or not they were suppressing articulation.

Post-EVET recall of the errand list

Recall of the errand list at the end of the test showed a clear ceiling effect, suggesting that task failures in EVET were rarely due to forgetting of the list. On average the recall in the single-task condition was 94.8% (SD = 8.3%), while the recall in the dual-task condition was 95.7% (SD = 7.9%). We also examined what percentage of the errands that were recalled in their original serial position on the list. The good group recalled 77.28% (SD = 35.01%) in the original position in the single-task condition, while this figure was 89.44% (SD = 23.96%) for the dual-task condition. The poor group were more likely to change the order in which errands were recalled, with 65.06% (SD = 36.36%) in the original position in the single task condition, and 64.56% (SD = 35.65%) in the dual. There was a marginal main effect of plan group, F(1,34) = 4.03, p = 0.053, partial η^2= 0.12, but no main effect of demand, F < 1, and no interaction, F(1,34) = 1.09, p = 0.303, partial η^2= 0.031.

Articulatory suppression performance

Participants maintained a rate of 1 utterance every 1.6 seconds on average across the 8 minutes of EVET. This was significantly slower than the rate of 1.3 per second that they achieved during baseline practice, t(38) = 4.59, p< 0.001. However, there was no significant difference between the good (M = 1.62, SD = 0.47) and poor (M = 1.64, SD = 0.36) plan groups on this measure, t(38) = -0.11, p> 0.9.

Discussion

It was predicted that participants given sub-optimal plans would be unable to re-organise the errands using on-line plan adjustment, when their ability to sub-vocally rehearse information in verbal working memory was inhibited by articulatory suppression. However, participants in the poor plan group deviated from the learned plan to an equal extent under single and dual task conditions. Across both, they deviated significantly more than participants given a good plan, demonstrating that they were aware of the need to re-order the errands for greater efficiency. They were successful in doing this, despite suppression of sub-vocal articulation, and overall they performed just as well as the good plan group.

Although Law et al. (2006) showed interference (in terms of overall score) from articulatory suppression on a previous virtual errands test, this was only when participants did the dual-task condition first. Here, there was no significant effect of suppression on EVET performance, even when only considering participants who suppressed articulation during their first EVET trial. It is possible that the more difficult mouse-based controls in the older (Law et al., 2006) task created extra cognitive demands (that were unrelated to multitasking) when participants were still new to the task and attempting it for the first time. With the EVET, controls are more user-friendly and the training procedure is more formalised, allowing participants to be more confident with the basic operation of the environment. Their multitasking performance could then be measured with greater precision and with fewer potential artefacts than in the previous virtual errands test.

However, it is notable that both Law et al. (2006) and the present study found that participants slowed their rate of articulatory suppression while performing the EVET. This finding suggests that there was some conflict between the tasks, and participants may have been trying to protect multitasking performance at the expense of articulatory suppression rate. Accordingly, we are not suggesting that verbal working memory plays no role in multitasking *at all*, but rather that multitasking activity does not necessarily require verbal working memory *at its full capacity* (see Logie, 2011). This is consistent with the finding in Logie et al. (2011) that participants' verbal working memory span was not a significant predictor of their multitasking performance. In any case, it was of interest that the rate of articulatory suppression did not differ between the good plan and the poor plan groups, and the presence or absence of suppression did not interact with plan group in the EVET data. Clearly, whatever role was played by verbal working memory in Experiment 1, it was not crucial for changing the order of errands from an initial poor plan. Executive resources in working memory might have been more involved with this aspect of the task, and this possibility was investigated in Experiment 2.

Experiment 2

The earlier study by Law et al. (2006) showed evidence (in secondary-task data) of increased task interference when the secondary task involved oral random generation. This is a task that is typically associated with executive functions, and in particular thought to be demanding of attention (Baddeley, 1966; Evans, 1978). To investigate the consistency of task requirements in EVET compared with the earlier task, we conducted a further study using the same executive secondary task as was used by Law et al. (2006), using the present good and poor plan manipulation to examine its effect on on-line plan adjustment. Given that the errand lists had been learned in advance, EVET performance might rely heavily on long-term memory for the original plan, rather than rely on rehearsal of the errand list in verbal working memory. This might explain why articulatory suppression did not disrupt performance in Experiment 1, or consistently throughout the Law et al., (2006) study. Therefore, according to the widely held assumption that working memory is constrained by a limited capacity attentional system (e.g. Cowan, 2005), the demands of oral random generation should be disruptive of the executive/attentional resources in working memory that would be required to re-arrange the order of the learned errand list during EVET performance. Therefore, participants in the poor plan group should be more likely to fall back on the strategy of adhering to the plan that they had learned rather than undertaking the more difficult task of re-arranging the errands on-line. (Or participants might continue to protect EVET performance, but have great difficulty in performing the concurrent random generation task.) In this case we would expect to see an interaction whereby the plan-following scores of the poor plan group only increase under dual-task conditions.

However, we might assume instead that working memory draws on a range of executive resources (Logie, 2011; Miyake, Friedman, Emerson, Witzki, Howerter & Wager, 2000) such as inhibition of learned or automatic responses, updating, or retrieval from long-term memory (e.g., Unsworth & Engle, 2007). If we consider the detailed task requirements of oral random generation, it seems to require rapid retrieval of items from a well learned item set (e.g. the alphabet, numbers, months of the year) in long-term memory, coupled with inhibition of retrieval from long-term memory for learned response sequences within that set (see Baddeley, 1996). In this case, oral random generation might disrupt retrieval of the learned EVET errands list, but have less impact on the updating function required for on-line re-ordering of the learned list during task performance, so would have little impact on on-line plan adjustment by the poor plan group, but might affect retrieval of the learned plan by both groups. The expected data pattern in this case would be an effect of secondary task on overall performance, but not on plan following. These alternative predictions were addressed in Experiment 2, in which participants performed EVET on its own, and concurrently with oral random generation of months of the year.

