Set Shifting in Problem-Solving
To appear in *Psychological Research*
[Accepted 10th February, 2015]

Set-Shifting as a Component Process of Goal-Directed Problem-Solving

Richard P. Cooper and Verity Marsh
Department of Psychological Sciences,
Birkbeck, University of London

Abstract: In two experiments, we compared secondary task interference on Tower of London performance resulting from three different secondary tasks. The secondary tasks were designed to tap three different executive functions, namely set-shifting, memory monitoring and updating, and response inhibition. Previous work using individual differences methodology suggests that, all other things being equal, the response inhibition or memory tasks should result in the greatest interference. However, this was not found to be the case. Rather, in both experiments the set-shifting task resulted in significantly more interference on Tower of London performance than either of the other secondary tasks. Subsequent analyses suggest that the degree of interference could not be attributed to differences in secondary task difficulty. Results are interpreted in the light of related work which suggests that solving problems with non-transparent goal/subgoal structure requires flexible shifting between subgoals – a process that is held to be impaired by concurrent performance of a set-shifting task.

1 Introduction

A substantial body of empirical work supports the view that cognitive processing involves the operation of multiple distinct executive or control functions that operate in concert to produce complex goal-directed behaviour. Some of the putative executive functions identified by this work (e.g, response inhibition: Logan, 1994; task setting and/or switching: Allport et al., 1994; Meiran, 1996; Monsell, 2003, Vandierendonck, Liefooghe, & Verbruggen, 2010; and working memory control functions of monitoring, maintenance, updating and gating: O’Reilly & Frank, 2006; Rave et al., 2007) are computationally relatively simple and plausibly analysed as “atomic” function. Others (e.g., rule induction: Reverberi et al., 2005; and planning: Shallice, 1982) would seem potentially to be computationally complex. For example, planning conceivably involves the operation of numerous sub-functions in the control of verbal or visuo-spatial working memory and in the generation and manipulation of hypothetical future states of the world. This paper concerns the relationship between the simpler or more basic executive functions and the executive requirements of one of these computationally more complex functions – planning – within the context of goal-directed problem-solving.

The Tower of London (ToL: Shallice, 1982) is a conceptually simple puzzle that involves rearranging objects (coloured balls on sticks of different lengths) from a given start state to obtain a desired or goal state, subject to a set of constraints. Participants are typically required

The final publication is available at Springer via http://dx.doi.org/10.1007/s00426-015-0652-2

to solve a series of ToL problems, each in the least number of moves. The task was originally developed as a means of assessing planning, which Shallice (1982) argued was subject to selective impairment following lesions to the left frontal lobe. While some subsequent neuropsychological studies have failed to replicate the left frontal deficit reported by Shallice (e.g., Cockburn, 1995, but see Glosser & Goodglass, 1990, and Morris et al., 1997), the task remains popular as a tool for assessing executive functioning, particularly in clinical and neuropsychological populations (e.g., Donders & Larsen, 2012; Köstering et al., 2012; Marzocchi et al., 2008; Rainville, et al. 2012; see Sullivan et al., 2009, for a meta-analysis) as well as in normally and atypically developing children and adolescents (e.g., Albert & Steinberg, 2011; Bull et al., 2004; Hartman et al., 2010; Kaller et al., 2008; Luciana et al., 2009).

Planning is a complex process that is likely to involve multiple subprocesses or subfunctions interacting in the generation, maintenance, and execution of a plan. Consistent with this, previous research has found that performance on the ToL is correlated with performance on a range of other tasks which have been held to tap more basic executive functions. For example, Asato et al. (2006) found age-related improvements in ToL performance of adolescents were related to improved working memory and response inhibition (see also Albert & Steinberg, 2011, for similar findings with adolescents). These results echo those of Welsh et al. (1999) who, using an individual differences methodology, found that processes or functions of response inhibition and visuo-spatial working memory were critical in the production of successful shortest-path plans for ToL problems. Moreover, using a related tower-task – the Tower of Hanoi (Simon, 1975) – Miyake and colleagues (2000) also found support for response inhibition as a component process of planning, as discussed in more detail below.

Notwithstanding the findings reported above, the evidence for response inhibition in planning is mixed. For example, in an individual differences study of 4-5 year old children, Bull et al. (2004) found no relation between the efficacy of response inhibition and performance on simple ToL problems. Bull and colleagues also tested their participants’ set-shifting ability, and this too was found to be unrelated to performance on simple ToL problems. On complex problems, however – those that required intermediate moves where balls were apparently moved away from the goal state – they found that both set-shifting and response inhibition were predictive of performance. The authors argue that the solution of such problems requires not only the inhibition of inappropriate but tempting moves, but also flexible shifting between goals and subgoals. The latter is held to underlie the correlation between performance on set-shifting tasks and complex ToL tasks.

