MEMORY &COGNITION

Multitasking: Multiple, Domain-Specific Cognitive Functions in a Virtual Environment

SCHOLARONE™
Manuscripts

1

Multitasking: Multiple, Domain-Specific Cognitive Functions in a Virtual Environment

Robert H. Logie, Steven Trawley

University of Edinburgh, UK

Anna Law

Liverpool John Moores University

Corresponding Author:

Robert H Logie

Centre for Cognitive Ageing and Cognitive Epidemiology

University of Edinburgh

7 George Square

Edinburgh

EH89JZ

Scotland

United Kingdom

e-mail: rlogie@staffmail.ed.ac.uk

Abstract

Multitasking among three or more different tasks is a ubiquitous requirement of everyday cognition, yet rarely is addressed in research on healthy adults who have had no specific training in multitasking skills. Participants completed a set of diverse subtasks within a simulated shopping mall and office environment, the Edinburgh Virtual Errands Test (EVET). The aim was to investigate how different cognitive functions such as planning, retrospective and prospective memory, and visuo-spatial and verbal working memory contribute to everyday multitasking. Subtasks were chosen to be diverse, and predictions were derived from a statistical model of everyday multitasking impairments associated with frontal lobe lesions (Burgess et al., 2000). Multiple regression indicated significant independent contributions from measures of retrospective memory, visuo-spatial working memory and on-line planning, but not from independent measures of prospective memory or verbal working memory. Structural Equation Modelling showed that the best fit to the data arose from three underlying constructs, with Memory and Planning having a weak link, but with both having a strong directional pathway to an Intent construct that reflected implementation of intentions. Participants who followed their pre-prepared plan achieved higher scores than those who altered their plan during multitask performance. This was true regardless of whether the plan was efficient or poor. Results substantially develop and extend the Burgess et al. (2000) model to healthy adults, and yield new insight into the poorly understood area of everyday multitasking. Findings also point to the utility of using virtual environments for investigating this form of complex human cognition.

Everyday multitasking among three or more different tasks pervades domestic and working life, yet rarely is the focus of empirical study or theory development. This form of multitasking refers to the management of several different on-going tasks to be completed within a limited time. Tasks often must be interleaved efficiently to maximise performance (Burgess, 2000), and there may be an optimum order for their completion, for example cooking a meal (e.g. Craik & Bialystok, 2006) or a time-limited shopping trip (e.g. Shallice & Burgess, 1991). Many occupations require multitasking, for example emergency medicine and medical decision making (Chisholm, Dornfeld, Nelson & Cordell, 2001; Law et al., 2005; van der Meulen et al., 2010), management (Seshadri &Shapira, 2001), or navigation (e.g. Garden, Cornoldi & Logie, 2002; Law, Logie & Pearson, 2006; Spiers & Maguire, 2006). We report here an investigation of how healthy young adults achieve everyday multitasking in a simulated environment. We start with a statistical model of everyday multitasking derived from studies of individuals with frontal lobe damage (Burgess, Veitch, de Lacy Costello & Shallice, 2000). We assess the generality of this model, and develop it for healthy young adults using a new paradigm designed to address this ubiquitous requirement of everyday cognition.

Most current theories of cognition address specific components of the cognitive system such as visual or auditory perception, focused or sustained attention, prospective memory, verbal or visuo-spatial working memory, episodic and semantic memory, task switching and cognitive bottlenecks. These topics are experimentally tractable and draw on well developed paradigms, allowing major advances in the understanding of each. Everyday multitasking requires co-ordinated and strategic deployment of several different cognitive functions, presenting a major challenge for paradigms and theories that focus on specific cognitive phenomena, and ensuring its rarity in studies of human cognition (Neisser, 1978). So, the research we describe is not concerned with understanding a particular cognitive function, or with cognitive bottlenecks between conflicting tasks (e.g. Borst, Taatgen & Rijn, 2010b). Rather, it addresses how the cognitive system avoids cognitive bottlenecks to perform multiple tasks (e.g. Craik & Bialystok, 2006). This requires new paradigms and development of theories concerned with how various parts of the cognitive system operate in concert rather than in isolation.

Studies of multitasking typically focus on deficits following brain injury (e.g., Alderman, Burgess, Knight & Henman, 2003; Miotto & Morris, 1998; Shallice & Burgess, 1991), on the

neuroanatomical correlates in healthy participants (Borst, Taatgen, Stocco & van Rijn, 2010a; Burgess, Dumontheil & Gilbert, 2007; Spiers & Maguire, 2006), or on highly trained experts such as in the military or aviation (e.g. Loukopoulos, Desmukes & Barshi, 2009; Wickens, 2008). Everyday multitasking in untrained adults is different from the microstructure of rapid switching between laboratory tasks (e.g. Koch, Gade, Schuch & Philipp, 2010; Monsell, 2003), or from concurrent dual-task demands (e.g. Borst et al., 2010a; 2010b; Logie, Cocchini, Della Sala & Baddeley, 2004a). Instead, it involves several subtasks that have different requirements, and participants decide themselves how to schedule subtask attempts. Other studies have assessed driving skills (e.g. Levy & Pashler, 2008; Strayer, Drews & Crouch, 2006), or planning and implementation of subgoals such as in the Tower of London task or simulated work (e.g. Hambrick, Oswald, Darowski, Rench & Brou, 2010; Phillips, Gilhooly, Logie, Della Sala & Wynn, 2003; Ward & Allport, 1997). However, these tasks do not address the broader demands of everyday multitasking, and despite its ubiquitous everyday requirement, there remains limited theoretical or empirical insight into how multitasking is achieved by healthy adults.

One theoretical approach to multitasking (Burgess et al., 2000) was derived from studies of individuals with frontal lobe damage, who often show impairments in everyday multitasking but intact performance on tests of attention, memory and executive functions. This suggests multitasking is not solely dependent on the latter functions. Burgess et al. (2000) used the 'Greenwich test', in which participants switch between three manual sub-tasks (sorting beads, sorting tangled lines on paper, and constructing plastic Meccano). Their statistical model (Figure 1) identified important, and largely independent roles for retrospective memory and intentionality (prospective memory). Planning did not reliably contribute to the model. However, planning deficits only appeared for patients with lesions in the right dorsolateral prefrontal cortex. Damage to anterior regions, Brodmann areas 8, 9 and 10, did not affect planning, but did affect task switching and task rule adherence. For these reasons, Burgess et al. included planning in their structural equation model, but they had only one measure of planning, so loading for the 'Plan' construct was set at '1.0'. Their model suggested that the 'Memory' construct drives separate constructs for 'Plan' and 'Intent'.

Figure 1 about here

Burgess, Simons, Coates and Channon (2005) suggested that planning is multifaceted and supported by a range of cognitive abilities. Phillips et al. (2003) noted that many healthy participants plan on-line during complex tasks rather than following a plan formed in advance. Therefore it would also be important to identify the individual cognitive functions that contribute to advance and on-line planning. Burgess et al. argued that forming and implementing a plan draws upon cognitive systems responsible for remembering instructions and rules for the multitasking paradigm. Therefore, planning and intentionality were represented downstream from retrospective memory in their model.

The roles for advance and on-line planning may be greater for subtasks with an optimum order, such as preparing a meal. Craik and Bialystok (2006) addressed this in a cognitive aging study using simulated breakfast making. Participants repeatedly set a simulated table by moving cutlery and plates on the computer screen, switching to alternate screens for starting and stopping preparation of foods with different cooking times. However, the reliance on prospective memory for starting and stopping foods, and similarity of sub-tasks, make it less well suited for multitasking assessment, which we assume to involve a range of different cognitive functions, only one of which is prospective memory.