Method

Participants

Participants were 32 undergraduates (21 female, 11 male) at the University of Edinburgh, with a mean age of 20.7 years (SD = 2.6 years). They had not previously taken part in Experiment 1, or in previously published experiments using EVET, and they received course credit in exchange for participation.

Design, tasks and procedure

Experiment 2 followed the same design and procedure as Experiment 1, but in this case participants were asked to randomly generate months of the year as a secondary task, at the rate of one per second. Random generation was also performed on its own for 120 seconds to provide a single task baseline measure of randomization performance. The oral random responses were recorded to allow analysis both of inter-utterance intervals and degree of randomness in the sequences generated.

Results

EVET Score

Table 3 shows a clear drop in performance for both good plan and poor plan groups in the dual-task condition. Data were analysed using a 2x2 mixed ANOVA, which confirmed a significant main effect of demand, *F*(1,30) = 24.22, *p*< 0.001, partial η^2= .45, but no significant main effects of plan, *F*< 1. There was no significant interaction between these factors, *F*< 1. Therefore, the concurrent demand of random generation had a detrimental effect on EVET score, regardless of the type of plan that participants had memorised.

Task completions, bonuses and penalties

The detailed breakdown on task completions, bonuses and penalties for each group in each condition is shown in Table 4. There appears to be little difference between the good and poor plan groups, but the scores do tend to be lower in the dual-task condition than the single-task condition. Again, normality was violated in almost all cases so non-parametric tests were utilised. Wilcoxon tests showed that single task performance was significantly better than dual task performance for the number of task completions, *Z* = -2.342, *p* = 0.019, and the number of bonuses awarded, *Z* = -2.205, *p* = 0.027. There was no significant difference in terms of the number of penalties awarded, *Z* = -0.948, *p* = 0.343. When collapsed across demand, there was no significant difference between the good and poor plan groups on any of the three measures as assessed by Mann-Whitney tests (all ps > 0.3).

Plan Following

Plan following scores were calculated in the same way as for Experiment 1, and are displayed in Table 3. A 2x2 ANOVA confirmed a significant main effect of plan group, F (1,30) = 17.471, p < 0.001, partial η^2 = 0.037, but there was no significant main effect of demand, F < 1, and no interaction, F < 1. As in Experiment 1, participants in the poor plan group tended to engage in more on-line plan adjustment of the errand list than participants in the good plan group, who tended to follow the memorised sequence more closely. This was the case in both the single and dual task conditions.

Post-EVET recall of the errands

Overall recall of the errands after the EVET was again very high (single task condition: Mean = 93.94%, SD = 7.57% and dual task condition: Mean = 92.26%, SD = 10.33%). For the good plan group, the percentage of errands recalled in the original serial position was 65.57% (SD = 34.88%) in the single task condition, which was almost identical to the dual-task condition at 65.57% (SD = 34.34%). For the poor plan group these percentages were lower, at 45.94% (SD = 37.23%) in the single-task condition and 47.19% (SD = 39.26%) in the dual-task condition. This trend did not reach significance, F(1,27) = 2.273, p = 0.143, partial η^2 = 0.075, and there was also no significant main effect of demand and no interaction, Fs < 1.

Randomness of secondary task responses

Redundancy, Random Number Generation (RNG) score and Ascending Adjacency were calculated using the computer programme RGCalc (Towse & Neil, 1998), and mean values are shown in Table 5. Redundancy expresses the extent to which participants sample equally from all the possible response alternatives, with lower scores indicating more equal sampling. RNG is based on the frequency with which particular response pairings occur in the data and varies between 0 and 1, with lower scores indicating greater equality of possible response pairings. Ascending adjacency examines the frequency with which participants followed a response with the next month in calendar order – i.e., the over-learned, stereotyped response that they should be trying to suppress.

[Table 5 about here please]

Baseline performance was compared to dual-task performance using mixed ANOVA (the between-participants factor being plan group). For Redundancy, there was no main effect of demand, no effect of group and no interaction (all Fs < 1). For RNG however, there was a highly significant main effect of demand, *F*(1,30) = 180.95, *p*<0.001, partial η^2= .86, indicating that participants were significantly less random in terms of the ordering of their responses during the dual-task condition than during baseline random generation performance. There was no main effect of plan group, *F*(1,30) = 1.16, $p>0.2$, partial $\eta^2 = 0.04$, and no interaction, $F< 1$, in the RNG scores. The ascending adjacency measure indicates that randomness suffered because participants often failed to inhibit the calendar order. They were also significantly less random on this measure during the dual-task condition, *F*(1,30) = 34.44, $p< 0.001$, partial $\eta^2 = 0.53$, but plan group had no effect, *F*(1,30) = 1.21, $p> 0.2$, partial η^2,= 0.04 , and there was no interaction, $F< 1$.

Participants were slower at generating months of the year under dual-task conditions (one word every 2.01 seconds) than during practice (one word every 1.36 seconds), a difference that was significant, t(30) = -5.671, $p< 0.001$. The poor plan and good plan groups did not differ in their mean rate of random generation while performing the EVET, t(30) = 0.711, p > 0.4.

Discussion

As in Experiment 1, participants in the good plan group were more likely to perform the errands in the well-learned order, than participants in the poor plan group, who were more likely to change the learned order of the errands on-line. However, there was no effect of demand, suggesting that the cognitive processes involved in oral random generation of months of the year did not overlap with those used for on-line plan adjustment of the order in which they completed the task list.