In a related vein, Kaller et al. (2008) found in a tightly controlled ToL study with 4-5 year old children that, while there were developmental improvements in planning ability, there was no direct relation between those improvements and measures of either (verbal or visuo-spatial) working memory or response inhibition in their sample of participants. Based on an analysis of performance on individual ToL problems with specific characteristics (e.g., those needing intermediate moves for their solution), the authors instead attribute developmental gains to improvements in the ability to look ahead, which, they argue, are distinct from those related to working memory or response inhibition.

These developmental results cannot necessarily be extrapolated to a model of adult performance, as the behaviour of younger participants might be compromised by immaturity of their executive system. Limited visuo-spatial working memory, for example, may alter the apparent association with set-shifting or response inhibition relative to the adult system. Nevertheless, Kaller et al. (2008) demonstrate that planning involves more than working memory and response inhibition, while Bull et al. (2004) provide both correlational evidence and a plausible role for the function of set-shifting in complex planning tasks.

The majority of studies cited above use individual differences methodology. Thus, their results are all correlational. In contrast, Phillips et al. (1999) adopted a dual-task experimental approach, where subjects were required to complete ToL problems while engaged in a secondary task. Four secondary tasks were considered: articulatory suppression, random number generation, spatial pattern tapping and random tapping. These tasks were held to involve different combinations of a) spatial working memory, b) verbal working memory, and c) central executive processes (in the sense of Baddeley, 1996). Phillips et al. found that ToL performance was impaired in conditions where the secondary task taxed central executive processes (their random number generation and random tapping conditions), but not in the articulatory suppression condition. This underlines the importance of executive processes in performance of the task. It also argues against any specific verbal mediation in the task. (See also Cheetham et al., 2012, for a similar demonstration using a dual-task methodology of the specific importance of visuo-spatial, rather than verbal, working memory in the solution of ToL problems.)

A further area of debate concerns whether similar cognitive processes are involved in the solution of the Tower of London task and the superficially similar Tower of Hanoi (ToH: Simon, 1975) task. The Tower of Hanoi also involves rearranging objects (in this case disks of different sizes onto three pegs, each of the same height) from a start state to match a goal state, but differs from the ToL in the constraints imposed and the subsequent structure of the solution space. It is therefore possible that the different tasks make different demands on executive processing, and direct comparisons of performance between the tasks (Welsh et al., 1999; Bull et al., 2004) suggest that ToL is more reliant on response inhibition than ToH (but see Zook et al., 2004, for evidence suggestive of the reverse).

Yet studies that have examined the executive requirements of ToH have also found response inhibition to be a key component. The most substantial such study is the individual differences study of Miyake et al. (2000) referred to above. Miyake and colleagues considered the role of three putative executive functions – *set-shifting*, *memory monitoring and updating*, and *response inhibition* – on performance of a range of complex tasks, one of which is the Tower of Hanoi task. The authors' analysis suggests that performance on that task, as measured by the number of moves required to solve two target problems, is associated with the construct of response inhibition, rather than either of two other executive constructs. Thus, participants who performed well on a range of simple response inhibition tasks tended to complete the target problems in fewer moves than participants who performed poorly on those tasks. A significant difficulty with this analysis is the statistical procedure used to support the conclusion (Cooper et al., 2012), which involves estimating latent factors corresponding to three executive functions and then comparing fits of structural equation models built upon these latent factors. Any error in estimating the latent factors will be compounded in the structural equation models, limiting the extent to which strong conclusions can be drawn from subtle differences in the fits of contrasting models.

Over the last 15 years the focus of work with the Tower of London has shifted from component subprocesses to the localisation of subprocesses. In one recent study, Kaller et al. (2011b) explored the effect of two specific factors relating to different ToL problems – whether the goal state unambiguously determines the order of subgoals and whether intermediate moves were required in order to solve a ToL problem – on the neural correlates of ToL solution. With regard to the former factor, previous research has shown that people typically solve Tower problems by focussing on one disk or ball at a time – getting that disk/ball into its correct target position before moving on to the next disk/ball (Gilhooly et al., 1999). Consistent with such a strategy, problems where the goal state involves all three balls on the same peg (i.e., problems with an unambiguous subgoal ordering) are more readily solved than those in which the goal state involves all three balls on different pegs (i.e., where the order in which subgoals should be achieved is ambiguous: Kaller et al., 2008; Klahr & Robinson, 1981; Newman & Pittman, 2007; Waldau, 1999). Both factors were found by Kaller et al. (2011b) to increase dorsolateral prefrontal cortical activity (DLPFC, as measured by BOLD), but ambiguous subgoal structure was found to differentially recruit left DLPFC while the need for intermediate moves was found to differentially recruit right DLPFC. While this study is suggestive of different cognitive processes being involved in different problems, it is difficult to identify the specific processes without the development of a complete process model of ToL performance and its application to the different types of problem.