The Multiple Errands Test (MET) (Shallice and Burgess, 1991; Alderman et al., 2003), has an optimum task order. Individuals with frontal lobe damage and healthy controls planned and attempted several tasks in a real shopping mall, following as efficient a route as possible. Although the MET is close to real life, and requires little or no initial practice, there are obvious drawbacks in real-life settings (e.g. Bailey, Henry, Rendell, Phillips & Kliegel, 2010). It is time consuming, requires transport for participants, and consent from local businesses. The lack of experimental control can compromise participant safety and data reliability, and tasks cannot easily be adapted for other clinical or research settings.

Simulated real-life tasks have been used for assessing brain damaged patients, including planning on a map (e.g. Burgess et al. 2005), moving furniture in virtual buildings (Morris, Kotitsa & Bramham, 2005), or selecting stores in a video of a high street (e.g. Knight, Titov & Crawford, 2006). Knight and Titov (2009) and others (e.g. Burgess et al., 2005; Bailey et al., 2010) note the need for ecologically valid and scientifically robust tests of multitasking. McGeorge et al. (2001) created a virtual version of the MET, the Virtual Errands Test (VET) that retained many advantages of the real environment while achieving experimental control.

The errands were tasks such as collect a book or meet a colleague. The VET was as sensitive as the real environment to executive dysfunction in brain damaged patients (see also Rand, Basha-Abu Rukan, Weiss & Katz, 2009). Thus, virtual environments may offer sufficient face and content validity for assessment of multitasking in a tractable and controlled setting. However, these studies have focused on impairments associated with brain damage (Knight & Titov, 2009). Law et al. (2006) modified the VET to be challenging for healthy adults. However, the graphics were unrealistic, and performance assessment involved video recording test sessions with subsequent manual scoring. The Edinburgh Virtual Errands Test (EVET), used in the current study, was developed from a widely available and inexpensive commercial games platform that permits non-profit development of virtual environments. It is well suited to creating an environment for multitasking with realistic graphics and a smooth interface as well as automatic recording of multiple performance measures.

In sum, little is known about how healthy young adults behave in everyday multitasking situations, and what cognitive variables affect their performance . We addressed the cognition that underlies this important activity, using sets of multiple errands performed by healthy young adults in a realistic simulation of everyday multitasking. EVET involves a higher memory load than did Burgess et al. (2000), because participants have to memorise an errand list. EVET requires a substantial degree of pre-planning, because inefficient ordering of the errands may result in time expiring before all tasks are completed. It requires navigation around a virtual environment, so participants may draw on the resources of visuo-spatial working memory (Logie, 1995). Therefore we expected that the Burgess et al. model might require modification to account for multitasking performance in healthy participants.

One additional important aspect of cognition is working memory, and König et al. (2005; see also Hambrick et al., 2010) found that working memory was a more important predictor of multitasking performance than fluid intelligence or attention. Baddeley and Logie (1999; Logie, in press) argued that working memory comprises multiple, domain-specific resources that are deployed selectively by participants according to task demands. This contrasts with the view that working memory is primarily a control system for focusing limited capacity domain-general attention on currently activated contents of episodic or semantic memory (e.g. Cowan, 2005). Therefore, we used independent measures of verbal and spatial working memory, and of verbal free recall to explore whether a single or a multiple factor model could best account for our multitasking data.

Participants completed an additional test of planning within EVET, and a battery of cognitive assessments to investigate relationships between individual differences in multitasking and separate measures of retrospective and prospective memory, on-line planning, and spatial and verbal working memory. From the Burgess et al. (2000) model and from our own previous work (e.g. Garden et al., 2002; Law et al., 2006; Logie, Baddeley, Mane, Donchin & Sheptak, 1989), we expected that retrospective memory and prospective memory, but also on-line planning (rather than pre-planning), spatial working memory and verbal working memory would make independent contributions to shared variance with EVET performance.

In summary, our goals were:

(a) To add to understanding of everyday multitasking in young healthy adults.

(b) To test the generality in healthy young adults of a statistical model originally developed to account for multitasking impairments associated with frontal lobe damage.

(c) To test the hypothesis that multitasking among several diverse subtasks draws on a range of several different cognitive functions, not only a single, general purpose attentional capacity.

(d) To develop a new methodology to address the above given that most current paradigms in experimental cognitive psychology would be ill suited to these goals.

Method

Participants

A total of 165 students at the University of Edinburgh (102 women, 63 men), mean age 19.59 (sd 2.43, range 16-32) participated.

Tests and Procedure

Participants first completed the EVET procedure and then five individual difference measures, administered in the order described below. Except for word recall, tasks were viewed on a 42

cm color monitor, on a Dell XPS PC with Intel Core Quad 2.33Ghz processor and 1GB ATI Graphics Card.

The Edinburgh Virtual Errands Task (EVET) environment was created using the Hammer environment editor supplied as a software development kit with the computer game Half Life 2™ to construct a 3D model of a four storey building with five rooms along the left and right of each floor around a central stairwell, with two sets of stairs (one left, one right) and a central elevator. Figure 2 shows a screen shot of the concourse on the ground floor (floor zero).

Figure 2 about here

Participants explored the virtual environment using the standard keyboard - w, s, a, d - for forward, backward, and lateral movement, and the mouse to look in any direction. Key "e" was used for actions such as picking up objects or opening doors. Participant movement within the virtual building was automatically recorded at 10Hz, represented as a series of XYZ coordinates, with actions recorded and time stamped. Participants were to complete eight errands within eight minutes (Table 1). Three of the errands were two-stage, requiring object collection and drop off. The remaining five required one action. Two tasks had time constraints (e.g. turn off cinema at 5:30). Sorting folders was the only open ended EVET task. It could be performed at any time for as long as the participant chose during the eight minute test period. Participants were informed that folder sorting was no more important than other tasks, but they should try to sort as many as possible. Half of the participants were given errand list A and started on the ground floor between the stairwells. The other half were given errand list B and started in the equivalent position on the top floor. Errands were given in a non-optimal order for completion, and participants were asked to plan the optimal order before commencing EVET.

Table 1 about here

First, each participant read EVET instructions detailing the task, building layout and rules. The experimenter described the layout of the building prior to the practice session. Building rules were to use left stairs for travelling down and right stairs for travelling up, to avoid entering non-task related rooms and to avoid picking up non-task related objects. Participants

practiced using the controls to move around the building to complete five practice errands; object collection and delivery, button pressing, unlocking stairwell door with key-code, and folder sorting. This also allowed participants to familiarise themselves with the building. None of the practice errands were used in the main testing session. The practice session took approximately five minutes.

Next, participants studied their allocated errand list for two minutes followed by free recall, then five minutes of further study and a test of cued recall. The total number of errands correct from free and cued recall comprised the errand list memory score, equivalent to a measure classified as 'Learn' by Burgess et al. (2000). Finally, participants were provided with a schematic building map and a copy of the errand list, and were asked to indicate the order in which they planned to perform each errand to achieve maximum efficiency in task completion. They were informed that they could change their plan during the actual test. Upon completion of their plan, the task list was removed along with their written plan, and they were asked again to verbally recall the errand list and building rules with any mistakes corrected. This process was repeated until 100% recall of the list. This last procedure was to minimise the chances that participants would fail to complete errands in EVET because they could not recall all of the errands.