The random generation task did have a clear effect on overall performance in terms of EVET score. Participants also generated sequences of months that were less random than those they were capable of producing at baseline, demonstrating that they were not protecting performance on the generation task at the expense of EVET. This drop in randomness occurred despite a rate of utterance that was significantly slower at test than at practice. Therefore, the drop in randomness cannot be attributed to a trade-off with response time. There was clear resource-competition between the tasks, a finding that is consistent with Law et al. (2006). In the prior study, the greater interference of random generation (relative to articulatory suppression) was seen in the secondary but not primary task data. However, the method of scoring the prior version of the virtual errands test (task completions minus errors) was cruder than the weighted scoring procedure adopted for the EVET, and may have been less sensitive to the effects of the secondary task. In the present study a more detailed analysis of how the individual components making up the overall score were affected by the secondary task, showed that random generation caused people to complete fewer tasks, and apply a less efficient strategy (resulting in fewer bonus points), but it did not cause them to make more errors such as entering an incorrect room or picking up a lure object.

These results are consistent with our expectations that the task requirements of oral random generation for retrieval of well learned items from long-term memory, coupled with inhibition of learned sequences, may be interfering with access to the learned sequence of errands while performing EVET, leading to poorer scores overall. The fact that the poor plan group undertook planning adjustment on-line, but was no more affected by random generation than the good plan group, suggests that random generation does not affect the updating and re-ordering process in working memory. It also suggests that random generation is not having a general effect on a limited capacity attentional system, but is having specific effects on selected and specific resources that are part of working memory function. This is consistent with the suggestion that on-line plan adjustment involves a different set of working memory functions than does oral random generation. It is also consistent with the conclusions from Experiment 1 that multitasking as studied here may involve selection of task-relevant, specific cognitive functions from a range of cognitive functions available for deployment within working memory.

If multitasking involves multiple cognitive functions acting in a co-ordinated way (Burgess et al., 2000, Logie, 2011; Logie et al., 2011), this may give participants the flexibility to adapt their approach to the task depending on the circumstances. When verbal rehearsal, retrieval from long-term memory and inhibition of prepotent responses are engaged with a secondary task like oral random generation, participants may be able to rely to a greater extent on visuo-spatial working memory (Logie, 1995) to assist with performance of EVET. Indeed, Logie et al. (2011) found that an independent measure of visuo-spatial working memory ability correlated with EVET performance, whereas a measure of verbal working memory did not, suggesting that the former is being used at the limits of its capacity, whereas the latter may be have a much less key role, perhaps contributing at well within its capacity. Phillips, Wynn, Gilhooly, Della Sala & Logie (1999) found that articulatory suppression actually enhanced performance on the Tower of London planning task, and concluded that it prevented participants from applying unhelpful verbal strategies to a task that is essentially visuo-spatial in nature. When participants in our poor plan group re-arranged their plan on-line they may have been able to use visuo-spatial resources to assist with this process despite the load on verbal and other resources within working memory. Therefore, Experiment 3 used a dual task paradigm to investigate the impact of visuo-spatial working memory load in on-line plan adjustment and EVET performance.

Experiment 3

Successful EVET performance involves navigating in a 3D virtual world displayed in perspective on a 2-D computer screen. It is therefore likely to draw on the resources of visuo-spatial working memory, as do other navigation tasks (e.g., Baumann, Skilleter & Mattingley, 2011; Deyzac, Logie & Denis, 2006; Garden, Cornoldi & Logie, 2002; Meilinger, Knauff & Bültoff, 2008). Baumann et al. (2011) and Meilinger et al. (2008) both showed that a spatial localisation secondary task disrupted performance on a primary navigation task more than a visual secondary task (with little spatial demand). Furthermore, individual differences in spatial working memory predicted EVET performance in Logie et al. (2011). In EVET, a secondary task with a spatial demand might therefore be expected to have a stronger disruptive effect for on-line plan adjustment than did the verbal secondary task of articulatory suppression. The secondary task in Experiment 3 required participants to localise the source of auditory tones emitted at a constant rate as they performed EVET.

Method

Participants

The participants were 33 undergraduates (17 male, 16 female) from the University of Edinburgh, with a mean age of 20.36 years (SD = 3.26), who had not taken part in the previous experiments. They received course credit for participation. One participant was found to have performed at below-chance levels on the sound localisation task, and was therefore excluded from all analyses.

Design, Tasks and Procedure

This experiment differed from the previous two only in respect of the secondary task. Participants were asked to localise the source of auditory tones emitted at the rate of approximately one every three seconds. Six speakers were located respectively in front, behind, above and below the participant, and on their left and right, and these emitted a tone in a pseudo-random order. Each tone was presented for 500 milliseconds at 70dB with an inter trial interval of 2.5 seconds. Participants gave an immediate verbal report indicating which speaker had emitted the tone.

Results

EVET Score

Table 3 shows the mean EVET scores in single and dual-task conditions for both the good plan and poor plan groups. Performance was lower in the dual-task condition for both groups, a main effect that was confirmed by a 2x2 mixed ANOVA, $F(1,30) = 25.922$, $p < 0.001$, partial $\eta^2 = 0.46$. However, there was no significant difference between the groups, $F < 1$, and no interaction, $F(1,30) = 1.433$, $p > 0.2$.

Task completions, bonuses and penalties

These data are shown in Table 4. Again the scores of the good and poor plan groups are similar, with performance tending to be worse in the dual-task condition than in the single-task condition. In this experiment, Wilcoxon tests showed a significant effect of the secondary task not only on task completions, $Z = -3.110$, $p = 0.002$ and bonus points awarded, $Z = -2.975$, $p = 0.003$, but also on number of penalty points, $Z = -2.815$, $p = 0.005$. In contrast, there was no effect of group on any of these three measures, all *p*s > 0.3, as determined by Mann-Whitney tests.

Plan Following

As in the previous two experiments, participants in the good planning group adhered more closely to the plan they memorised at the start of the session than did the poor planning group (see table 3). A 2 x2 ANOVA confirmed a significant main effect of plan group, $F(1,30) = 34.082$, $p< 0.001$, partial $\eta^2 = .532$, and no significant main effect of demand, $F < 1$, or interaction, $F < 1$.

