The impact of secondary tasks on multitasking in a virtual environment

Anna S. Law*

(Department of Psychology, University of Stirling)

Robert H. Logie

(Human Cognitive Neuroscience, PPLS, University of Edinburgh)

David G. Pearson

(School of Psychology, University of Aberdeen)

Running Head: Secondary tasks during multitasking

*Corresponding Author: Dr Anna Law, Department of Psychology, University of Stirling, Stirling, FK9 4LA, U.K.

Tel: + 44 (0) 1786 466375, Fax: + 44 (0) 1786 467641

Email: a.s.law@stir.ac.uk

Acknowledgements: We are grateful to Dr Peter McGeorge and Dr Louise Phillips at the University of Aberdeen for making the Virtual Errands Test available to us, and for their help in setting up the equipment and materials.

Abstract

One experiment is described that examined the possible involvement of working memory in the Virtual Errands Test (McGeorge et al., 2000), which requires participants to complete errands within a virtual environment, presented on a computer screen. Time was limited, therefore participants had to swap between tasks (multitask) efficiently to complete the errands. Forty-two undergraduates participated, all attempting the test twice. On one of these occasions they were asked to perform a concurrent task throughout (order of single and dual task conditions was counterbalanced). The type of secondary task was manipulated between-groups. Twenty-one participants were asked to randomly generate months of the year aloud in the dual-task condition, while another twenty-one were asked to suppress articulation by repeating the word "December". An overall dual-task effect on the virtual errands test was observed, although this was qualified by an interaction with the order of single and dual task conditions. Analysis of the secondary task data showed a drop in performance (relative to baseline) under dual-task conditions, and that drop was greater for the random generation group and the articulatory suppression group. These data are interpreted as suggesting that the central executive and phonological loop components of working memory are implicated in this test of multitasking.

Keywords: multitasking; working memory; phonological loop; central executive

PsycINFO Classification: 2340 - Cognitive Processes

1. Introduction

The term “multitasking” can be used to refer to a situation where a person has to complete multiple tasks, but cannot execute them sequentially (due to time limitations) or simultaneously (due to physical or cognitive limitations). The tasks must therefore be interleaved with one another, each being suspended and then resumed after appropriate intervals (Burgess, 2000a, 2000b). Everyday domestic examples are cooking and shopping, but multitasking is also necessary for many jobs, for example emergency medicine (Chisholm, Collison, Nelson, & Cordell, 2000) or management (Seshadri & Shapira, 2001). A number of studies have shown that patients with brain damage, particularly in the frontal lobes, can have great difficulty in applying efficient strategies in multitasking situations (Alderman, Burgess, Knight, & Henman, 2003; Burgess, Veitch, de Lacy Costello, & Shallice, 2000; Crépeau, Belleville, & Duchesne, 1996; Fortin, Godbout, & Braun, 2003; Goldstein, Bernard, Fenwick, Burgess, & McNeil, 1993; Knight, Alderman, & Burgess, 2002; Levine, Dawson, Boutet, Schwartz, & Stuss, 2000; Levine et al., 1998; Shallice & Burgess, 1991). However, little research has yet been conducted into the factors that constrain multitasking performance among healthy adults, and how the cognitive system deals with these complex situations (Law et al., 2004). Therefore, our aim was to investigate the cognitive demand of multitasking using a well-established theoretical framework - the multiple component model of working memory (e.g., Baddeley & Logie, 1999). Specifically, we used dual-task methodology to investigate the involvement of the phonological loop and central executive components of working memory in a test of multitasking. The test used was a “Virtual Errands Test” created by McGeorge and colleagues (McGeorge et al., 2001) and based on the “Multiple Errands Test” of Shallice and Burgess (1991).

The Multiple Errands Test (MET) was created by Shallice and Burgess (1991) in response to the observation that some patients with frontal lobe lesions who had disruptions to everyday life skills nevertheless performed normally on traditional "executive" tests, which were supposed to be sensitive to frontal lobe damage. The idea was to tap cognitive processes analogous to those involved in real life open-ended planning situations, but to have a quantifiable measure of performance. The original version of the Multiple Errands Test involved taking participants to a local shopping centre and giving them a list of tasks. Most of these were easy, for example "buy a brown loaf", but others were more difficult, for example "find out the name of the coldest place in Britain yesterday". An important part of a time-limited shopping trip is finding an efficient route through the shopping precinct. Excessive backtracking will result in the time elapsing before all errands are completed. Therefore the errands have to be interleaved in an efficient manner, rather than simply tackled in the sequential fashion in which they are presented. Shallice and Burgess found that their 3 patients were impaired in their ability to attempt this task effectively compared to a group of control participants – the patients tended to break more rules and to be more inefficient.

The sensitivity of the Multiple Errands Test as a measure of executive impairment has also been demonstrated in other studies with brain-lesioned patients (Alderman et al., 2003; Goldstein et al., 1993; Knight et al., 2002). These studies ask patients to attempt a task with high ecological validity in a real world setting, but this of course makes it a difficult and time-consuming test to administer. Some patients have become distressed or behaved in socially inappropriate ways while attempting the Multiple Errands Test (Alderman et al., 2003). In addition, unforeseen events in a real shopping centre can result in little experimental control. With these difficulties in mind, McGeorge and colleagues (2001) assessed the ability of five dysexecutive patients and five matched controls to undertake a version of the Multiple

Errands Test in a virtual environment presented on a computer screen. They compared performance in this condition with performance in a real building on which the virtual environment was based. Patients were recruited on the basis that care staff reported that they had problems with planning. In the “real” condition, which was attempted first, participants had to move around inside the physical building completing a series of simple office-type errands (e.g., meet a colleague in one room and take him to another, send a fax from the main office). Participants then attempted a similar set of errands within the virtual replica of the building. Patients completed significantly fewer errands than controls, and the type of environment (real or virtual) had no effect on the results.