Participants next performed the EVET test for 8 minutes (neither task list nor plan were present during the test). On completion, they were asked to recall what errands they had attempted, or failed to complete, and any building rules they had broken (see Table 2). This assessed whether participants could recall all of the actions they performed in EVET and is equivalent to the measure labelled 'Recount' by Burgess et al (2000). It was followed by free recall of the complete errand list regardless of whether or not all of the errands had been completed. Participants were then cued about any errands that they had omitted in this post-test free recall. A point was awarded for each errand correctly recalled by both methods to generate a measure equivalent to the "remember" variable in Burgess et al (2000). Finally, all participants were presented with the alternative errand list (set B if they had used set A and vice versa) with a fresh sheet showing the layout of each floor, and were asked to generate another plan for the order of errands in this alternative list. This was used as a second measure of planning within the context of EVET, but without performing EVET a second time. "EVET travel time" indicated the total amount of time each participant spent travelling in the EVET building, but excluded time spent inside rooms. This measure was intended to

index efficiency of navigation through the building. Finally, the EVET score indicated errand completion efficiency. Points were awarded for errand completion and were removed for breaking the building rules (Table 2).

The Word Recall Task required participants to recall orally, in any order, twelve words read out by the experimenter, one word per second following the standardised procedure in Capitani, Della Sala, Logie and Spinnler (1992). This was repeated for five different lists (60 maximum). This was used as an independent measure of retrospective memory.

Working Memory Verbal Span was based on working memory sentence span from Baddeley, Logie, Nimmo-Smith and Brereton (1985) and Duff and Logie (2001). Participants verified each of a set of sentences as they appeared consecutively on a computer monitor, and were to memorize the last word of each sentence. At the end of the set of sentences they were asked to orally recall, in order, the sentence-final words. The task began with sequences of two sentences, with each set size repeated three times, after which the set size increased by one. All participants continued, regardless of performance, until the maximum set size (seven sentences) was completed. All sentences were presented for three seconds, preceded by a fixation cross for one second. Total correct recall of the sentence-final words was calculated as a proportion of maximum possible (81).

Working Memory Spatial Span was based on a task from Shah and Miyake, 1996). Participants verified whether block capital letters that appeared consecutively were shown in their normal configuration or as a mirror image. Each letter was shown in a different orientation within a circular area. Participants were instructed to memorize the orientation of each letter for recall at the end of each set. The task began with sequences of two letters, with each set size repeated three times, after which the set size increased by one. All participants continued, regardless of performance, until the maximum set size (five letters) was completed. All letters were presented for 3 seconds, and preceded by a fixation cross for one second. Total correct recall was calculated as a proportion of maximum recall score (70).

The Travelling Salesman Task (TSP) involved presenting nine colored target shapes along the bottom of the screen, with the first shape labelled "Start/End" and the rest "Target Locations". The nine target shapes for each trial were also placed in the main section of the screen at random locations in a 5 x 5 array, and different colored shapes were placed as distracters in

the remaining 16 locations. Participants were asked to click on each of the target locations in the order that formed the shortest possible path connecting the start shape with the locations of each of the targets in the array, finally returning to the start location. As each target was clicked in the array it was marked to show which targets in the array had been visited, and it disappeared from the target list at the bottom of the screen to indicate which targets had yet to be visited in the array. When all nine target locations had been visited and the participant returned to the start point, the next array was presented, using a different selection of nine target shapes from the full set of 25, and with the targets shown in new random locations in the array. There were two practice trials, both with nine different targets, but the first showed only an array of the nine targets in random locations, and no distracters. The second showed nine targets along with 16 distracters within the full 5x5 array. Optimum distance for each array was calculated using an algorithm for travelling salesman problems (Kirk, 2007). Each participant completed a total of ten arrays. There was only one possible optimum solution for each combination of array and target set, and the average proportion of distance longer than the optimum was calculated for each array completed, with the average proportions across the ten arrays comprising the score.

The Breakfast Task was used as a measure of prospective memory ability. It was devised by Craik and Bialystok (2006) who kindly provided a copy of the programme. In this task participants had to swap between setting plates and cutlery on a virtual table on the computer screen, and starting and stopping the virtual cooking of eggs, coffee, pancakes, sausages and toast, each with different cooking times ranging from 120 seconds to 330 seconds. The goal was for all foods to complete cooking at the same time. The five foods were each shown on a separate screen from each other and from the table being set. To start cooking, the participant clicked on an icon for the food with the longest cooking time (330 seconds). This took them to a screen for that food where they clicked to start a timer displaying the progression of cooking time for that food. They then had to return to setting the virtual table until it was time to start the food with the next longest cooking time (240 seconds). This continued until the time when all the foods should be ready and the participant had to visit each screen to stop cooking each food. Participants practiced with two foods. Craik and Bialystok reported a range of different measures of cooking performance, which, in our data were highly correlated. Our chosen outcome measure was the average deviation between the actual start time for each food and the time that each food should have started. Given that 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 an independent measure of prospective memory ability. <mark>We chose this task rather than established laboratory measures of prospective memory, because it was a simulation of everyday prospective memory, and our overall goal was to investigate simulations of everyday complex cognition in multitasking.</mark>

Results

Twelve participants were unable to complete either the breakfast or travelling salesman task due to technical problems. Their data were excluded, leaving a final total of 153 participants (95 women and 58 men) for subsequent analysis.

The EVET is a completely novel task and a novel paradigm. There are therefore no clear precedents for generating scores to be used as dependent measures. We describe below the rationale for, and the procedure followed to generate scores. Our aim was to generate indicator variables that were as close as possible to those considered by Burgess et al. (2000), with additional variables that were intended to assess the contribution to multitasking of additional cognitive functions that were not considered in the earlier study.

EVET Score

The overall *EVET Score* comprised a weighted score procedure devised to emphasise task efficiency. Following Burgess et al. (2000), the general principle was to award points for tasks completed, and remove them for rule breaks. Errands that could be completed at any time were awarded one point for each successful action (maximum two points for two-part errands), yielding a maximum action score of 8. Finding the door code and unlocking the stair door counted as one action for this purpose. The time-restricted (cinema and meeting) and open ended (folder sort) tasks were weighted on a five point scale (0, 1, 2, 3 and 4), yielding a possible maximum bonus of 12 points, which were added to the action score for a potential overall maximum of 20 points. There were no obvious a-priori criteria for allocating bonus points. Rather than generate arbitrary criteria, we based the allocation on inspection of the frequency distribution of raw scores on each of these measures (Table 2). The rationale was to generate a distribution of scores that fairly reflected the distribution of actual performance across participants on each task. We acknowledge that the precise cut-off criteria for allocating scores to adjacent bonus categories remain somewhat arbitrary, but given that the same criteria were applied across all participants, we have no grounds for

suspecting that this procedure resulted in systematic bias in the overall EVET score, or distorted its use as a measure of individual differences.

A similar scoring procedure was used for breaking EVET rules, except minus points were allocated for going up the down stairs and vice versa, and for going into rooms or picking up objects that were not part of the errand list. We allocated minus points (0, -1,-2,-3,-4) on the basis of inspection of the frequency distributions of each error type (Table 2), and to fairly reflect the error performance across participants. For example, over 80% of participants broke the stair rule once, and so this attracted zero penalty. Across the three error types, the maximum possible penalty score was -12, which was combined with the action+bonus score for successful performance of errands in EVET. A negative score was possible if participants failed to completed most of the errands and incurred a large penalty score, although this never occurred within our participant sample.

Table 2 about here

EVET learn score ("Learn"). This is a sum of the free and cued recall scores that participants took prior to starting the EVET test, based on number of points for each errand. For example, "Pickup computer in G4 and take to T7" contained three elements (the item, collection point, and delivery point), so was worth 3 points. For free recall 23 points were available and for cued recall 14 points (1 point was removed from each errand due to the cue) making a possible maximum of 37 points.