Post-EVET recall performance

There was again a clear ceiling effect in the overall recall of the errands after EVET performance in the single task condition (M = 97.00%, SD = 3.21%) and dual task condition (M = 97.38%, SD = 4.26%). In terms of whether errands were recalled in their original serial position, there was a very clear difference between the good and poor plan groups in this experiment. For the good plan group, 77.54% (SD = 25.33%) were recalled in their original position in the single-task condition, and 72.19% (SD = 31.24%) in the dual. For the poor plan group, these figures were 24.68% (SD = 27.48%) and 27.27% (SD = 34.01%) for the single and dual task conditions respectively. The main effect of plan group was highly significant, $F(1,29) = 27.73$, $p < 0.001$, partial $\eta^2= 0.489$, but there was no main effect of demand or interaction, Fs < 1.

Spatial localisation performance

Performance on the spatial localisation task was calculated as the percentage of correct responses to sounds presented during baseline and dual-task test periods. Baseline performance (*M*=61.01, *SD*= 9.57) was significantly better than dual-task performance (*M*=52.17, *SD*=14.22), $t(31) = -4.062$,

$p < 0.001$. The good plan and poor plan groups did not differ significantly under dual-task conditions, $t(30) = -0.889$, $p > 0.3$.

Discussion

The hypothesis that concurrent performance of the spatial localisation task would lead to greater plan following, especially for the poor plan group, was not supported, as plan following behaviour was similar across single and dual-task conditions. Good plan and poor plan participants also achieved a similar level of performance on the spatial localisation task.

However, this secondary task clearly had an overall disruptive impact on EVET score. It is notable that single-task performance in this experiment was higher than in Experiments 1 and 2. This can be explained by the fact that participants in this study were particularly highly motivated and organised undergraduates who were available for research participation during the exam period. Nevertheless their performance was clearly impaired by concurrent spatial localisation. An additional finding was that participants scored more penalties under dual-task conditions in Experiment 3. Penalties are mainly awarded for navigational failures such as entering an incorrect room or travelling the stairways in the wrong direction. While the rate of these was increased by the spatial concurrent task, this pattern was not seen when verbal rehearsal, long-term memory retrieval and inhibitory functions were loaded in Experiment 2. As in Experiment 2, the secondary task load also decreased the number of tasks completed and bonuses awarded.

General Discussion

The present study was aimed at determining which working memory resources are critical for on-line plan use, re-ordering, and manipulation during multitasking as measured by a virtual multiple errands test (EVET). An important finding was that participants were able to use flexible, on-line plan adjustment processes to re-organise pre-learned sequences of errands into a different order. Participants who memorised the errands in a sub-optimal (*poor*) order were more likely to change the order of errand execution, as demonstrated by their lower plan following scores. They also achieved EVET scores that were just as good as participants who memorised an optimal errand sequence from the outset. This was the case regardless of whether or not a secondary task was being performed at the same time, and regardless of which secondary task was involved. High levels of recall for the errands after EVET was completed suggest that task failures and errors were generally not due to forgetting of the list. Across the three experiments there was a general tendency for the poor plan group to change the order in which they recalled the errands (relative to the original) more than did the good plan group. This is consistent with the fact that the poor plan group were also more likely to change the order in which they executed the tasks. Taken together, the plan following and recall order data show that participants' internal representation of the task list underwent a transformation from what was initially learned – a re-ordering and adjustment of the sequence of task goals. We have referred to this process as on-line plan adjustment.

Overall performance in terms of score was impaired by a secondary task that involved generating a stream of randomised responses (Experiment 2) and making a judgement about the spatial location of sounds (Experiment 3). In contrast, the results of Experiment 1 showed that participants were able to maintain EVET performance under conditions of articulatory suppression, with no difference between single and dual task performance in terms of EVET score. Taken together, the results of three experiments suggest that successful completion of EVET draws heavily on visuo-spatial working memory, long-term memory retrieval and inhibition of prepotent responses, but does not depend on unhindered access to verbal working memory for the purposes of sub-vocal rehearsal.

The process of on-line plan adjustment in multitasking appears to rely on additional resources within working memory, which might involve maintaining and updating current goals. As none of the secondary tasks affected plan-following, it might be concluded that on-line plan adjustment does not involve the resources of working memory. However, we think it is more likely that the cognitive system is able to deploy flexibly the resources of working memory according to the demands of the situation (Logie, 2011). When people had a verbal load in Experiments 1 and 2, they may have used spatial working memory to assist with the re-ordering of the errands according to an internal map. When people had a spatial load in Experiment 3, they may have relied more on verbal resources or executive functions in working memory to assist with on-line plan adjustment. This hypothesis is consistent with the data pattern observed, but would merit investigation in future studies. For example, it would predict that verbal working memory capacity might predict EVET performance only when EVET is performed concurrently with a spatial load.

In any case, it seems that having a good plan in advance of the task is not a necessary requirement for successful multitasking, as assessed by the virtual multiple errands methodology. This aptitude for on-line plan adjustment is a finding that is consistent with research using the Tower of London task (Phillips, et al., 2001), where pre-planning was not found to result in faster or more accurate performance of the task than planning on-line with no pre-planning. Those authors also point out that while pre-plans in Tower of London and Tower of Hanoi tasks may rely on verbal rehearsal, visuo-spatial codes may be more important during task execution. This fits with their earlier finding (Phillips et al., 1999) that while articulatory suppression reduced time spent pre-planning on the Tower of London task, it actually led to quicker execution times (with no significant change in accuracy). During errand-based multitasking, in EVET but more importantly in everyday environments such as shopping centres or office buildings, a list of tasks is learned and rehearsed verbally, but when sufficiently well learned, retrieval from long-term memory and visuo-spatial working memory resources are crucial when putting the plan into action.