As discussed above, previous behavioural approaches to investigating the involvement of executive processes in the solution of tower tasks have largely used individual differences methodology. Such approaches may be criticised for their inherently correlational nature. The dual-task studies of Phillips et al. (1999) and Cheetham et al. (2012), in contrast, involved experimental manipulation. However, in the Phillips et al. study the secondary tasks were relatively complex. Random generation, for example, appears to draw upon multiple executive functions (Towse & Neil, 1998; Miyake et al., 2000; Cooper et al., 2012). And in the dual-task experiment of Cheetham et al. (2012) the focus was on verbal versus visuo-spatial working memory rather than on executive functions more generally. The approach of the present study, following Cooper et al. (2012), focuses on the differential involvement of specific executive functions. We investigate the interference effects of relatively simple executive tasks on primary task performance. In particular, we contrast interference resulting from simple set-shifting, memory monitoring / updating, and response inhibition tasks on completion of ToL problems. The approach therefore assumes that the central executive is fractionable or componential and aims to determine the role of specific executive components in ToL performance.

2 Experiment 1

If good performance on the Tower of London does indeed require response inhibition, and if response inhibition is a finite cognitive resource, then concurrent performance of a response inhibition task while also completing the ToL should lead to greater interference on the ToL than concurrent performance of a simple executive task that taps some other executive function (such as tasks taping set-shifting or verbal memory monitoring and updating). On the other hand, if shifting task set is a major factor in successful solution of ToL problems, then ToL performance should be impaired when the ToL is performed while also performing a set-shifting task – more so than if it is performed whilst also performing equivalently difficult tasks that do not involve set-shifting (such as response inhibition or memory monitoring / updating tasks). Experiment 1 was designed to test this hypothesis by measuring interference effects on Tower of London performance when simultaneously performing a range of simple executive function tasks.

2.1 Materials and Methods

2.1.1 Participants

Forty-two participants (10 male, 32 female; average age 29 years 1 month; age range 18 years to 48 years) completed the experiment. Participants were recruited from the Birkbeck's volunteer participant panel (which includes mature students at various levels and lay-people with a general interest in psychology) and received either partial course credit or £5 for their participation.

2.1.2 Design

The experiment used a repeated measures design where each participant completed a set of Tower of London problems four times. In block 1, which served as a baseline condition, participants completed the problems with no secondary task. In blocks 2, 3 and 4, they also completed one of three secondary tasks – the digit-switching task, the 2-back task or the go/no-go task (each as described below) – with the order of those secondary tasks fully counterbalanced across participants. For each ToL problem, each participant's solution path and time per move was recorded, allowing extraction of a range of dependent variables as described below. Accuracy and response time on the secondary tasks was also recorded.

2.1.3 The Tower of London (ToL) Task

The Tower of London task was administered to participants via a graphical computer program written in MatLab and running on a standard IBM-compatible PC with a 17" monitor. On each trial, the screen showed the current state of the ToL apparatus and, in the top left corner, a depiction of the target or goal state. Participants manipulated the current state by using the mouse to drag and drop the coloured balls from one peg to another. When the current state matched the goal state, a button icon on the lower left of the screen displayed the text "Next Trial". When clicked, this button advanced the participant to the next problem. The computer program did not tell subjects how many moves were required for each problem, but ensured that only physically possible moves could be made. It also recorded all dependent variables (moves made and time of each move).

A total of eight variant ToL problems were used in the experiment – two for the practice trials and six (referred to as A to F) for each block of experimental trials (see appendix 1). The experimental trials included a range or problem types. Thus, problem A required 4 moves, problems B, D and E required 5 moves, and problems C and F required 6 moves. Moreover, problem A had a flat goal state (meaning that the order in which balls should reach their target positions was completely ambiguous), while all other problems had goal states consisting of two balls on one peg and one on another. Finally, successful completion of problems E and F in the minimal number of moves involved in both case resisting the temptation, on the first move, to move a ball to its desired location. In order to reduce the likelihood of participants learning solution paths for specific problems, the ball colours were randomly permuted on each block (so for experimental problem A, for example, the ball initially on the smallest peg might be red on one block but green on another block; cf. Berg & Byrd, 2002). This method has been used in previous ToL research to present multiple ToL problems to participants while controlling for problem structure (e.g., Kaller et al., 2008, 2011b). The order of presentation of the six problems was also randomised within each block.

2.1.4 Secondary Tasks

Set-shifting task: In the digit-switching task (DS: see Monsell, 2003), held to primarily tax set-shifting, participants heard digits (1, 2, 3, 4, 6, 7, 8 or 9) and were required to categorise them as either high/low (magnitude) or odd/even (parity), with responses given vocally. Initially, participants were asked to perform the magnitude categorisation, with digits above five classed as “high” and those below five as “low”. After the first four trials (and subsequently after each four trials), a tone was presented indicating that participants should switch their categorisation rule, alternating between magnitude and parity. Digits were presented at a rate of one every 2.5 seconds. The dependent measure was accuracy (the number of correct trials divided by the total number of trials).