EVET Plan Efficiency: ("Plan"). The score for plan efficiency was different from that used for the Greenwich test (Burgess et al., 2000). There are more possible permutations of order for EVET sub-tasks, and, unlike the Greenwich test, inefficient ordering may result in failure to complete all errands within the time limit. Assessment of efficiency of individual EVET plans involved comparison with an optimum plan for each set of errands. There were no clear a-priori criteria for identifying the optimum plan. Therefore we examined the rank order in which errands were actually completed by each of the five highest scoring participants (scoring 19 or 20) for each errand list. The rank order correlations between these five individuals ranged from 0.791 to 0.955. We therefore used the average rank order as the optimum plan for each list against which to compare the efficiency of the order in which errands were completed by individual participants. For each errand order for each participant,

one point was awarded for each errand that was in the same serial position as in the optimum plan. If there was a mismatch of serial position, then a point was awarded if a pair of errands were in the same sequential order as on the optimum plan, even if, as a pair of errands, they were completed earlier or later in the sequence. ==This ensured that there was credit given for partial use of the planned errand sequence.== The maximum possible overlap score was 11. Each individual overlap score was divided by 11 to derive a plan efficiency score. We defined "EVET pre-test plan" as the efficiency score for the plan participants made before they attempted the EVET, and defined "EVET post-test plan" as the efficiency score of the plan that participants made at the end of the EVET procedure for the alternative errand set.

*EVET follow score ("Follow").*This was designed to measure the correspondence between pretest plan and the order in which participants actually completed the errands. It was based on allocating one point for each errand that was completed in the same sequential position as planned, and a point if a pair of errands were in the same sequential order but not in the same overall sequential position as on the pret-test plan, following an analogous scoring procedure to that for the plan efficiency score. Each individual score was divided by the total possible points that could have been gained from the number of tasks actually completed by each individual participant to index how closely participants followed their initial plan.

Recall EVET actions ("Recount"). After completion of EVET, a point was awarded for each errand or type of rule break that participants recalled actually carrying out; equivalent to the Burgess et al. (2000) 'Recount' variable. Only the number of rule break types was recorded. For example, if the participant had gone down the 'up' stair 5 times, they were asked to recall if this had happened on at least one occasion. The maximum possible score was 26.

EVET remember score ("Remember"). Here, the scoring procedure for the measure "Learn" was repeated but was based on free and cued recall of the errand list after completing the EVET.

==The six measures above were designed to be comparable to those considered by Burgess et al. (2000): Score, Learn, Plan, Plan follow, Recount, Remember.== We included additional measures that we hypothesised would address the broader theoretical constructs that might be incorporated within or added to the Burgess model (EVET travel time, verbal and spatial working memory span, travelling salesman task, EVET post test plan, word recall, and

breakfast task). The two planning measures (EVET pre-test plan and EVET post-test plan), were intended to allow free estimation of the error variance of the planning construct indicators in our structural equation model. This was not possible for Burgess et al. (2000) who included only one indicator of planning.

Scores were converted to percentages for analysis. Descriptive statistics and intercorrelations are shown in Table 3.

Table 3 about here

Analyses

First we examined which of the five independent measures contributed unique variance to prediction of EVET score, using multiple regression with backwards stepwise elimination (Table 4). As expected from the Burgess et al. (2000) model, retrospective memory (word recall) was a significant predictor of EVET score. However, unlike the earlier model, our measure of prospective memory (breakfast task) did not reliably share variance with EVET, nor did it correlate with any other variables. Like the breakfast task, EVET involves unique items for each of the errands, and so it is not straightforward to calculate split half reliability across equivalent test items. However, Table 3 shows significant correlations between several different measures of EVET - Travel Time, Learn, Recount, Remember and Plan Follow, indicating a reasonable level of internal consistency. EVET Score also correlates with established independent measures of word recall, verbal and spatial working memory.

Our independent measure of planning ability (TSP) was a significant predictor, and so unlike Burgess et al., (2000), we included planning in the model on statistical grounds rather than inferring its contribution indirectly from associations with brain lesion sites. Unlike our other independent measures, the TSP was a novel task, that allowed calculation of internal reliability. This revealed a Chronbach Alpha of 0.737. Spatial working memory span was also a reliable predictor, indicating that it made a contribution that was independent of word recall, and of planning.

Verbal working memory did not make an independent contribution in this initial analysis. Although Table 3 shows that this measure correlated with verbal free recall, when verbal

working memory was forced into the regression equation before verbal free recall, the verbal working memory measure did not share significant variance with EVET (Table 4 notes).

Insert Table 4 about here

To further examine the role of planning in EVET performance, we explored the partial correlations for the efficiency measures of the two pre-test plans relative to the optimum plan, the plan follow score, and the EVET score. Controlling for plan following, a significant correlation was found between pre-test plan efficiency and EVET performance (partial r = 0.16, p=0.04), but a stronger correlation was found between EVET performance and the efficiency with which participants *followed* their plan (partial r = .32, p<.01), controlling for pre-test plan efficiency. Including the interaction between pre-test plan efficiency and plan following added no significant contribution to the prediction of EVET score, so these appear to be separate main effects.

A graphical demonstration of the importance of plan-following can be seen in Figure 3 that illustrates the length of time participants spent at each specific set of x,y,z co-ordinates in the 3-D space. Based on the plan following measure, participants were split into upper (N =38) and lower quartiles (N =37) and their movements across all four floors during the eight minute EVET test were characterized using kernel density estimation. In the figure, a peak indicates that participants remained in that particular position for a period of time, whereas the absence of a peak indicates movement through the virtual building. It is clear from Figure 3A that participants who did not follow their plan spent more time travelling along the building corridors (lower peaks). Figure 3B, in contrast, shows that participants who adhered to their plan had a strong tendency to focus their time in the folder sorting room (high peak on lower right side) with corridor time and movement kept to a minimum.

Insert Figure 3 about here

An exploratory factor analysis confirmed our theoretical expectations in identifying the three factors that were included in the Burgess et al. (2000) model, labelled here as 'Memory', 'Plan' and 'Intent'. It was clear that the breakfast task had no relationship with any of the factors so could not be used as an indicator. The EVET-learn measure was not easily identified with a single factor (first and second factor loadings, .39 and .34 respectively) so was also excluded.

Structural equation modelling (SEM) with maximum likelihood estimation was carried out using EQS 6.1 (Bentler, 2004) to assess the fit of the Burgess et al. model to the EVET data. The following model fit indices were used: $\chi 2$ (Bollen. 1989), which tests the hypothesis that an unconstrained model fits the covariance/variance matrix better than the proposed model (non-significant values indicate good fit); Bentler's (1990) comparative fit index (CFI), which compares the proposed model with the null model that assumes all variables are uncorrelated (values between .90 and .95 are acceptable); root mean square error of approximation (RMSEA) is a measure of the closeness of fit, with values below .08 indicating reasonable model fit and below .05 a good fit.

The Burgess et al. structural framework (Figure 1), positioned 'Memory' upstream from both 'Plan' and 'Intent'. Fit indices indicated a reasonable fit to our data: χ^2 ($df = 42$, $N = 153$) = 74.09, p =<0.01, CFI = .87 RMSEA = .07 (.04 - .10). Modification indices were used to evaluate possible model changes towards improved model fit. On the basis of these modification index recommendations and theoretical relevance, a path between 'Plan' and 'Intent' was included, which was not possible for Burgess et al. (2000), given that they had only one index for planning. This second structural model (Figure 4), therefore assumed that the 'Intent' factor was predicted by both 'Memory' and 'Plan' factors, with a weak link between 'Memory' and 'Plan'. Fit indices suggested a better fit for this second model: χ^2 ($df = 41$, $N = 153$) = 64.91, p =<0.01, CFI = .90 RMSEA = .06 (.03 - .09). A chi-square difference test, (χ^2Model 1 - χ^2Model 2, dfModel 1–dfModel 2), showed a significant difference between the models, χ^2 ($df = 1$) = 9.18, p =0.01. This suggested that the second model offered a significantly different and better fit, hence supporting the important relationship between 'Plan' and 'Intent' latent variables, and between 'Memory' and 'Intent', but with 'Memory' and 'Planning' only weakly related.