The finding from the performance data that EVET taps executive resources is in line with previous research from the neuropsychology literature, which shows that patients with executive dysfunction have multitasking difficulties both in everyday life and during laboratory tests (Alderman et al., 2003; Burgess et al., 2000; Crépeau, Belleville, Duchesne, 1996; Fortin, Godbout & Braun, 2003; Goldstein, Bernard, Fenwick, Burgess, & McNeil, 1993; Knight et al., 2002; Law et al., 2004; Levine, Stuss, Milberg, Alexander, Schwartz & MacDonald, 1998; Levine, Dawson, Boutet, Schwartz & Stuss, 2000; McGeorge et al., 2001; Rand et al., 2009; Shallice & Burgess, 1991). The present results are also broadly compatible with the findings of Law et al. (2006), who also demonstrated some conflict between random generation and virtual errands performance. However, in the present study the conflict is evident in both the primary and secondary task data, rather than in the secondary task data alone, as was found in the earlier study. We have also suggested which specific executive resources might be involved.

We did not observe a decrement in EVET performance under conditions of articulatory suppression, therefore it appears that the task may be achievable with only limited sub-vocal rehearsal. Emerson and Miyake (2003) showed that articulatory suppression was less disruptive (in a task-switching paradigm) when participants were provided with explicit cues to guide switches, and more disruptive when participants had to rely on what they termed “internal self-cueing”. In EVET, there are some partial cues on screen that may help to support performance. For example, once an object has been collected a word label for the object appears on the side of the screen and remains there until it is delivered (see figure 1B). However, the participant has to retrieve from long-term memory, the correct destination for the object and the remaining errands to be completed. Future research using EVET could manipulate the availability of on-screen retrieval cues to further examine the role of internal self-cueing compared with external cueing, while investigations of the effects of unexpected interruptions could yield additional insight into on-line plan adjustment.

A possible caveat to our interpretation of these findings is that the secondary tasks used may have differed in overall levels of difficulty, in particular the articulatory suppression task may have been too easy. However, participants were asked to repeat a three syllable word at a demanding rate of once per second, and a range of previous studies have shown that articulatory suppression is highly disruptive of verbal immediate memory task performance (e.g. Baddeley, Lewis & Vallar, 1984; Murray, 1965), but not of non-verbal tasks (e.g. Saito, Logie, Morita & Law, 2008). So, its difficulty depends on the task with which it is combined, not on any overall demand on the cognitive system. Another possible limitation is that the findings might not generalise to real-world multitasking situations. However, EVET retains essential features of real errand-based multitasking situations such as, for example, a shopping trip (particularly one where time is limited) or a series of tasks being executed by an office worker during a sojourn from their desk. In these cases a list of tasks has to be mentally assembled, rehearsed and then executed. During execution, adjustments and task-ordering may be required, necessitating the use of on-line cognition as well as pre-planning processes. There is clearly much work yet to be done to understand the full complexity of how cognitive functions act in concert to support multitasking activities in everyday life, but we would argue that the research presented here yields important insights into the flexibility of on-line plan adjustment processes during errand-based multitasking, and suggests that multitasking performance is not wholly constrained by general attentional demands. We would argue that the use of a virtual environment strikes a balance, retaining a good degree of experimental control while achieving reasonable ecological validity. We also think that in terms of multitasking, there is an important distinction to be made between errand-based multitasking such as in the Multiple Errands Test and variants (including the research presented here), and 'table-top' multitasking where movement around an environment is not required, such as in the Six Elements Test and variants (both tests originally developed by Shallice & Burgess, 1991). Real-world corollaries of the latter type of laboratory test might include cooking (e.g. Craik & Bialystok, 2006), or a series of tasks being executed by an office worker at their desk. An important priority for future research will be to establish the extent to which both types of multitasking situation draw on the same cognitive resources.

In conclusion, our results point to the importance of retrieval from long-term memory and visuo-spatial working memory for errand-type multitasking, but suggest that the task is achievable without heavy reliance on sub-vocal rehearsal. Even under conditions of high concurrent demand, participants were able to implement flexible on-line plan adjustment and re-ordering processes. For participants given a poor plan at the outset, these processes may have allowed them to achieve comparable scores to participants given a good plan at the outset. These experimental results are consistent with our previous findings from a multivariate individual differences approach (Logie et al., 2011) suggesting that on-line plan adjustment and implementation of intentions can be identified as a separate latent variable from working memory and from forming an initial plan. The relative lack of disruption of on-line plan adjustment by demanding secondary tasks could therefore suggest that human adults spontaneously develop specific skills in on-line plan adjustment and re-ordering of everyday activities, and that these skills offer specific resources that can be recruited flexibly within a multiple resource working memory system (Baddeley & Logie, 1999; Logie, 2011; Logie & Niven, 2012). Our findings certainly indicate that on-line plan adjustment cannot easily be attributed to the operation of spatial or verbal working memory, but might draw on either, or on acquired everyday skills, depending on which resources are currently available and most effective for performing the current task (Logie, 2011). EVET and similar controlled and systematic research tools offer a promising means to yield additional insight into this ubiquitous but complex feature of everyday on-line cognition.

References

Alderman, N., Burgess, P. W., Knight, C., & Henman, C. (2003).Ecological validity of a simplified version of the multiple errands shopping test. *Journal of the International Neuropsychological Society, 9*, 31 - 44.

Baddeley, A.D., (1966). The capacity for generating imformation by randomization. *Quarterly Journal of Experimental Psychology*, 18, 119-129

Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 49*, 5-28.

Baddeley, A.D., Lewis, V.J. & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*, 36, 233-252.

Baddeley, A. D., & Logie, R. H. (1999). Working Memory: The Multiple Component Model. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 28-61). Cambridge: Cambridge University Press.