The present study used secondary task methods (random generation and articulatory suppression) to investigate the working memory demands of the Virtual Errands Test (VET) for healthy adults. Given that McGeorge et al.’s “dysexecutive” patients performed poorly on the VET, a direct prediction is that performance will be sensitive to an executively-demanding secondary task. However, it is important to demonstrate this with behavioural data from a sample of healthy participants. Random Generation (RG) was chosen because it is becoming widely used in the working memory literature (Towse & Neil, 1998) and is thought to engage executive resources within working memory (Baddeley, 1996). Letters or numbers are often chosen as the response set from which participants have to generate random sequences, while inhibiting the well learned sequences such as alphabetical or numerical order. Neither letters nor numbers were suitable for the present experiment however, because the Virtual Errands Test involved remembering room identifiers that involved a letter and a number (e.g., F5, T15). Therefore, the response set chosen for the RG task was months of the year; these were not involved in the Virtual Errands Test but fulfilled the requirements for a RG task because there is a limited number of alternatives in the set with an over-learned sequence (i.e., calendar order).

For the present experiment, secondary task responses were generated orally because manual responding might have interfered with the use of the mouse-based interface in the Virtual Errands Test. This meant that the random generation task would also prevent sub-vocal rehearsal during the Virtual Errands Test, and any disruption to multitasking could be attributed to participants being unable to use inner speech. This possibility was investigated by asking half of the participants to suppress sub-vocal rehearsal by repeating the word 'December'. Articulatory suppression is thought to have less executive involvement than random generation, but still places an extra demand on participants because they are unable to use inner speech to rehearse current task goals. The word 'December' was chosen for the articulatory suppression task because this matched one of the longest alternatives in the random generation set. Some studies (e.g., Baddeley, Chincotta, & Adlam, 2001; Phillips, Wynn, Gilhooly, Della Sala, & Logie, 1999) have used the over-learned sequence as the articulatory suppression control to a random generation task, for instance they have asked participants to count from 1 to 9 if the task is random number generation, or to recite the months of the year in calendar order. However, using a single word created a bigger difference in executive demand between the two secondary tasks, while keeping them comparable in terms of overt vocalization.

In the task switching paradigm, where participants are required to switch frequently between two simple cognitive tasks, articulatory suppression has been shown to increase the time-cost of these switches (Baddeley, Chincotta, & Adlam, 2001; Emerson & Miyake, 2003; Goschke, 2000; Miyake, Emerson, Padilla, & Ahn, 2004). Emerson and Miyake (2003) argue that people make use of inner speech to help retrieve and keep active a phonological representation of the next goal they have to accomplish. If this is the case in task switching, then it is also likely that people will recruit inner speech during multitasking, where participants have to determine their own schedule of switching between multiple sub-tasks, rather than switching back and forth between two tasks in response to a cue. It was therefore expected that there would be interference between the Virtual Errands Test and the secondary task of articulatory suppression, but that the interference would be greater with the secondary task of random generation. In terms of the multiple component model of working memory (e.g., Baddeley & Logie, 1999), the articulatory suppression task would be expected to load the Phonological Loop, while the random generation task would be expected to load both the Phonological Loop and Central Executive. All participants attempted two versions of the Virtual Errands Test, one version without a concurrent task (single task condition) and the other version with a concurrent task (dual task condition). Half the participants were given articulatory suppression as their concurrent task, while the other half were given random generation. Secondary task performance was also recorded.

2. Method

2.1. Participants

All 42 participants were first year psychology undergraduates who received course credit for their participation. There were 26 women and 16 men, with equal proportions allocated randomly to each group (Articulatory Suppression (AS) or Random Generation (RG)). They ranged in age from 18 to 26, with a mean age of 19.52 years (SD = 2.05).

2.2. Design

The experiment employed a 2x2 mixed design where the within participants factor was the condition under which the Virtual Errands Test was attempted (single vs. dual task) and the between participants factor was the type of secondary task (articulatory suppression or random generation). The type of secondary task was manipulated between participants so that each participant would only complete the VET twice, thereby minimising possible effects of practice. The order in which participants attempted single and dual task conditions was counterbalanced. Performance on the Virtual Errands Test was scored as the number of errands completed minus the number of errors. A baseline measure of secondary task performance was collected in order that this could be compared with performance under dual task conditions. Rate of utterance was examined for both types of secondary task, and degree of randomness was analysed for the random generation task.

2.3. Equipment/Materials

The experiment used exactly the same computer hardware and software as McGeorge et al. (2001). A Personal Computer with a 350MHz Pentium II processor and 8MB ATI Graphics Card ran the virtual environment. The environment was created using Superscape 3D-Webmaster to construct a series of WebPages, which could then be viewed through a web browser program, in this case Microsoft Internet Explorer. It was viewed on a 17 inch colour monitor, and movement through the environment was controlled by the participant using a mouse. The monitor ran at 75Hz (resolution 600 x 800) and was connected to a LCD Flat Panel Projector. This allowed the screen to be projected onto a white wall from which it was videotaped using a JVC Camcorder. There was therefore a video record of the performance of each participant.

The virtual environment was a replication of a University building, which consists of three floors connected by two stairwells. Each floor consists of a long corridor, with consecutively numbered offices along one side. Stairwells are reached through doors on the other side of the corridor, and participants were told to travel in one direction only for each staircase. An example screenshot from the virtual environment is shown in Figure 1. The view of the environment was presented in the centre of the screen, and there was a black frame around the side which was divided into boxes with grey lines. The boxes on the left displayed the items which the participant was "carrying" and the ones on the right displayed information that he or she had collected. Along the top, feedback was provided when they completed a task successfully. In the top right corner was a display of the number of the last door that the participant had "clicked" with the mouse arrow. This very important feature allowed participants to navigate their way around the building, and to examine room numbers on doors as they would have been able to do in a real building.