Working memory monitoring and maintenance task: In the 2-back task (2B), held to primarily tax verbal working memory maintenance and monitoring functions, participants heard a series of digits (in the range 1 to 9) and were required to respond vocally (by saying “yes”) when a digit was identical to the digit presented two items earlier in the sequence. Digits were presented at a rate of one every 2.0 seconds. The dependent measure was accuracy, defined as the number of hits and correct rejections divided by the total number of trials.

Response inhibition task: In the go/no-go task (GnG), held to primarily tax the response inhibition function, participants heard a series of single or double tones and were required to indicate as quickly as possible (by saying “yes”) whenever the stimulus was a single tone, and to withhold their response whenever the stimulus was a double tone. Single tones occurred on 5 out of every 6 trials (thus establishing a prepotent response), and the interval between stimuli varied randomly from 1.5 seconds to 2.5 seconds. The dependent measure was again accuracy, as defined in the 2-back task.

2.1.5 Procedure

The procedure followed closely that used by Cooper et al. (2012) in their study of dual-task interference, with the only difference relating to administration of the primary task (the Tower of London). Thus, the nature of the experiment was first described to participants. Once participants had given their informed consent they sat at a desk on which was positioned the PC monitor, a mouse for interacting with the PC, and a microphone (placed between the participant and the monitor) to detect the onset of participant’s vocal responses to the secondary tasks. Participants then completed the two practice ToL problems to familiarise themselves with the interface and the task. Before attempting these trials, they were informed of the ToL rules and the method of dragging balls from one peg to another was explained. They were then told that they should “try to solve each problem in as few moves as possible” (but not how many moves were required for each problem). Participants then completed the six experimental trials.

In blocks 2, 3 and 4 participants completed the six experimental ToL problems while also completing one of the three auditory-vocal tasks. On each of these blocks participants first received practice on the auditory-vocal task. They then attempted to complete both the ToL problems and the auditory-vocal task concurrently. Order of problems was randomised within each block, and the order of secondary task in blocks 2, 3 and 4 was fully counterbalanced such that one third of participants completed the digit-switching task in block 2, one third completed the 2-back task in block 2, and one third completed the go/no-go task in block 2, and so on.

2.2 Results

In the analyses that follow we first focus on performance measures related to the primary task (Tower of London), considering four specific dependent variables: proportion correct (i.e., proportion of problems solved in the minimum number of moves), excess moves (i.e., total number of moves beyond the minimum number taken to solve the six problems), mean first move time (i.e., average time between problem presentation and making the first move), and mean time per subsequent move (i.e., total time taken, after the first move, to solve all six problems divided by total number of moves taken beyond the first move). In each case we report results from a one-within analysis of variance of the dependent variable over conditions. This analysis addresses the question of whether the different secondary tasks have different effects on ToL performance as a whole. Following this we report an item analysis to explore whether performance on particular ToL problems was differentially affected by experimental condition. Lastly, we report an analysis of secondary task data, including correlational statistics concerning relations between primary and secondary task performance.

2.2.1 Effects of Secondary Task Performance on ToL Dependent Measures

Figure 1 shows condition means for the four ToL dependent measures. One-way within-subjects analyses of variance, with condition as the independent variable, revealed significant main effects of condition on all four dependent measures (problem solved in minimum moves: $F(3, 123) = 3.31$, $p = 0.019$, $\eta^2 = 0.077$; excess moves: $F(3, 123) = 6.788$, $p < 0.001$, $\eta^2 = 0.142$; first move time: $F(3, 123) = 20.862$, $p < 0.001$, $\eta^2 = 0.337$; and mean time per subsequent move: $F(2.361, 96.806) = 11.474$, $p < 0.001$, $\eta^2 = 0.219$)[1]. Planned comparisons across the secondary task conditions, assuming a critical p-value of 0.008 to correct for multiple comparisons (i.e., 0.05 divided by 6), confirmed that:[2]

a) While fewer problems were solved in the minimum number of moves during the set-shifting condition than any other condition, the between-condition differences were not significant at the corrected level of α (set-shifting < memory: $t(41) = 0.498$, $p = 0.311$; set-shifting < response inhibition: $t(41) = 2.125$, $p = 0.020$; memory < response inhibition: $t(41) = 1.704$, $p = 0.048$);

b) More excess moves were produced during the set-shifting and the memory conditions than in the response inhibition condition (set-shifting > response inhibition: $t(41) = 2.895$, $p = 0.003$; memory > response inhibition: $t(41) = 2.487$, $p = 0.009$), though the latter difference did not quite survive correction for multiple comparisons;

c) Time to first move did not differ significantly between the set-shifting and memory conditions, but was significantly longer in those conditions than in the response inhibition condition ($t(41) = 3.999$, $p < 0.001$, and $t(41) = 2.756$, $p = 0.004$, respectively); and

---

[1] Because of substantial skewness in some of the dependent measures, excess moves was square root transformed prior to analysis, while first move time and time per subsequent move were log transformed. The sphericity assumption was violated for the ANOVA for time per subsequent move, and so Greenhouse-Geisser corrections to degrees of freedom have been made for the analysis of variance of this measure.