Baumann, O., Skilleter, A. J., & Mattingley, J. B. (2011). Short-term memory maintenance of object locations during active navigation: Which working memory subsystem is essential? PLoS One, 6, e19707. DOI:10.1371/journal.pone.0019707

Burgess, P. W. (2000). Strategy application disorder: The role of the frontal lobes in human multitasking. *Psychological Research-Psychologische Forschung, 63*, 279 - 288.

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

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

Crépeau, F., Belleville, S., & Duchesne, G. (1996). Disorganisation of behavior in a multiple subgoals scheduling task following traumatic brain injury. *Brain and Cognition, 32*, 266-268.

Cowan, N. (2005). *Working memory capacity*. Hove, UK: Psychology Press.

Dawson, D. R., Anderson, N. D., Burgess, P. W., Cooper, E., Krpan, K. M., & Stuss, D. T. (2009). Further development of the Multiple Errands Test: Standardized scoring, reliability, and ecological validity for the Baycrest version. *Archives of Physical and Medical Rehabilitation, 90*, S41 - S51.

Deyzac, E., Logie, R.H. & Denis, M. (2006). Visuospatial Working Memory and the Processing of Spatial Descriptions*. British Journal of Psychology,* 97, 217-243*.*

Emerson, M. J., & Miyake, A. (2003). The role of inner speech in task switching: A dual task investigation. *Journal of Memory and Language, 48*, 148-168.

Evans, F.J. (1978). Monitoring attention deployment by random number generation: An index to measure subjective randomness. *Bulletin of the Psychonomic Society*, 12, 35-38.

Fortin, S., Godbout, L., & Braun, C. M. J. (2003). Cognitive structure of executive deficits in frontally lesioned head trauma patients performing activities of daily living. Co*rtex, 39*, 273 - 291.

Garden, S. E., Cornoldi, C., & Logie, R. H. (2002). Visuo-spatial working memory in navigation. *Applied Cognitive Psychology, 16*, 35 - 50.

Garden, S. E., Phillips, L. H., & MacPherson, S. E. (2001).Midlife Aging, Open-Ended Planning, and Laboratory Measures of Executive Function. *Neuropsychology, 15*, 472-482.

Goldstein, L. H., Bernard, S., Fenwick, P. B. C., Burgess, P. W., & McNeil, J. (1993). Unilateral frontal lobectomy can produce strategy application disorder. *Journal of Neurology, Neurosurgery, and Psychiatry, 56*, 274-276.

Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in a synthetic work paradigm. *Applied Cognitive Psychology, 24*, 1149 - 1167.

Knight, C., Alderman, N., & Burgess, P. W. (2002).Development of a simplified version of the multiple errands test for use in hospital settings. *Neuropsychological Rehabilitation, 12*, 231-255.

Law, A. S. (2004). Human adult multitasking: Developing and applying a methodology. Unpublished PhD thesis. University of Aberdeen.

Law, A. S., Logie, R. H. & Pearson, D. G. (2006). The impact of secondary tasks on multitasking in a virtual environment. *Acta Psychologica, 122,* 27-44.

Law, A. S., Logie, R. H., Pearson, D. G., Cantagallo, A., Moretti, E., &Dimarco, F. (2004). Resistance to the impact of interruptions during multitasking by healthy adults and dysexecutive patients. *Acta Psychologica, 116*, 285-307.

Levine, B., Dawson, D., Boutet, I., Schwartz, M. L., & Stuss, D. T. (2000). Assessment of strategic self-regulation in traumatic brain injury: Its relationship to injury severity and psychosocial outcome. *Neuropsychology, 14*, 491-500.

Levine, B., Stuss, D. T., Milberg, W. P., Alexander, M. P., Schwartz, M., & MacDonald, R. (1998). The effects of focal and diffuse brain damage on strategy application: Evidence from focal lesions, traumatic brain injury and normal aging. *Journal of the International Neuropsychological Society, 4*, 247-264.

Logie, R.H. (1995). *Visuo-spatial working memory*. Hove, UK: Lawrence Erlbaum Associates.

Logie, R. H. (2011). The functional organisation and the capacity limits of working memory. *Current Directions in Psychological Science*, *20*, 240-245.

Logie, R. H., Law, A. S., Trawley, S. & Nissan, J. (2010). Multitasking, working memory and remembering intentions. *Psychologica Belgica, 50*, 309-326.

Logie, R.H. & Niven, E.H. (2012). Working Memory: An Ensemble of Functions in On-Line Cognition. In V. Gyselinck and F. Pazzaglia (Eds.). *From Mental Imagery to Spatial Cognition and Language. Essays in Honour of Michel Denis*, pp77-105. Hove, UK: Psychology Press.

Logie, R. H., Trawley, S. & Law, A. S. (2011). Multitasking: Multiple, domain specific cognitive functions in a virtual environment. *Memory and Cognition*. DOI: 10.3758/s13421-011-0120-1.

McGeorge, P., Phillips, L. H., Crawford, J. R., Garden, S. E., Della Sala, S., Milne, A. B., et al. (2001). Using virtual environments in the assessment of executive dysfunction. *Presence, 10*, 375-383.

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

Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. (2004). Inner speech as a retrieval aid for task goals: the effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica, 115*, 123-142.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100.

Murray, D. (1965). Vocalization-at-presentation, with varying presentation rates. *Quarterly Journal of Experimental Psychology*, 17, 47-56.

Phillips, L. H., Wynn, V., Gilhooly, K. J., Della Sala, S., & Logie, R. H. (1999). The role of memory in the Tower of London task. *Memory, 7*, 209-231.

Phillips, L. H., Wynn, V. E., McPherson, S., & Gilhooly, K. J. (2001). Mental planning and the Tower of London task. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 54*, 579-597.