Figure 4 about here

DISCUSSION

Our primary goal was to investigate the cognitive factors that contribute to a realistic simulation of everyday multitasking. This was intended to address a major lacuna in understanding of complex cognition in which a several cognitive functions are required to operate in a co-ordinated fashion, rather than focus on each function in isolation. We drew on

previous research on cognitive impairments that are sequellae of frontal lobe damage, and developed a novel paradigm using a virtual environment (EVET) and novel behavioral measures to investigate the general cognitive principles of multitasking in the healthy young adult brain.

The only existing statistical model of everyday multitasking (Burgess et al., 2000) was used as an initial theoretical framework. However, that framework was based on a set of table-top tasks carried out by individuals with frontal lobe damage and older controls, with a limited range of measures and with only one measure of planning. This model was broadly useful in explaining our data from simulated multiple-errand multitasking with young healthy adults, but required modification. We identified separate constructs for memory, pre-planning, and plan implementation. This suggests that memory and pre-planning may involve different cognitive functions, both of which are required for this form of multitasking in young, healthy adults. This separation between memory and planning also is consistent with the finding by Shallice and Burgess (1991) and by Burgess et al. (2000) that planning can be selectively impaired in individuals with frontal lobe damage, while leaving other cognitive functions, including memory, intact. Further, our results are consistent with the Burgess et al. (2005) suggestion that planning is not a unitary construct, and we identified separate constructs for pre-planning and for on-line planning during task performance. However, these latter two factors were not uncorrelated. Participants who formed an efficient plan also tended to achieve a higher EVET score. In addition, participants who closely followed their plans during EVET performance also tended to achieve higher scores than those who modified their plans on-line. This was true regardless of whether their initial plan was poor or was close to the optimum. Given that errors driven by a change in plan would arise after a plan change, this suggests that changing a plan on-line in an attempt to make it more efficient at the point of the change might have disrupted the plan for performance of errands yet to be completed.

Our final model comprised three factors: 'Memory', 'Plan' and 'Intent', as did the Burgess et al. (2000) model. However, for the latter, the 'Plan' construct was added on neuroanatomical grounds, and 'Memory' was thought to drive 'Plan' and 'Intent' as separate constructs. Also, all of the tasks in the Greewich test were visible, and could be performed in any order. In the current study with the Edinburgh Virtual Errands Test, the 'Plan' construct was added on statistical grounds, and a better fit was obtained when 'Memory' did not drive the 'Plan'

construct, but when both drove the 'Intent' construct for which there were several indicator variables. Moreover, participants were required to memorise a list of errands and create a plan of the order for completion, intended to simulate the everyday requirement of scheduling a list of tasks for completion. These operations might be expected to require resources from memory and planning, so the separate sets of indicator variables for 'Memory' and for 'Plan' suggest that these two constructs reflect different cognitive functions that act together, but have little mutual dependence. EVET required participants to maintain their internal list of delayed intentions (plan), and then realise those intentions within the virtual environment. Presumably this would require frequent consultation of the plan. Clearly, high scoring participants followed their plans rather than deviating from them on-line. This argument is consistent with the Model in Figure 4, which illustrates strong links between the INTENT construct and the indicator variables 'EVET remember', 'EVET plan follow' and 'EVET recount'. 'EVET remember' refers to the ability to remember, after EVET completion, the errand list, whereas a high score on 'EVET plan follow' requires participants to keep track of which errands had been completed, and to remember to perform the remaining errands in the planned order. So, a combination of good memory for the errands, coupled with consultation and updating of the current representation of the position on the errand list both appear to contribute to the 'INTENT' construct. A high score on 'EVET recount', suggests that participants had an accurate memory for actions they had actually performed during EVET, so they could recall what errands had actually been completed and which had not. This is consistent with the idea that they could update a representation of which errands had been completed from their planned sequence, and which had not as they were performing EVET. We would then expect that participants who effectively consulted and followed their plans would have shorter EVET travel times with fewer deviations from the planned route, hence the negative loading for the 'EVET travel time', and would achieve more errands as a result (EVET score). These considerations suggest a factor that reflects plan following should draw on resources of memory and preplanning. This is apparent in the final model that includes a path between 'Plan' and 'Intent'. This was possible because we included two estimators for planning.

Also intriguing was that the Travelling Salesman Problem loaded on 'Intent' rather than 'Plan'. Our model therefore suggests it is an indicator variable for plan implementation. A detailed analysis of the TSP task is beyond the scope of this paper. However, previous studies of the Tower of London task, and of chess have shown that most planning for complex tasks is on-

line rather than in advance (e.g. Phillips et al., 2001; Saariluoma, 1995). Yet, we found that sticking with an original plan resulted in better scores than modifying plans during EVET performance. It is possible that participants created a planned order for the errands, but then did not pre-plan the actual moves between errands. For example when moving around the virtual building, they might have found shorter, or more efficient routes between the locations they had to visit for each errand, so the errand order remained as planned, but the movements between errands were modified on-line. This would also be consistent with EVET travel time loading on the INTENT variable.

The model we propose does not exclude the possibility that other models, based on alternative theoretical rationales, might offer a better fit. However, the results support for our suggestion that multitasking requires the operation of different cognitive functions acting in concert. One possible account, driven by our overall theoretical rationale is that different cognitive functions are required to support performance, and they tend to correlate because they act together to achieve a common goal, not because they reflect the operation of a single construct such as a general attentional resource. This adds to our confidence that use of complex paradigms such as EVET, together with a theoretical rationale that assumes the co-ordinated operation of multiple cognitive functions to achieve task goals, can generate robust and interpretable data to address the complex cognition that supports everyday multitasking.

The multiple regression analysis suggested that overall EVET score was predicted by independent measures of retrospective memory, of visuo-spatial planning (travelling salesman problem), and of spatial working memory. It may be that a participant's ability to manipulate spatial information is a key factor for this particular type of multitasking. This may be less important in the Greenwich test, where sub-tasks are in full view of the participants. No additional unique variance in EVET score was explained by an independent measure of prospective memory (simulated breakfast making) or verbal working memory. The lack of a contribution from verbal working memory appeared to be due, at least in part to co-linearity with the measure of retrospective memory, suggesting that these two measures reflect a common, domain-specific cognitive function, that is different from visuo-spatial working memory and planning, and consistent with the multicomponent framework for working memory (Baddeley & Logie, 1999; Logie, in press).

It is striking that performance on the Breakfast Task (Craik & Bialystok, 2006) was unrelated to overall EVET performance. Craik and Bialystok demonstrated that it showed clear effects of cognitive ageing and could detect an advantage for older bilingual compared with older monolingual individuals. Participants have to monitor the progress of a number of foods that cook at different rates, and stop them cooking at the appropriate time. This is very different from a situation where a participant has a specific, planned order in which they are attempting a memorised list of quite diverse tasks, and this could be a crucial difference between EVET and the breakfast task. It may also be the case that young healthy participants correct their prospective memory errors by returning to complete 'forgotten errands' within the eight minute period. In this case, the EVET score might be insensitive to individual variation in prospective memory performance within EVET. However, this is speculative, and the general issue of prospective memory in multitasking using EVET clearly merits further investigation. However, detailed consideration of this issue is beyond the scope of the current paper, and has been addressed elsewhere (Trawley, Law & Logie, in press).