Figure 1 about here please

There were two different errand lists that each consisted of 7 errands (see Appendix 1). Participants were required to memorise the list of errands before attempting the test. This was a departure from the procedure of McGeorge et al., who used a list of 12 errands to be completed in 20 minutes, and allowed participants to keep a copy of the list in front of them at all times. This change was made because previous data showed that undergraduates tended to perform at ceiling on the original version of the task (Law, 2004). Three of the 7 errands had two steps to them (i.e., pick up an object in one room and deliver it to another) while four had only one step. Each errand list asked participants to complete one of the errands *before* 4 minutes into the test, and another *at* 7 minutes through the test. Participants completed both of the errand lists, one under single task conditions and the other while attempting a secondary task. The order of presentation of the conditions and the errands lists was fully counterbalanced.

Additional materials consisted of a digital clock (a Sportline Giant Timer) that kept the time for the participants, and a stopwatch which the experimenter used to keep the time on the Virtual Errands Test. The digital clock was placed underneath the computer screen, so that it would be clearly in view while the participants attempted the Virtual Errands Test.

Secondary task performance was recorded using a voice-activated Sennheiser microphone to record rate of utterance, into software running on a Pentium II computer. For participants in the random generation condition, responses were recorded onto cassette tape. These data were later analysed for degree of randomness using the software program RGCalc (Towse & Neil, 1998). An electronic metronome was used in the explanation and practice of the secondary task.

2.4. Procedure

All experimental sessions began with a practice session on using the virtual environment. The movement controls were explained to the participants, as well as how objects could be manipulated. Participants were then guided on a tour through the virtual environment, which involved using the staircases and locating the rooms which were the most difficult to find. The procedure then differed for participants depending on whether they were attempting the single or dual task condition first. There were two different errands lists (Appendix 1) so that they worked on a different one in each condition, and the order in which these errands lists were given was also counterbalanced.

Single task condition: Participants were given a list of errands to learn, and the aim and rules of the Virtual Errands Test were explained to them. They then had 2 minutes to read through the task list before it was removed and they had to try to recall freely as much of the information as possible. The list was then returned to them for further study (for a maximum of 5 minutes) until they indicated that they were satisfied that they had learned the errands. Participants were then asked a series of cued recall questions, and informed of the answers to any that they answered incorrectly. The experimenter then started the video camera to record errands task performance, and started the participant's clock. Participants worked on the errands task for 8 minutes until asked to stop what they were doing. In a small minority of cases, participants finished all the errands and returned to their starting point before the time was up. When finished, participants were asked to recall everything they could from the task list once more, and given the same list of cued recall questions that they were given before the test.

Dual task condition: In the dual task condition, the secondary task was explained to participants first and they were given some practice at randomly generating months of the year or repeatedly saying "December". A metronome was used to demonstrate the rate of one per second in initial practice. Participants then had to execute the secondary task for 2 minutes without the benefit of the audible beep, and their performance was recorded. Following this measure of baseline secondary task performance, participants were given instructions for the errands task along with a list of errands to learn. They were then given the same free and cued recall procedure as described above for the single task condition. Participants were then told that when the clock was started they should begin both the tasks and perform each as well as they possibly could. The experimenter then allowed participants to hear the audible beep of the metronome once more before starting the video recorder, the tape recorder (when the task was random generation) and the software to record rate of utterance. When the clock was started participants began both tasks and continued until 8 minutes were up and they were asked to stop (or they finished the errands task, in a very small number of cases). Free and cued recall procedures for task list were repeated once more.

After participants completed the Virtual Errands Test under both single and dual task conditions, another two minutes of secondary task performance (articulatory suppression or random generation on their own) was recorded. This provided a second measure of baseline performance in addition to the one recorded before the VET.

3. Results

3.1. Virtual Errands Score

The overall score for each participant was the number of correct errands (or part-errands where the errand involved two steps) completed, minus the number of errors. Errors were either going into an incorrect room, or picking up an incorrect object. A high score therefore indicates better Virtual Errands performance. As an initial exploration of the data suggested possible interactions with the order of single and dual task conditions, a 2x2x2 mixed analysis of variance was carried out with the within-participants factor being the Task Condition under which the Virtual Errands Test was attempted (single vs. dual task) and the two within participants factors being Group (random generation vs. articulatory supression) and the Order in which the conditions were attempted. The mean scores for each cell of this design are shown in Table 1.

Table 1 about here please

The main effect of Task Condition was significant, F (1,38) = 8.035; p = 0.007, MSE = 4.485. There was also a significant main effect of Group, F (1,38) = 9.106; p = 0.005, MSE = 4.095, i.e., articulatory suppression (AS) vs. random generation (RG), but no significant main effect of Order, F(1,38) = 2.895, ns, MSE = 4.095. There was no interaction between Group and Task Condition, F < 1, but Order did significantly interact with both Group, F (1,38) = 11.269; p = 0.002, MSE = 4.095, and Task Condition, F (1,38) = 5.04; p = 0.031, MSE = 4.485. The Group by Order interaction was caused by relatively low single-task performance of RG participants who did the single task condition first, the reason for which is not clear. However, across all participants who did the single task condition first, performance did not significantly change between single and dual task. In contrast, amongst participants who did the dual task condition first, performance was significantly poorer in the dual task condition, for both random generation and articulatory suppression groups.

The data were also analysed according to the percentage of one-step errands and the percentage of two-step errands that participants in each Group completed. Over both conditions, the RG group completed 71.43% (SD = 20.21%) of the available one-step errands, and 65.14% (SD = 21.08%) of the available two-step errands. The AS group completed 83.33% (SD = 14.97%) of the one-step errands and 73.02% (SD = 20.79%) of the two-step errands. A 2x2 ANOVA conducted on these data showed that there was a significant main effect of errand type, F (1, 40) = 6.896, p = 0.012, MSE = 0.012, and a marginally non-significant effect of Group, F (1,40) = 3.771, p = 0.059, MSE = 0.054, with no interaction, $F < 1$. Therefore, participants were more successful in completing the errands that had only one step.