Rajendran, G., Law, A. S., Logie, R. H., van der Meulen, M., Fraser, D. & Corley, M. (2011). Investigating multitasking in high-functioning adolescents with autism spectrum disorders using the Virtual Errands Task. *Journal of Autism and Developmental Disorders*, *41,* 1445-1454.

Rand, D., Basha-Abu Rukan, S., Weiss, P. L., & Katz, N. (2009). Validation of the virtual MET as an assessment tool for executive functions. *Neuropsychological Rehabilitation, 19*, 583-602.

Saito, S., Logie, R.H., Morita, A. & Law, A. (2008). Visual and phonological similarity effects in verbal immediate serial recall: A test with kanji materials. *Journal of Memory and Language*, 59, 1-17.

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

Towse, J. N., & Neil, D. (1998). Analyzing human random generation behavior: A review of methods used and a computer program for describing performance. *Behavior Research Methods Instruments & Computers, 30*, 583-591.

Trawley, S. L., Law, A. S. & Logie, R. H. (2011). Event based prospective remembering in a virtual world. *Quarterly Journal of Experimental Psychology*. DOI: 10.1080/17470218.2011.584976

Unsworth, N. & Engle, R.W. (2007) The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104-132.

Figure 1. Upper screenshot (A) of EVET building taken from the back wall looking towards the building entrance. Lower screenshot (B) taken during EVET test shows current time and inventory displays.

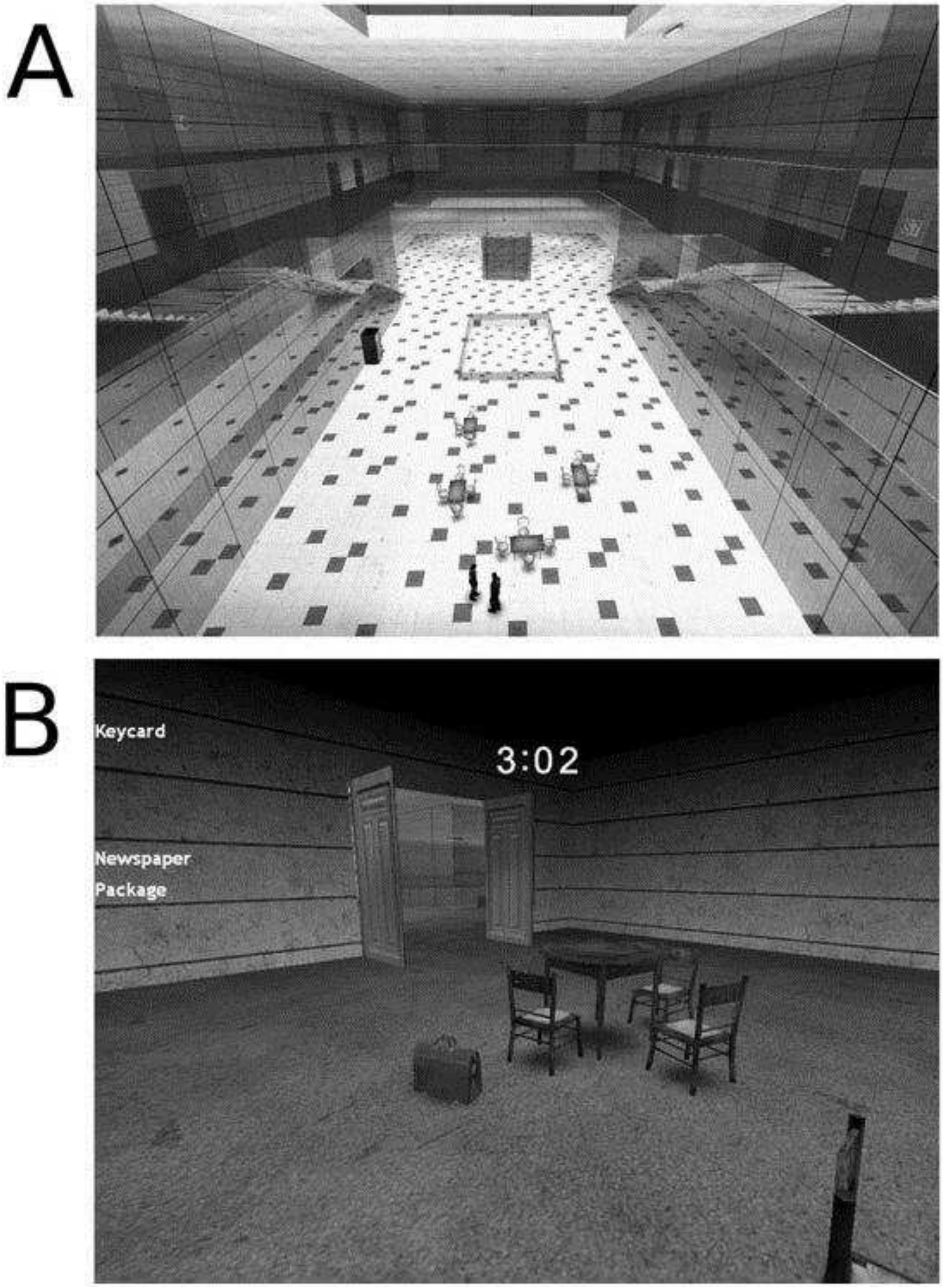
A
B
Keycard
3:02
Newspaper
Package

Table 1: EVET errand lists for good plan group (A1, B1) and poor plan group (A2, B2)