A possible criticism is that the novelty of EVET also comes with uncertainty as to its reliability. This is an issue with any novel paradigm, particularly when it is more complex than is common in studies of human cognition. However, if we are to understand how human cognition deals with everyday complexity, then the experimental paradigms have to be complex as well as robust. The results of the current study illustrate the utility of exploring this form of complex cognition using multiple errands in a virtual environment, retaining a high degree of experimental control together with a degree of realism. Some confidence in the reliability of our paradigm comes from the degree of internal consistency between the different measures of EVET performance, and from the robust relationships with more established measures of retrospective memory and working memory. However, further assessment of its reliability would be useful to address in future studies.

One possible limitation of the implementation of EVET in this study is that the results might not generalise to other scenarios. However, with minor modifications to the procedure EVET could be used to explore a wide range of research questions, for example when multiple errands are carried out without pre-planning, or when plans are interrupted. We have completed experiments of this kind that will be reported elsewhere. Unlike the original multiple errands test (Shallice and Burgess, 1991), the same environment can be used in a

range of different laboratories, and the environment and the data extraction utilities are available for free[1] for not-for-profit research on request from the authors.

The study has developed and demonstrated the utility of a novel methodology to study everyday multitasking. It has added insight into how different cognitive functions act in concert to achieve complex cognitive goals, rather than studying each function in isolation. There are new findings reported regarding the relationship between pre-planning and plan following, and the relationship between three constructs (Memory, Planning and Intent) in multitasking by healthy young adults, thereby substantially developing research that has focused previously on the cognitive impact of frontal lobe damage. The study has also demonstrated that this form of complex cognition can successfully be addressed and thereby yield insights into how human cognition can meet and manage multiple requirements of daily living.

---

[1] Setting up and running the EVET environment requires researchers to purchase a software licence for the widely available computer game Half-Life 2. The software development kit licence that comes with the game allows not-for-profit creation of virtual environments. The data extraction utilities require a local licence for MATLAB. A detailed technical description of the EVET and data extraction utilities is given in Trawley, Law, Logie and Logie (under review).

23

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Acknowledgements

24

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. & Logie, R.H. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (eds.) *Models of Working Memory*, pp28-61. New York: Cambridge University Press.

Baddeley, A.D., Logie, R.H., Nimmo-Smith, I & 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*.

Bollen, K.A. (1989).Structural Equations with Latent Variables.New York: John Wiley.

Bentler, P.M., (1990). Bentler, Fit indices, LaGrange multipliers, constraint changes, and incomplete data in structural models, *Multivariate Behavioral Research*25 (1990), pp. 163–172.

Bentler, P.M., (2004). EQS for Windows, 6.1 Encino, CA: Multivariate Software,Encino, CA..

Borst, J.P., Taatgen, N.A., Stocco, A. & van Rijn, H. (2010). The Neural Correlates of Problem States: Testing fMRI Predictions of a Computational Model ofMultitasking. PLoS ONE 5(9): e12966. doi:10.1371/journal.pone.0012966

Borst, J.P., Taatgen, N.A. & van Rijn, H. (2010). The Problem State: A Cognitive Bottleneck in Multitasking. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 36, 363–382.

Burgess, P.W. (2000). Real-world multitasking from a cognitive neuroscience perspective. *Attention and Performance, 18*, 465–472.

Burgess, P.W., Alderman, N., Forbes, C., Costello, A., Coates, L. M-A., Dawson, D.R. (2006), Anderson, N. D., Gilbert, S. J., Dumontheil, I., & Channon, S. The case for the development and use of "ecologically valid" measures of executive functions in experimental and clinical neuropsychology. *Journal of International Neuropsychological Society 12*, 194–209.

Burgess, P.W., Dumontheil, I., & Gilbert, S. (2007). The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends in Cognitive Science, 11,* 290-298.

Burgess, P.W., Simons, J.S., Coates, L.M.-A., & Channon, S. (2005). The search for specific planning processes. In R. Morris & G. Ward (Eds.), *The Cognitive Psychology of Planning* (pp. 199–227). Hove, U.K.: Psychology Press.

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.

Chisholm, C.D., Dornfeld, A.M., Nelson, D.R., & Cordell, W.H. (2001). Work interrupted: A
  comparison of workplace interruptions in emergency departments and primary care offices.
  Annals of Emergency Medicine, 38, 146–151.

Cohen, J. & Cohen, P. (1983). *Applied Multiple Regression/Correlation Analyses for the Behavioral
  Sciences*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates.

Craik, F.I., & 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.

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

Hambrick, D.Z., Oswald, F.L., Darowski, E.S., Rench, T.A. & Brou, R. (2010). Predictors of
  Multitasking Performance in a SyntheticWork Paradigm. *Applied Cognitive Psychology* 24,
  1149–1167.

Imbo, I. &LeFevre, J.-A. (2010). The role of phonological and visual working memory in complex
  arithmetic for Chinese- and Canadian-educated adults. *Memory and Cognition*, 38 (2), 176-
  185.

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

Knight, R.G., & Titov, N. (2009). Use of virtual reality tasks to assess prospective memory:
  Applicability and evidence. *Brain Impairment*, *10,* 3-13.

Knight, R.G., Titov, N., & Crawford, M. (2006).The effects of distraction on prospective
  remembering following Traumatic Brain Injury assessed in a simulated naturalistic
  environment. *Journal of the International Neuropsychological Society, 12*, 8–16.

Koch, I., Gade, M., Schuch, S. & Philipp, A.M. (2010). The role of inhibition in task switching: A
  review. *Psychonomic Bulletin and Review*, *17*, 1-14.

König, C.J., Bühner, M. & Mürling, G. (2005). Working Memory, Fluid Intelligence, and Attention
  Are Predictors of Multitasking Performance, but Polychronicity and Extraversion Are not.
  Human Performance, 18, 243–266.

Law, A. S., Freer, Y., Hunter, J., Logie, R. H., McIntosh, N., & Quinn, J. (2005). A comparison of graphical and textual presentations of time series data to support medical decision making in the Neonatal Intensive Care Unit. *Journal of Clinical Monitoring and Computing, 19,* 183-194.

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

Levy, J., Pashler, H. (2008) Task prioritization in multitasking during driving: Opportunity to abort a concurrent task does not insulate braking responses from dual-task slowing. *Applied Cognitive Psychology, 22*, 507-525.

Logie, R.H. (1995). *Visuo-Spatial Working Memory*. Hove, UK: Lawrence Erlbaum Associates.

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

Logie, R.H., Baddeley, A.D., Mane, A., Donchin, E. and Sheptak, R. (1989). Working memory and the analysis of a complex skill by secondary task methodology. *Acta Psychologica, 71*, 53-87.

Logie, R.H., Cocchini, G., Della Sala, S., & Baddeley, A.D. (2004a). Is there a specific executive capacity for dual task co-ordination? Evidence from Alzheimer's Disease. *Neuropsychology, 18,* 504-513.

Logie, R.H., Gilhooly, K.J. & Wynn, V. (1994b). Counting on working memory in arithmetic problem solving. *Memory and Cognition* 22, 395-410.

Loukopoulos, L.D., Dismukes, K. &Barshi, I. (2009). *The Multitasking Myth: Handling Complexity in Real-World Operations.* Farnham, UK: Ashgate.

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

Miotto, E.C., & Morris, R.G. (1998).Virtual planning in patients with frontal lobe lesions. *Cortex, 34,* 631-657.