3.2: Time-constrained Errands

Two errands on each list had instructions about timing - the first was to pick up and deliver something by 4 minutes through the test, while the second was to deliver something as close to 7 minutes through the test as possible. However, the VET program allowed these errands to be completed at any time, and in the analysis of VET score above, these errands were counted as correct even if they were completed at the wrong time. The frequency with which these errands were successfully completed on time was then examined. In the single task condition, sixteen participants completed neither timed errand on time, twenty-one managed one of them and five managed both. In the dual task condition, nineteen completed neither, seventeen completed one, and six completed two. The proportion of participants completing the first timed errand was 52.38% in both the single and dual task conditions, while the proportion completing the second timed errand was 21.43% in the single task condition and 19.05% in the dual task condition. A Chi-square test showed that there was no association between dual task type (RG or AS) and number of timed errands completed in the dual task condition, $\underline{X}^2$ (2, N=42) = 1.199, ns. Sign tests showed that there was not a significant number of participants who did better at the timed tasks in the single task condition (10 participants did better in the single-task condition, compared to 6 who did better in the dual task condition). Therefore it was not the case that participants were more likely to complete the errands on time in the single-task condition, or that the AS group were more likely to complete them than the RG group.

3.3. Recall of the errand list

The mean free and cued recall data (shown in Table 2) were examined for any differences between single and dual task conditions, and also to check the random generation and articulatory suppression groups did not differ in their learning of the errands before they attempted the VET. Using a 2x2 ANOVA to examine initial free recall of the task list, there was no main effect of Condition, $\underline{F}$ (1,40) = 1.794, ns, MSE = 0.016, and no main effect of Group, $\underline{F}$ (1,40) = 1.297, ns, MSE = 0.031, but there was an interaction between Condition and Group, $\underline{F}$ (1,40) = 4.855, $\underline{p}$ = 0.033, MSE = 0.016. Simple effects analysis showed that there was a significant difference between the groups for the single task condition only, $\underline{F}$(1,80) = 4.867, p = 0.030, MSE = 0.024. However, on the cued recall measure taken just before participants began the test there was no main effect of Condition, $\underline{F}$ < 1, no main effect of Group, $\underline{F}$ (1,40) = 1.585, ns, MSE = 0.022, and no interaction, $\underline{F}$ < 1. Therefore, initial recall of the errand list after a 2 minute exposure differed for the groups in the single-task condition, with the articulatory suppression group recalling about 10% more, but five minutes further study eliminated this difference so that the groups began the VET at the same level, for both the single-task and dual-task conditions. As for the recall measures taken after the

VET, there were no significant main effects or interactions in these data, with all F-values less than 1. Therefore, it was not the case that performing either type of dual task during the VET caused participants to have poorer recall of the list of errands at the end.

Table 2 about here please

Due to the procedure of asking participants to memorise the task list, some relationship between recall performance and multitasking performance is to be expected. The correlations were examined for each condition of the experiment, using one-tailed tests. Across both groups, single task VET performance significantly correlated with free recall of the task list before the test, $r = 0.311$, $p = 0.022$, but not with cued recall of the task list before the test, $r = 0.121$, ns. There were also significant correlations with free recall of the task list after the test, $r = 0.345$, $p = 0.013$, and cued recall of the task list after the test, $r = 0.451$, $p = 0.001$. The correlation between dual task VET performance and free recall of the task list before the test was marginally non-significant, $r = 0.246$, $p = 0.058$, but cued recall performance before the test was significantly correlated, $r = 0.386$, $p = 0.006$. There was no significant correlation between dual task VET performance and free recall of the task list after the test, $r = 0.156$, ns, but there was a significant correlation with cued recall, $r = 0.260$, $p = 0.048$.

3.4. Secondary Task Data

3.4.1. Rate of Utterance

The rate of utterance was measured for both the Random Generation and Articulatory Suppression groups at the two baseline time-points and during dual-task execution of the VET. Firstly, the two baseline measures were compared. The AS group maintained a rate of one utterance every 1.131 seconds (SD = 0.112) at Baseline Time 1 and one utterance every 1.115 seconds (SD = 0.092) at Baseline Time 2. The RG group maintained a rate of one utterance every 1.300 seconds (SD = 0.181) at Time 1 and every 1.211 seconds (SD = 0.140) at Time 2. A 2x2 ANOVA was carried out on these data which showed a significant main effect of dual-task Group, $F(1, 40) = 11.855$, $p = 0.001$, MSE = 0.031, a significant main effect of Time, $F(1,40) = 10.009$, $p < 0.003$, MSE = 0.006, and a significant interaction, $F(1,40) = 4.997$, $p = 0.031$, MSE = 0.006. Simple effects analysis showed that the performance of the AS group did not significantly differ between the two time points, $F < 1$, but that the RG did make utterances at a significantly faster rate at the second time point, $F(1,40) = 14.579$, $p < 0.001$, MSE = 0.006. The variance in inter-response interval was also examined; for the AS group the mean was 0.066 (SD = 0.082) at Time 1 and 0.049 (SD = 0.078) at Time 2. For the RG group the mean was 0.109 (SD = 0.122) at Time 1 and 0.057 (SD = 0.039) at Time 2. A 2x2 ANOVA revealed no main effect of Group, $F(1,40) = 1.761$, ns, MSE = 0.008, a main effect of Time that only approached significance, $F(1,40) = 3.647$, $p = 0.063$, MSE = 0.007, and no interaction, $F(1, 40) = 1.014$, ns, MSE = 0.007.

For the next step the average of the two baseline measures was taken for each participant, and this was compared with rate of utterance under dual-task conditions. The data for mean inter-response interval, and mean variance in inter-response interval, are shown in Figure 2.

.