[2] Recall that in all cases the baseline condition was performed before the dual task conditions, but that the order of dual task conditions was counterbalanced over participants. Performance of the primary task on the dual task conditions therefore potentially benefits from prior practice on the single task condition but suffers from the requirement to concurrently perform a secondary task. It is consequently not possible to interpret pairwise comparisons between the baseline condition and each dual task.

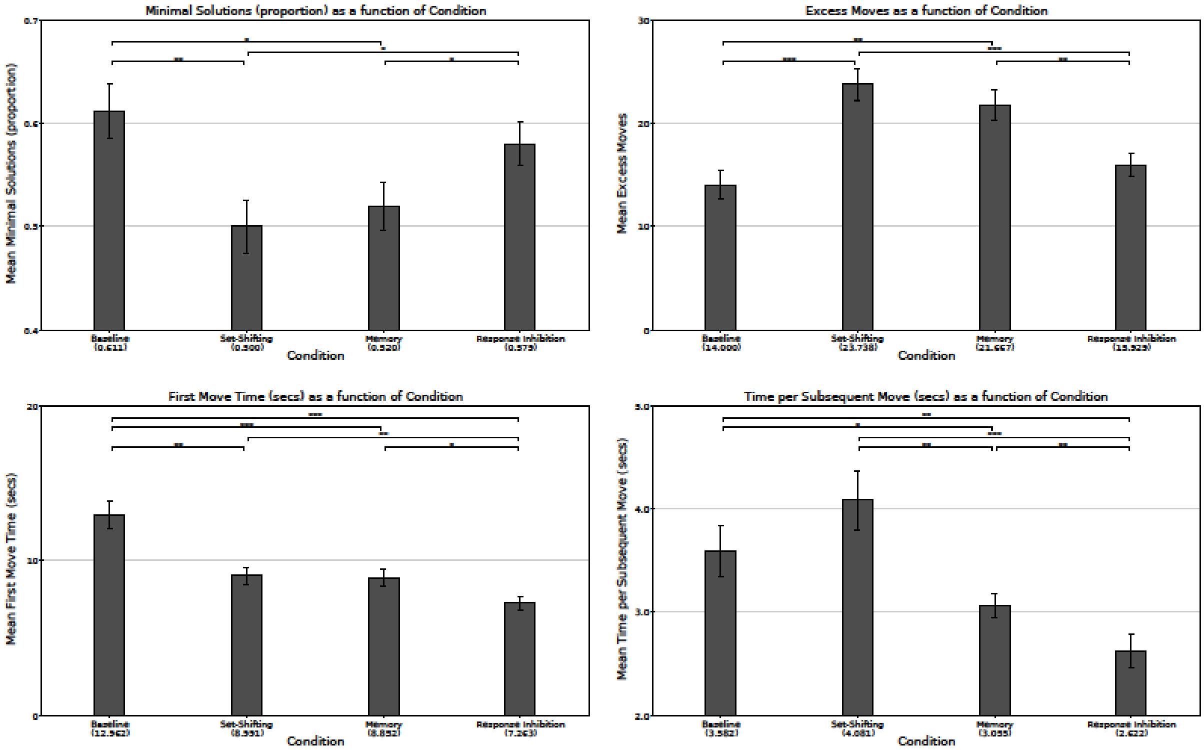

Fig. 1: Bar charts showing mean values of all dependent measures in the four conditions of experiment 1. Error bars indicate one standard error from the mean, corrected for the within-subject comparison. * = significant at p = 0.05 (uncorrected); ** = significant at 0.01 (uncorrected); *** = significant at 0.001 (uncorrected).

d) Time per subsequent move was significantly longer in the set-shifting condition than in the memory condition ($t(41) = 3.861$, $p < 0.001$), and in the memory condition than in the response inhibition condition ($t(41) = 2.832$, $p = 0.004$). Unsurprisingly, the difference between the extremes (set-shifting condition and response inhibition condition) was also statistically significant ($t(41) = 6.099$, $p < 0.001$).

To summarise these results, across the three secondary task conditions, while pair-wise comparisons were not in all cases significant, a consistent rank ordering of interference effects was observed, with concurrent performance of the response inhibition task resulting in least interference and concurrent performance of the set-shifting task resulting in greatest interference. The only exception to this was for first move time, which did not differ between the set-shifting and memory conditions.