Monsell, S. (2003).Task switching. *Trends in Cognitive Sciences, 7,* 134–140.

Morris, R., Kotitsa, M. & Bramham, J. (2005). Planning in patients with focal brain damage: From simple to complex task performance. In R. Morris & G. Ward (Eds.), *The Cognitive Psychology of Planning* (pp. 153–179). Hove, U.K.: Psychology Press.

Morris, R., & Ward, G. (2005). *The Cognitive Psychology of Planning*. Hove, UK: Psychology Press.

Neisser, U. (1978). Memory: What are the important questions? *In* M.M. Gruneberg, P.E. Morris and R.N. Sykes (Eds.) *Practical Aspects of Memory*. London: Academic Press.

Phillips, L.H., Gilhooly, K.J., Logie, R.H., Della Sala, S. & Wynn, V. (2003). Age, working memory, and the Tower of London task. *European Journal of Cognitive Psychology*, 15, 291-312.

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, 54A*, 579–597.

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.

Saariluoma, P. (1995). *Chess Players' Thinking*. London: Routledge.

Seshadri, S. & Shapira, Z. (2001). Managerial Allocation of Time and Effort: The Effects of Interruptions. *Management Science*, *47*, 647-662.

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.

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

Spiers, H.J. & Maguire, E.A. (2006). Thoughts, behaviour, and brain dynamics during navigation in the real world. *Neuroimage*, 31, 1826-1840.

Strayer, D.L., Drews, F.A., & Crouch, D.J. (2006). Comparing the cellphone driver and the drunk driver. *Human Factors, 48*, 381?391.

Subrahmanyam, K., & Greenfield, P.M. (1994). Effect of video game practice on spatial skills in girls and boys. *Journal of Applied Developmental Psychology, 15*, 13-32.

Trawley, S.L., Law, A.S. & Logie, R.H. (in press). Event based prospective remembering in a virtual world. *Quarterly Journal of Experimental Psychology*.

Trawley, S.L., Law, A.S. & Logie, M.R. & Logie, R.H. (under review). Virtual reality for a realistic price: methodology for using the Source 3D game engine in psychological research.

Van der Meulen, M., Logie, R.H., Freer, Y., Sykes, C., McIntosh, N. & Hunter, J. (2010). When a Graph is Poorer than 100 Words: A Comparison of Computerised Natural Language Generation, Human Generated Descriptions and Graphical Displays in Neonatal Intensive Care. *Applied Cognitive Psychology, 24*, 77-89.

Ward, G., & Allport, A. (1997).Planning and problem-solving using the 5-disc Tower of London task. *Quarterly Journal of Experimental Psychology*, *50*, 49–78.

Wickens, C.D. (2008). Multiple Resources and Mental Workload. *Human Factors, 50,* 449–455.

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	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. Number of bonus points added, and penalty points deducted for EVET score based on percentage of participants performing within the ranges shown in the table.

Bonus points added	**+4**	**+3**	**+2**	**+1**	**0**
Number of folders sorted	30+ (4%)	23-29 (11%)	15-22 (16%)	8-14 (21%)	1-7 (46%)
Cinema (absolute time discrepancy in seconds from 5:30 min)	0-2 (40%)	3-5 (10%)	6-7 (7%)	8-10 (5%)	11+ (4%)
Meeting (time discrepancy in seconds over 3:00min)	<=3mins (28%)	1-12 (9%)	13-25 (11%)	26-37 (6%)	38+ (24%)
Penalty points deducted	**-4**	**-3**	**-2**	**-1**	**0**
Number of objects picked up that were not on the errand list	4+ (1%)	3 (1%)	2 (6%)	1 (33%)	0 (59%)
Number of rooms entered that were not on the errand list	4+ (12%)	3 (5%)	2 (7%)	1 (15%)	0 (62%)
Number of times stair rule broken	5+ (10%)	4 (2%)	3 (7%)	2 (14%)	1 (67%)

Table 3. Descriptive Statistics and Correlation matrix of EVET performance and predictive measures

		M	SD	2	3	4	5	6	7	8	09	10	11	12	13
1	EVET score (%)	49.30	28.46	.27**	.29**	.23**	-.32**	-.02	-.44**	.46**	.27**	.46**	.23**	.23**	.35**
2	Word recall (%)	48.65	7.66		.44**	.13	-.02	.02	-.11	.27**	.07	.07	.17*	.22**	.09
3	Working memory verbal span (%)	82.85	16.09			.29**	-.18*	.09	-.14	.27**	.17*	.06	.02	.01	-.06
4	Working memory spatial span (%)	73.19	24.23				-.09	.02	-.19*	.18*	.08	.06	.00	.06	-.06
5	Travelling salesmanproblem (% above optimum distance)	10.10	6.42					.06	.29**	-.15	-.12	-.20*	-.08	-.20*	-.20**
6	Breakfast task (seconds)	16.12	14.75						-.01	.05	-.12	-.02	-.00	-.08	.04
7	EVET travel time (seconds)	305.10	42.08							-.20*	-.17*	-.27**	-.02	-.15	-.37**
8	EVET learn (%)	65.51	15.49								.06	.48**	.04	.18*	.04
9	EVET recount (%)	81.96	9.35									.22**	.13	-.04	.25**
10	EVET remember (%)	89.46	13.03										.10	.09	.28**
11	EVET pretest plan (%)	46.36	18.59											.36**	.23**
12	EVET posttest plan (%)	40.92	17.63												.22**
13	EVET plan follow (%)	62.29	24.09												

**p< 0.01 (2-tailed).

*p< 0.05 (2-tailed).

Table 4. Results of multiple regression with backwards stepwise elimination to assess the contribution to common variance between scores on the Edinburgh Virtual Errands Test and scores on five different measures of mental ability as described in the text.

Variable	*B*	*SE B*	*Beta*	*t*	*p*
Travelling Salesman	23.746	6.707	.27	3.54	<.001
Word Recall	.30	.09	.25	3.32	<.001
Spatial working memory span	.05	.02	.15	1.96	<.055

F=10.69, *df*=3, 152, *p*<.001 *R²*=.18
Model selection procedure: backwards stepwise elimination.
Excluded (Not Significant): Prospective memory task.
Verbal working memory (VWM) did not make a significant additional contribution.
When VWM was forced into the regression equation ahead of word recall,
values for VWM were: B=0.05, SE B=0.04, Beta=0.11, t=1.24, p=0.22
See text for explanation.

Figure Captions

Figure 1. An illustration of the structural equation model of multitasking from Burgess et al. (2000), redrawn from the original with permission.

Figure 2. Screen shot of the ground floor concourse area of the Edinburgh Virtual Errands Test (EVET)

Figure 3. Kernel density estimates of participant movement for lower (A) and upper (B) quartile plan followers.

Figure 4. A proposed model of EVET multitasking. Selected model fit indices: χ^2 ($df = 41$, $N = 153$) $= 64.91$, p < 0.01, CFI $= .90$ RMSEA $= .06$ (.03-.9), AIC $= -17.09$.

Figure 1.

Figure 2.