SET A1 - 8 MINUTES	**SET A2 - 8 MINUTES**
START GROUND FLOOR	START GROUND FLOOR
(1) Get stair code in G8 from noticeboard	**(1)** Get stair code in G8 from noticeboard
(2) Turn off lift on G-Floor	**(2)** Pickup brown package in T4
(3) Pickup newspaper in G3	**(3)** Get keycard in F9
(4) Drop newspaper off on desk in S4	**(4)** Use keycard to unlock G6 (via G5)
(5) Meet person in S10 before 3:00minutes	**(5)** Turn on Cinema in S7 at 5:30minutes
(6) Get keycard in F9	**(6)** Drop brown package in G6
(7) Pickup brown package in T4	**(7)** Sort red and blue binders in room S2
(8) Use keycard to unlock G6 (via G5)	**(8)** Meet person in S10 before 3:00minutes
(9) Drop brown package in G6	**(9)** Pickup newspaper in G3
(10) Turn on cinema in S7 at 5:30minutes	**(10)** Drop newspaper off on desk in S4
(11) Sort red and blue binders in room S2.	**(11)** Turn off lift on G-Floor.
G = Ground Floor, F = First Floor,	G = Ground Floor, F = First Floor
S = Second Floor, T = Third Floor	S = Second Floor, T = Third Floor
SET B1 - 8 MINUTES	**SET B2 - 8 MINUTES**
START THIRD FLOOR	START THIRD FLOOR
(1) Get stair code T10 from noticeboard	**(1)** Get stair code in T10 from noticeboard
(2) Turn off lift on T-Floor	**(2)** Pickup computer in G4
(3) Pickup milk carton in T3	**(3)** Get keycard in S9
(4) Take milk carton to desk in F4	**(4)** Use keycard to unlock T7 (via T6)
(5) Meet person F10 before 3:00minutes	**(5)** Turn on Cinema in F7 at 5:30minutes
(6) Get keycard in S9	**(6)** Drop computer in T7
(7) Pickup computer in G4	**(7)** Sort red and blue folders in F2

(8) Use keycard to unlock T7 (via T6)	**(8)** Meet person F10 before 3:00minutes
(9) Drop computer in T7	**(9)** Pickup milk carton in T3
(10) Turn on Cinema in F7 at 5:30minutes	**(10)** Drop milk cartoon off on desk in F4
(11) Sort red and blue folders in F2.	**(11)** Turn off lift on T-Floor.
G = Ground Floor, F = First Floor	G = Ground Floor, F = First Floor
S = Second Floor, T = Third Floor	S = Second Floor, T = Third Floor

Table 2. EVET point allocation (20 point maximum)

Bonus Weightings	4 +points	3 +points	2 +points	1 +points	0 points
Folder sorting	30+	23-29	15-22	8-14	1-7
Cinema (time discrepancy from 5:30min)	0-2sec	3-5sec	6-7sec	8-10sec	11+
Meeting (time discrepancy over 3:00min)	<3mins	1-12sec	13-25sec	26-37sec	38+sec
Penalty Weightings	4 -points	3 -points	2 -points	1 -points	0 points
Picking up objects not on list	4+	3	2	1	0
Entering rooms not on list	4+	3	2	1	0
Breaking stair rule	5+	4	3	2	1

Table 3: Mean EVET and plan following scores in Experiments1, 2 and 3 (standard deviation in parentheses)

		Good Plan		Poor Plan	
		EVET Score	Plan Following	EVET Score	Plan Following
Experiment 1	Single Task Performance	10.60 (5.35)	0.91 (0.11)	10.05 (4.97)	0.35 (0.32)
	Dual Task Performance	10.55 (4.27)	0.91 (0.11)	10.50 (4.27)	0.40 (0.34)
Experiment 2	Single Task Performance	11.44 (5.48)	0.73 (0.28)	12.06 (3.71)	0.28 (0.25)
	Dual Task Performance	6.38 (4.69)	0.68 (0.30)	7.00 (5.44)	0.30 (0.38)
Experiment 3	Single Task Performance	14.53 (4.90)	0.79 (0.31)	13.53 (3.64)	0.10 (0.44)
	Dual Task Performance	8.82 (5.37)	0.81 (0.25)	10.00 (5.16)	0.11 (0.38)

Table 4: Mean number of Task Completions, Bonus and Penalty points awarded in Experiments 1, 2 & 3 (standard deviation in parentheses)

		Good Plan			Poor Plan		
		Tasks (max=8)	Bonus Points (max=12)	Penalty Points (max=12)	Tasks (max=8)	Bonus Points (max=12)	Penalty Points (max=12)
Experiment 1	Single Task	7.15 (1.42)	5.75 (2.65)	1.80 (2.76)	7.30 (1.03)	4.40 (3.36)	1.35 (1.87)
	Dual Task	7.50 (1.00)	4.85 (2.28)	1.20 (1.54)	7.20 (1.44)	5.35 (3.01)	1.45 (1.57)
Experiment 2	Single Task	7.31 (1.45)	4.56 (3.40)	1.75 (1.77)	7.06 (1.34)	5.63 (2.91)	1.44 (1.93)
	Dual Task	5.81 (1.76)	3.50 (3.11)	1.81 (2.34)	6.50 (2.37)	3.94 (2.81)	2.44 (2.45)
Experiment 3	Single Task	7.76 (0.56)	7.41 (3.50)	0.88 (1.45)	7.33 (0.98)	6.60 (2.64)	0.93 (1.16)
	Dual Task	6.71 (1.65)	4.88 (2.39)	2.06 (2.38)	6.40 (1.45)	6.13 (2.88)	2.27 (2.12)

Table 5: Randomness measures from secondary task data in Experiment 2 – Mean with standard deviation in parentheses

	Redundancy		Random Number Generation		Ascending Adjacency	
	Baseline	Test	Baseline	Test	Baseline	Test
Good Plan	**2.94** (2.61)	**3.03** (1.77)	**0.26** (0.07)	**0.41** (0.09)	**14.34** (4.53)	**23.49** (8.81)
Poor Plan	**2.63** (2.19)	**2.82** (1.52)	**0.28** (0.06)	**0.45** (0.10)	**15.95** (6.52)	**27.15** (11.83)