2.2.2 Item Analysis

Each of the four dependent measures described above was subjected to a further analysis of variance with two within-subjects factors: condition and problem. In three out of four cases, problem did not interact with condition, though in all cases there were significant main effects of problem (see Table 1). Further analysis suggests that in the one case where an interaction was observed (first move time), the effect is due to a difference between the baseline condition and the second task conditions, rather than a result of any specific secondary task. These results suggest a general effect of problem difficulty, with some problems being easier (e.g., problem A), regardless of condition, than others (e.g., problem F). Thus, while care must be taken in interpreting what is essentially a null result, particularly as structural characteristics of problems were not systematically varied within the problem set, the effect of problem was not modulated by the secondary task.

Table 1: Analysis of variance results for two-within (condition by problem) factor analyses for each dependent variable of experiment 1. Note the lack of interaction effects in all cases except first move time.

Dependent variable	*Effect*	*F statistic*	*Significance*	*Effect Size*
Problems solved in minimum number of moves	Condition	F(3, 123) = 3.431	0.019	0.077
	Problem	F(5, 205) = 28.273	0.001	0.408
	Condition × Problem	F(15, 615) = 1.313	0.188	0.031
Excess moves	Condition	F(3, 123) = 6.725	0.001	0.141
	Problem	F(5, 205) = 19.408	0.001	0.321
	Condition × Problem	F(15, 615) = 1.327	0.180	0.031
First move time	Condition	F(3, 123) = 22.436	0.001	0.354
	Problem	F(5, 205) = 6.429	0.001	0.136
	Condition × Problem	F(15, 615) = 2.023	0.012	0.047
Time per subsequent move	Condition	F(3, 123) = 17.425	0.001	0.298
	Problem	F(5, 205) = 19.683	0.001	0.324
	Condition × Problem	F(15, 615) = 1.466	0.112	0.035

2.2.3 Analysis of Secondary Task Data

Accuracy on the auditory vocal tasks was generally high (DS: 78.5%; 2B: 85.3%; GnG: 87.1%), with the digit-switching task being performed significantly more poorly than either of the other auditory vocal tasks (DS versus GnG: $t(41) = 4.624$, $p < 0.001$; DS versus 2B: $t(41) = 3.086$, $p = 0.002$). Accuracy on all three auditory vocal tasks was correlated (DS versus 2B: $r = +0.206$, $N = 42$, $p = 0.048$; DS versus GnG: $r = +0.533$, $N = 42$, $p < 0.001$; 2B versus GnG: $r = +0.244$, $N = 42$, $p = 0.030$).

During the set-shifting condition, accuracy on the secondary task (the digit-switching task) was not significantly correlated with any dependent measure on the primary task. In contrast, during the memory condition accuracy on the secondary task (the 2-back task) was positively correlated with time to first move ($r = +0.269$, $N = 42$, $p < 0.05$) and marginally negatively correlated with number of excess moves ($r = –0.187$, $N = 42$, $p = 0.059$), suggesting that participants who did well on the 2-back task generally solved ToL problems relatively quickly and with relatively few excess moves. In the response inhibition condition, a positive correlation between accuracy on the go/no-go task and number of problems solved in minimum number of moves ($r = +0.209$, $N = 42$, $p < 0.05$) was found in conjunction with a marginally negative correlation between go/no-go accuracy and number of excess moves ($r = –0.187$, $N = 42$, $p = 0.059$), again suggesting that good performance on the secondary task was generally indicative of good performance on the primary task. Consequently there is no support for the conjecture that participants traded off performance on the two tasks.

2.3 Discussion

When participants were asked to complete Tower of London problems while simultaneously performing one of three simpler tasks, each held to tap a different executive function, it was