Figure 2 about here please

For mean inter-response interval, a 2x2 mixed ANOVA showed a highly significant main effect of Condition, F (1,40) = 62.779, p <0.001, MSE = 0.046, a highly significant main effect of Group, F (1,40) = 20.532, p < 0.001, MSE = 0.112, and a highly significant interaction, F (1,40) = 18.065, p < 0.001, MSE = 0.046. Simple effects analysis showed that both the AS group, F (1,40) = 6.746, p = 0.013, MSE = 0.046, and the RG group, F (1,40) = 74.097, p<0.001, MSE = 0.046, did significantly slow down in their rate of utterance under dual task conditions compared to baseline. However, this slowing was significantly greater for the RG group. The variance in inter-response interval was also examined using a 2x2 ANOVA which showed a highly significant main effect of Condition, F (1,40) = 29.575, p < 0.001, MSE = 0.280, a highly significant main effect of Group, F (1,40) = 14.351, p = 0.001, MSE = 0.320, and a highly significant interaction, F (1,40) = 14.681, p < 0.001, MSE = 0.280. Simple effects analysis showed that while the variance was significantly higher for the RG group under dual task conditions, F (1,40) = 42.964, p < 0.001, MSE = 0.280, the AS group did not become significantly more variable, F (1,40) = 1.291, p = 0.2627, MSE = 0.280, despite their overall slowing. However, the mean variance in inter-response interval for the AS group did increase from 0.057 as an average of the two baselines, to 0.242 in the dual-task condition.

The rate of utterance data were also analysed for order effects to see whether performance of these tasks was worse when participants had to combine them with the Virtual Errands Test the first time they attempted it. The impact of order was examined separately for each group (RG and AS) using one-way ANOVA. There was no effect of order on the mean inter-response interval in the rate of utterance data for either secondary task (Random Generation Group, F < 1, Articulatory Suppression Group, F(1,19) = 2.048, ns, MSE = 0.022). However, the effect of order on the variance in rate of utterance for the Random Generation Group approached significance, F (1,19) = 3.903; p = 0.063, MSE = 0.998. When these participants did the dual task condition first, the variation in rate of utterance tended to be slightly greater (M = 1.605, SD = 1.281) than when they did the single task condition first (M = 0.743, SD = 0.648). Possibly the benefit of practice on the Virtual Errands Test meant that they were able to keep up a more steady rate of responding on the random generation task. However, there was no effect for the Articulatory Suppression Group, who were able to keep up a steady rate regardless of whether they were doing the secondary task the first or the second time they did the Virtual Errands Test (F(1,19) = 1.455, ns, MSE = 0.049).

3.4.2. Degree of Randomness

A two minute section of dual task performance (between 3 and 5 minutes through the Virtual Errands Test) was analysed and compared with the average baseline performance. The computer program RgCalc (Towse & Neil, 1998) was used to analyse the degree of order in the random generation responses of the participants in that dual-task group. Three of the measures that RgCalc provides, and that have been used elsewhere (Towse & Valentine, 1997; Salway, 1991), were selected for analysis. The first, Redundancy, measures how equally the participant has chosen from the different response alternatives. A low score indicates better performance, i.e. how well distributed the choices were. The second measure chosen was Evan's RNG which is a commonly reported measure based on the distribution of response pairs. A higher number indicates better performance (more equality of response pair distribution). The third measure chosen was Ascending Adjacency (of response pairs), because the temptation when saying months will be to choose the next one in the calendar year. For this measure, a low score indicates better performance as there are a lower number of ascending adjacent pairs.

Using one-tailed paired-samples t-tests, the differences between average baseline performance and dual task performance were examined. For Redundancy, average baseline score was 3.373 (SD = 1.609) and the dual-task score was 5.705 (SD = 3.238), a difference that was significant, t (20) = 3.513, p = 0.001. The difference between Evan's RNG score at baseline (M = 0.317, SD = 0.052) and under dual-task conditions (M = 0.347, SD = 0.106) failed to reach statistical significance, t (20) = 1.552, p = 0.068. For Ascending Adjacency, there was a clear increase in the percentage of adjacent response pairs between baseline (M = 16.750%, SD = 8.472%) and dual-task performance (M = 28.070%, SD = 11.799%) – a difference that was highly significant, t (20) = -5.409, $p < 0.001$.

4. Discussion

Analysis of the Virtual Errands Test data offers some support for the hypothesis that the concurrent tasks would have a negative impact on performance, in that there is an overall impact of a secondary task. This conclusion is qualified by the interaction with order of conditions, in that the dual task impairment was most evident for those participants who had attempted the dual task condition first. One possible account is that participants relied heavily on working memory resources when they first encountered the virtual errands test, but learned strategies for task performance that reduced the working memory load when they were allowed to perform the Virtual Errands Task first on its own. There was no evidence that random generation had a greater disruptive effect on VET performance than articulatory suppression. There was also no evidence that either of the secondary tasks had an impact on the punctuality of the performance of the two errands that were given specific time constraints. There was a tendency for participants to complete a higher percentage of the errands that had only one step, than the errands that required two steps to complete.

A clear picture of the cognitive interference in the dual task condition can only be gained from consideration of the overall pattern of data from both primary and secondary tasks – it is not sufficient to rely on the primary task data alone (e.g., Logie, Cocchini, Della Sala, & Baddeley, 2004; Phillips, Gilhooly, Logie, Della Sala, & Wynn, 2003; Phillips et al, 1999). In the present experiment, participants appear to have been protecting performance on the Virtual Errands Test at the expense of the secondary tasks. This is an entirely rational approach for participants when finding themselves in a situation of cognitive overload. If they were not able to achieve optimum performance on both tasks, they may (despite instructions to the contrary) have chosen to give greater priority to the more engaging, realistic Virtual Errands Test. However, the fact that this trade-off had to occur is evidence of interference between the tasks, and the secondary task data show that the degree of that interference varied with the type of secondary task. The rate of utterance data show that the overall rate of random generation was slowed further by concurrent performance with the VET than was the overall rate of articulatory suppression, although the latter did slow significantly compared to baseline. The slowing of articulatory suppression performance suggests that sub-vocal rehearsal may be required for performance of the VET. For the random generation group, the variance in inter-response interval and the degree of randomness also increased significantly. Participants may have slowed down their random generation when performance of the VET became more demanding, and then speeded up when the concurrent demands eased. These data suggest that in addition to sub-vocal rehearsal, the Virtual Errands Test also requires the involvement of executive resources. In terms of the multiple-component model of working memory, the data suggest that both the Phonological Loop and Central Executive may be implicated in multitasking.