1
2
3   Figure 3
4
5
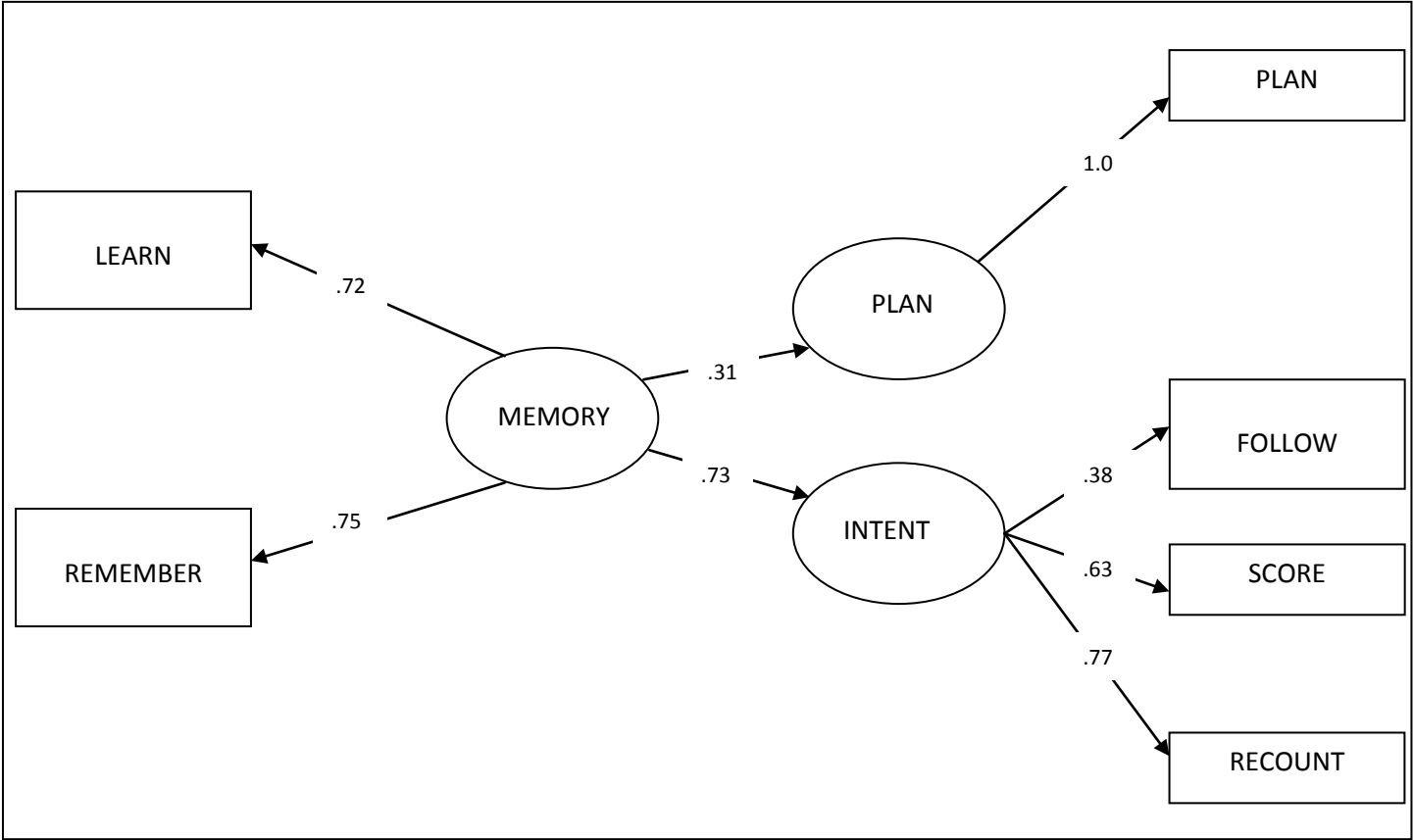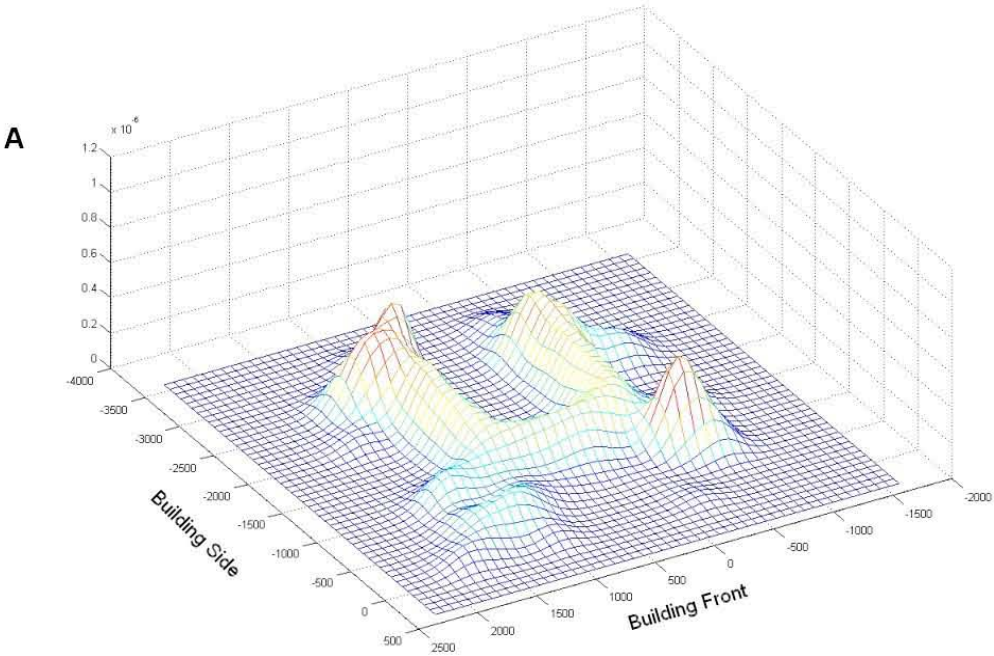6
7
8
9   Kernel density map, Lower quartile plan followers, 3D visualization   Kernel density map, Upper quartile plan followers, 3D visualization
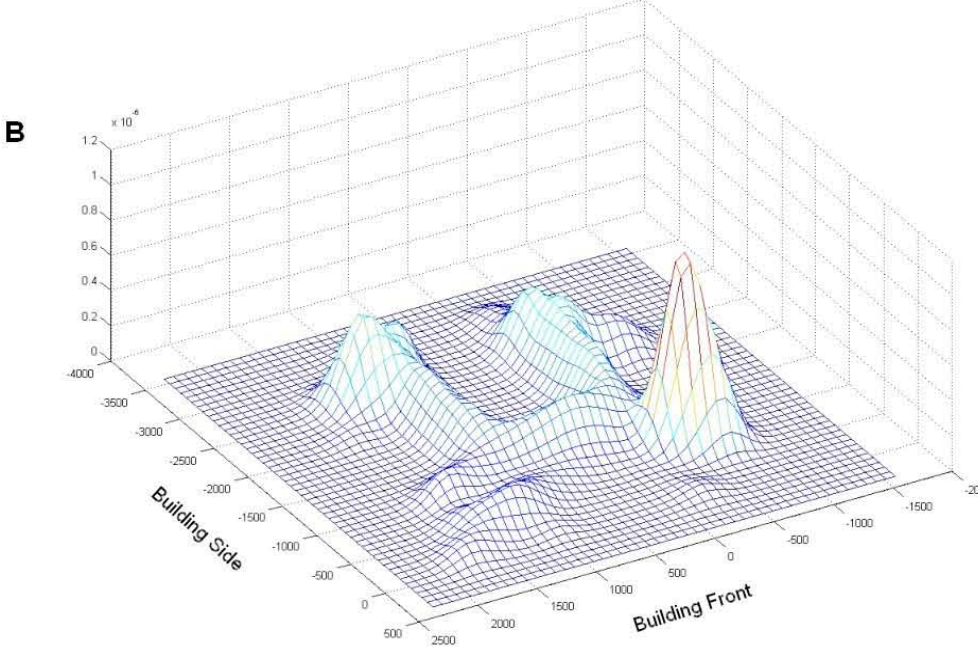10
11
12

Figure 4.