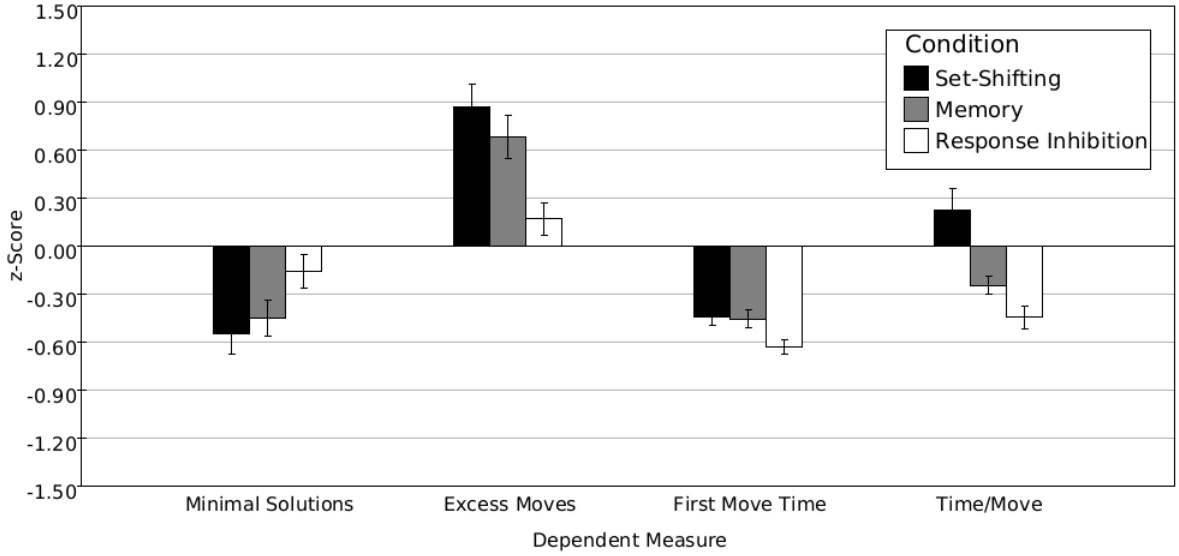

Fig. 2: Interference effects of the three secondary tasks on the four dependent measures. In order to allow comparison across dependent variables and across conditions, interference effects are calculated as the difference between the baseline mean and the mean in the experimental condition, converted to a z score by dividing by the standard deviation of the dependent measure in the baseline condition. Error bars represent one standard error of the dependent variable, corrected for within-subject comparisons.

found that greatest interference (as determined by multiple measures) occurred with a set-shifting task (the digit-switching task of Monsell, 2003), while least interference occurred with a response inhibition task (the go/no-go task). The working memory monitoring and updating task (the 2-back task) lead to moderate interference. This ordering of interference is seen most clearly in Figure 2, which shows interference effects as *z*-scores calculated based on the mean and standard deviation of the baseline (i.e., single-task) condition. Across three of the four dependent variables, there is a decline from the set-shifting condition to the memory condition to the response inhibition condition.[3] These results may seem surprising, given that Tower of London performance has been held to depend on response inhibition (e.g., Albert & Steinberg, 2011; Asato et al., 2006; Welsh et al., 1999).

There are at least five hypotheses that might be advanced to account for the results. First, the observed effects may simply reflect secondary task difficulty: Possibly digit switching is the hardest secondary task, while go/no-go is the easiest, and so the rank ordering of interference effects may reflect the rank ordering of secondary task difficulty. In support of this hypothesis is the fact that secondary task performance on digit-switching (78.5%) was significantly poorer than on either of the other secondary tasks (2B: 85.3%; GnG: 87.4%). Moreover, during debriefing many subjects spontaneously reported that the go/no-go task was least effortful. Note though that when in another experiment the same auditory-vocal tasks were paired with a different primary task – random generation (Cooper et al., 2012) – it was the 2-back (memory) task and not the digit-switching (set-shifting) task that was performed most poorly and that led on several measures to greatest interference. Thus, the digit-switching task does not appear to be intrinsically more difficult than the 2-back task.

---

[3] The difference between the dependent variable values in the baseline condition and the experimental conditions is presumably due to a combination of the baseline condition being easier because it does not involve dual tasking, but harder because it is unfamiliar. Recall that the baseline condition was completed first by all participants but order of the three dual-task conditions was counterbalanced, so the dependent variables across the dual-task conditions are directly comparable.

A second possibility is that the observed effect in some way reflects the different time course of the secondary tasks. The trials of the secondary tasks were not all of equal duration, with digit-switching trials lasting 2.5 seconds, 2-back trials lasting 2.0 seconds, and go/no-go trials varying between 1.5 seconds and 2.5 seconds. These differences may have impacted upon primary task performance in some unanticipated way. For example, the slightly longer trial time for the digit-switching task (2.5 seconds) may have resulted in participants exploring potential solution paths on the ToL for slightly longer before being interrupted by the secondary task than when the secondary task was the 2-back task (where trials lasted 2.0 seconds). However given that trials were longer in the digit-switching task than the 2-back task it is unclear why this could have favoured the memory condition at the expense of the set-shifting condition.

A third possibility is that the secondary tasks are not “process-pure”. That is, they do not tap a single executive function. This is very likely to be the case. The phenomenological experience of performing the digit-switching task, for example, is that it requires response inhibition as well as set-shifting, as when one has to switch from high/low responses to odd/even responses, and given a stimulus such as “3” on an odd/even trial one must inhibit the “low” response in favour of the “odd” response. Given the secondary tasks employed in this study, this possibility cannot be ruled out.

Fourthly, the methodology assumes that executive functions are divisible cognitive resources that can be shared between concurrent tasks, and that this sharing results in less of the resource for each of the concurrent tasks. This assumption may not apply for all simple executive functions. For example, suppose response inhibition is a more global “stop” process that results in suspension of *all* prepared responses (cf. Coxon et al., 2009). If this is the case, then inhibiting a response in the go/no-go task would result in inhibition of all responses (including those in the primary task). This could act to *facilitate* performance on the ToL task, by helping to prevent the selection of locally desirable but globally incorrect moves (as in problems E and F). The lack of interaction between experimental condition and problem in the item analysis argues against such a differential affect.