Performing the VET once gave participants enough practice on this complex task to perform it well under dual task conditions (although the secondary task performance continued to suffer), and this raises an important methodological point. The Multiple Errands Task was originally developed (along with the Six Elements Test) by Shallice and Burgess (1991) because traditional executive tests were not tapping into the difficulties experienced by some patients in everyday life. They identified several characteristics of these traditional tests that potentially made them less demanding than many real-life tasks. Trials were typically short, with clear signals from the experimenter about when they were to begin and end. Also, participants did not have to choose between multiple tasks. The Six Elements and the Multiple Errands Tests were designed to be open-ended and therefore more demanding. The Six Elements Test was included in the Behavioural Assessment of the Dysexecutive Syndrome (BADS) test battery, which was reported to have good test-retest reliability (Wilson, Evans, Alderman, Burgess, & Emslie, 1997). Many other executive tests do not achieve this, because part of the cognitive demand of an executive test necessarily arises from its novelty (see Rabbitt, 1997), and the task can only be novel the first time it is attempted. The current data would suggest that the Multiple Errands methodology does not overcome the problem of rapid practice and learning effects. Nevertheless, the Multiple Errands methodology has ecological validity as it is similar to the real-life task of shopping, and it does overcome the other problems (i.e., short, explicit trials with one clear goal) identified by Shallice and Burgess (1991). The procedure used here was somewhat different to previous studies, in that participants had to memorise the list of errands. This is analogous to real-life multitasking situations where people find themselves without external aids, and makes the task suitably demanding for a sample of undergraduates. It does mean that part of the variability in performance can be attributed to failures in memory for the errands (rather than on-line planning), as the significant correlations between VET performance and memory for the errand list show. However, we would argue that working with a heavy memory load is a typical part of the cognitive demand of multitasking, both in laboratory tests and in everyday situations.

It seems likely that if an executive control system such as the central executive of working memory exists, then it would be required for a complex activity like multitasking. However, the extent to which the central executive is a unitary construct is a matter of debate (e.g., Baddeley, 1996; Miyake et al. 2000). It is argued that executive resources may fractionate, and could be a related set of abilities rather than a unified system. The findings relating to "strategy application disorder" patients, who have multitasking deficits but can perform well on other executive tests, may be consistent with this argument (e.g., Shallice & Burgess, 1991, Goldstein et al., 1993). It could be that tests of multitasking draw on a wider range of executive abilities than these other tests, and so allow the deficits of patients to become apparent.

When considering executive functioning as a related set of abilities, it is worth considering the types of processes that may be shared by the VET and random generation. One similarity is that both tasks involve the control of a stream of events. In random generation participants have to keep track of the items that they have recently produced, and avoid stereotyped sequences of responses. Baddeley (1996) has argued that random generation may require the constant switching of plans that guide the retrievals of items from long-term memory, in order to produce a sequence that is suitably random. Also in the VET, participants have to keep track of those errands that have been completed and those that have yet to be attempted, and relate this to the locations involved and the time remaining. The list of errands should have been encoded through elaborative rehearsal into long-term memory during the few minutes that participants had to learn the list. As with random generation, participants may need to switch between different plans that guide the retrieval of the relevant information from long-term memory. For example, they might switch between retrieving errands based on their location, to retrieving the errands with a specific time limit. Executive resources would therefore be required to keep track of the sequence of errands, and switch between retrieval strategies for the ones that still had to be completed. If as Baddeley (1996) suggested, the random generation task required the same executive switching process, then it would be no surprise that there was interference between the two tasks.

As mentioned above, both the VET and random generation involve retrieving items from long-term memory. It has been shown (Rohrer, Pashler, & Etchegaray, 1998) that while items from the same category may possibly be retrieved concurrently, items from different categories must be retrieved in a serial fashion. In this case the errands and months, being completely unrelated, should not have been able to be retrieved concurrently. This is likely to have been a factor in the slowing of the random generation - when it was necessary to retrieve errands to make progress on the VET, so participants may well have given this priority and slowed the retrieval of the months. As participants had been instructed to name months at the reasonably brisk pace of 1 per second, this concurrent task had what Rohrer and Pashler (2003) describe as “temporal density”, which they showed to be important in determining whether a current task impacted on free recall. In the time-based resource-sharing model of

Barrouillet, Bernardin, & Camos (2004), an on-going task that requires frequent retrievals from long-term memory (as does random generation) will be particularly disruptive to the maintenance of information in working memory (e.g., the errand currently being attempted), due to a central bottleneck of retrieval. According to this model information in working memory decays until it is refreshed by a focusing of attention which also requires the retrieval bottleneck. As well as maintaining current goals in working memory, the VET also requires that uncompleted errands be retrieved from long-term memory, another process that would require the retrieval bottleneck. The data suggest that in this case retrievals relating to the VET were given priority, as the interference between the two tasks was seen in the slower and more variable random generation performance.

The suggestion that both the VET and random generation involve the on-line control of a sequence of events assumes that participants did not make a complete plan of how to order the errands while they learned the list. They were not instructed to do so, and evidence from the Tower of London (TOL) planning literature that suggests in general participants do not need or tend to form detailed pre-plans. The TOL task involves moving coloured discs on three pegs from a start arrangement to a goal arrangement in the smallest possible number of moves. This was initially thought of as a task requiring pre-planning for efficient performance (Shallice, 1982; Owen, 1997), but subsequent studies indicated that where pre-planning occurred it was of little benefit (Phillips, Wynn, MacPherson, & Gilhooly, 2001; Gilhooly, Phillips, Wynn, Logie, & Della Sala, 1999; Phillips, Wynn, Gilhooly, Della Sala, & Logie, 1999). Ward and Allport (1997) made a distinction between the pre-planning that occurs before a particular trial on the TOL, and the planning that occurs on-line during execution of the task. They found that during on-line planning, the longest move latencies occurred when there were two conflicting move options, either of which would advance a sub-goal of moving one disk to the correct location (Ward & Allport, 1997, Experiment 3). In the VET there will often be a number of competing options to select as the next goal; this would again suggest an increase in demand at times when the next errand had to be retrieved.

On-line planning would be likely to require sub-vocal rehearsal, and the rate of utterance data for the articulatory suppression group do suggest that inner speech was being recruited by the VET. Miyake et al. (2004) have argued that inner speech may serve as a self-cueing device, particularly where the forthcoming goal is difficult to retrieve due to the absence of salient cues. Consistent with this argument, Baddeley, Chincotta and Adlam (2001) found that concurrent articulatory suppression did not affect the speed of responses when participants had to switch between adding and subtracting digits, when the plus and minus signs were included on the list of sums. However, when participants had to keep track of whether to add or subtract the two numbers for themselves, articulatory suppression did slow performance. In the VET, there were few salient cues in the environment that could have helped participants maintain current goals; it would not be possible to rely on these in order to complete the test.

We have used the working memory framework to investigate the cognitive demand of multitasking. Finally it is worth considering how the other proposed components of working memory (the visuo-spatial sketchpad and the episodic buffer) might have contributed to VET performance. Scheduling the tasks efficiently in the Virtual Errands Test involves efficient route planning, which could require the visuo-spatial sketchpad as well as the central executive. If a spatial secondary task had been introduced in this experiment, it would likely have had a disruptive effect on Virtual Errands Test performance, but this might well have been due only to the particular demands of this task. Other multitasking tests reported in the literature do not have such an obvious spatial component as the Multiple or Virtual Errands

Tests, and the visuo-spatial sketchpad might not be expected to play a large role in other types of multitasking situation. It is possible that the other proposed component of working memory, the episodic buffer (Baddeley, 2000), might be used in the Virtual Errands Test to bind particular tasks (stored in a speech-based code) with locations (stored in a spatial code) into a unitary representation. However, research on the episodic buffer is at an early stage, and the secondary tasks needed to investigate its role in the Virtual Errands Test have yet to be developed. Also, the binding of articulatory and spatial codes might not be relevant for other tests of multitasking.

In summary, the addition of a concurrent task to the Virtual Errands Test did not have as dramatic an effect on performance as might have been expected. Participants were able to achieve a reasonable score on the VET, especially when they had the advantage of attempting it under single-task conditions at first. However, the secondary task data provided evidence of interference, and that interference was greater when the secondary task placed substantial demands on executive resources.

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

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417-423.

Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, 130: 641-657.

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

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133*: 83-100.

Burgess, P. W. (2000a). Real-World Multitasking from a Cognitive Neuroscience Perspective. *Attention and Performance, 18*, 465 - 472.

Burgess, P. W. (2000b). Strategy application disorder: The role of the frontal lobes in human multitasking. *Psychological Research-Psychologische Forschung, 63*(3-4), 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.

Capitani, E., Della Sala, S., Logie, R. H., & Spinnler, H. (1992). Recency, Primacy, and Memory - Reappraising and Standardizing the Serial Position Curve. *Cortex*, *28*, 315-342.

Chisholm, C. D., Collison, E. K., Nelson, D. R., & Cordell, W. H. (2000). Emergency department workplace interruptions: Are emergency physicians "interrupt-driven" and "multitasking"? *Academic Emergency Medicine, 7*, 1239-1243.

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.

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.

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. *Cortex, 39*, 273 - 291.

Gilhooly, K. J., Phillips, L. H., Wynn, V., Logie, R. H., & Della Sala, S. (1999). Planning processes and age in the five-disc Tower of London task. *Thinking and Reasoning, 5*: 339-361.

Goel, V., & Grafman, J. (1995). Are the Frontal Lobes Implicated in Planning Functions - Interpreting Data from the Tower-of-Hanoi. *Neuropsychologia, 33*, 623-642.

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.

Goschke, T. (2000). Involuntary persistence and intentional reconfiguration in task-set switching. In S. Monsell & J. Driver (Eds.), Attention and Performance XVIII: Control of Cognitive Processes (pp. 331-355). Cambridge, MA: MIT Press.

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

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.

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.

Owen, A. M. (1997). Cognitive planning in humans: Neuropsychological, neuroanatomical and neuropharmacological perspectives, *Progress in Neurobiology, 53*: 431-450.

Phillips, L. H., Gilhooly, K. J., Logie, R., 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., Gilhooly, K. J., Della Sala, S., & Logie, R. (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.

Rabbitt, P. (Ed.). (1997). *Methodology of frontal and executive function*. Hove: Psychology Press.

Rohrer, D., & Pashler, H. E. (2003). Concurrent task effects on memory retrieval. *Psychonomic Bulletin & Review, 10*: 96-103.

Rohrer, D., Pashler, H., & Etchegaray, J. (1998). When two memories can and cannot be retrieved concurrently, *Memory & Cognition*, *26*: 731-739.

Salway, A.F.S. (1991). Random generation in the working memory dual-task paradigm. Unpublished PhD thesis, University of Aberdeen, Scotland, UK.

Seshadri, S., & Shapira, Z. (2001). Managerial allocation of time and effort: The effects of interruptions. *Management Science*, *47*, 647-662.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B, 298*: 199-209.

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.

Towse, J. N., & Valentine, J. D. (1997). Random generation of numbers: A search for underlying processes. *European Journal of Cognitive Psychology, 9,* 381-400.

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

Wilson, B. A., Evans, J. J., Alderman, N., Burgess, P. W., & Emslie, H. (1997). Behavioural assessment of the dysexecutive syndrome. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 239-250). Hove: Psychology Press.

Table 1: Virtual Errands Test score

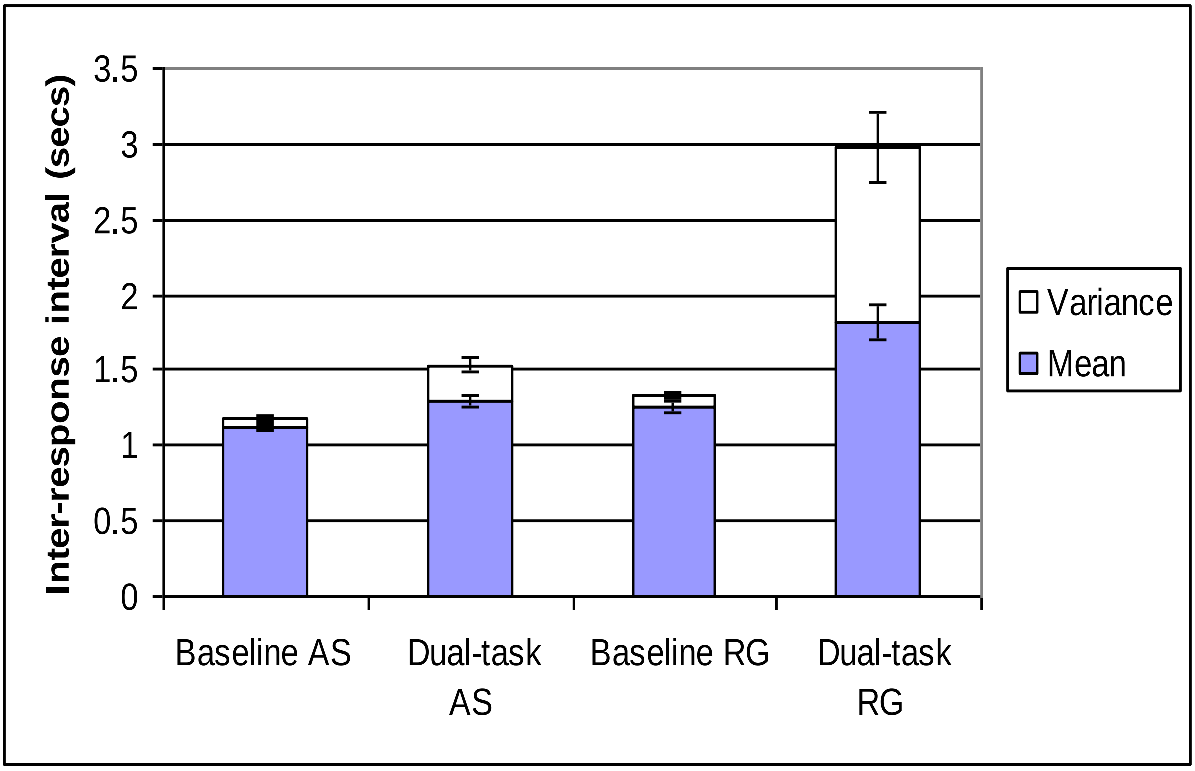
	Single task performed first		Dual task performed first	
	Random Generation Group	Articulatory Suppression Group	Random Generation Group	Articulatory Suppression Group
Single Task Performance **Mean** (SD)	**5.70** (1.83)	**7.45** (1.21)	**8.20** (1.23)	**8.20** (1.81)
Dual Task Performance **Mean** (SD)	**4.20** (3.65)	**7.91** (1.76)	**6.00** (1.89)	**5.70** (2.11)

Table 2: Measures of recall of the errand list in percentages

		Before the VET		After the VET	
		Free recall	Cued recall	Free recall	Cued recall
Random Generation Group **Mean** (SD)	Single Task Performance	**51.19** (14.13)	**87.05** (14.32)	**81.90** (17.78)	**85.00** (18.42)
	Dual Task Performance	**53.57** (13.89)	**85.38** (13.19)	**81.19** (13.13)	**86.95** (13.09)
Articulatory Suppression Group **Mean** (SD)	Single Task Performance	**61.67** (19.32)	**90.43** (11.08)	**80.48** (16.95)	**88.95** (11.57)
	Dual Task Performance	**51.90** (13.46)	**90.14** (10.42)	**81.67** (17.27)	**88.95** (11.95)

Figure 1: A screenshot from the virtual environment

Figure 2: Rate of utterance in the secondary tasks



Appendix 1: Errand Lists

Task Set 1

Start: 3 pm in room F12
End: 3.08pm promptly, back in room F12

Stairwell A = down only   Stairwell B = up only

- Collect a phonecard from the cupboard in room **T10**.
- Find out the date of a memory exam from the door of **F15,** and then pass this information on to a colleague in **F17**.
- Collect some milk from room **F5**.
- Pick up a borrowed computer disk from **S15** and return it to **S12** where it is needed, **before 3.04pm**.
- Find out the day the laboratory in **T15** is vacant.
- Collect some exam papers from **T1** and deliver them to **S1, as close to 3.07pm as possible.**
- Find out how many chairs are in **T20**.

Task Set 2

Start: 1pm in Room S1
End: 1.08pm promptly, back in Room S1

Stairwell A = down only   Stairwell B = up only

- Pick up a folder from **S7**.
- Go to the door of **T10** and find out what time a new lecturer will be arriving at the airport today.   Then go to **T1** to phone and book a taxi for him.
- Pick up some bread from **F15**.
- Find out if there is a camera in **T15**, then pass this information on to a technician in **T13**, **before 1.04pm.**
- Find out what time the laboratory in **T14** is free.
- Pick up at class register from **F18** and deliver it to **S10, as close to 1.07pm as possible**.
- Pick up a poster from **S9**.