Finally, and as suggested in the introduction, it may be that the critical executive function for successful ToL performance is not response inhibition, as the studies by Albert and Steinberg (2011), Asato et al. (2006) and Welsh et al. (1999) suggest, but set-shifting. More specifically, as argued by Bull et al. (2004), it may be that the ability to shift set plays a critical role, and indeed that the set-shifting requirement has a greater effect on the various performance measures than any possible response inhibition requirement. There are two reasons to be cautious about accepting this hypothesis. First, Bull et al.’s result concerned only the solution of relatively complex problems that required intermediate moves where balls were apparently moved away from the goal state and not the solution of simple problems. Second, generalisation or extension of the results from developmental studies of executive function to adult executive function is not straightforward given that executive functions continue to develop throughout adolescence.

3 Experiment 2

Given the above arguments, further empirical work is required to determine which, if any, of the previous hypotheses are plausible. A key concern is that the interference effects observed in experiment 1 are a consequence of some specific aspect of the secondary tasks beyond their executive requirements (e.g., task difficulty or time course of the different secondary tasks). Experiment 2 addresses this concern by coupling the Tower of London with a further set of secondary tasks: the plus/minus task, a variant on the 2-back task, and a stop-signal task. As in experiment 1, these secondary tasks have been held to tap the different executive functions of set-shifting, memory updating and monitoring, and response inhibition. Moreover, as in experiment 1 these secondary tasks can be administered aurally with verbal responses. If the results of experiment 1 are an artefact of the specific secondary tasks used, then there is no reason to expect that repeating the study with different secondary tasks would yield similar patterns of interference. If, on the other hand, the interference pattern is a function of the executive requirements of the secondary tasks, then similar patterns of interference should occur. Thus, based on the results of experiment 1, we hypothesise that simultaneous performance of the plus/minus task will lead to greater interference than simultaneous performance of either of the other secondary tasks.

3.1 Materials and Methods

3.1.1 Participants

Forty-three participants (13 male, 30 female; average age 29 years 11 months; age range 19 years to 74 years) completed the experiment. Participants were mature students and lay-people with a general interest in psychology. Participants received either partial course credit or £5 for their participation.

3.1.2 Design

As in experiment 1, a repeated measures design was used with the baseline condition followed by three dual task conditions. Again, the order of dual task conditions was counterbalanced. The primary task was again the Tower of London, though a different set of ToL problems were used (see appendix 1, figure 6). Again, different auditory-vocal secondary tasks were used in each of the three dual task conditions. While these differed from the tasks used in experiment 1, they were designed primarily to tap the same three executive functions.

3.1.3 Secondary Tasks

Set-shifting task: In the plus-minus task, adapted from Jersild (1927), participants were presented on each trial with a number in the range of 12 to 98 (inclusive), spoken in either a male or female voice. Participants were instructed to add one to the number if it was spoken in a female voice and to subtract one if it was spoken in a male voice. Voices alternated on successive trials, meaning that participants were required to switch every trial between addition and subtraction, and this was explained to participants prior to commencing the task. Participants were instructed to say their answer as quickly as possible into the microphone provided.

Working memory monitoring and maintenance task: The 2-back task was again used as a secondary task to tax memory monitoring and updating. The version used in experiment 2 differed from that used in experiment 1 in two respects. First, participants were required to respond on every trial (saying “yes” if the stimulus was a match and “no” otherwise), rather than just responding on matching trials. Second, the rate of presentation was slowed to one item every 2.7 seconds. Both changes were designed to improve comparability of the three secondary tasks. Thus, the former ensured that a verbal response was required on all trials of all secondary tasks while the latter allowed a consistent pace across the three secondary tasks.

Response inhibition task: In the categorisation/stop task, adapted from Logan (1994), participants were presented with nouns, and required to say “yes” if the noun corresponded to a food item and “no” otherwise. On one in six trials the noun was followed (1000 milliseconds after its onset) by a tone. Participants were instructed to withhold their response on these trials. The length of delay was set following pilot work which suggested that participants were generally, but not always, successful at stopping their response on such trials. For the stimuli, 116 nouns (including 46 food items) of three to six letters were selected from the MRC psycholinguistics database (Wilson, 1988).

All three secondary tasks were performed at the same pace (2.7 seconds per trial), and in all cases the dependent measure was accuracy, defined as the number of correct trials divided by the total number of trials.

3.1.4 Procedure

As in experiment 1, after giving informed consent participants completed two practice ToL problems followed by six experimental problems (see appendix 1). Following this, participants completed the three dual-task blocks. Each of these blocks began with the participant being given 20 practice trials of the secondary task. The participant was then asked to complete the six experimental ToL problems while performing the secondary task. As in experiment 1, the order of the three secondary tasks was fully counterbalanced and subjects were given the opportunity to rest between blocks. Moreover, prior to each dual-task block participants were reminded to respond to the secondary task as quickly and accurately as possible. Furthermore, as in experiment 1, the mapping of colours to balls within the ToL problems was randomised on each block, as was the order of problems. Subjects were tested individually, with the subject sitting at a desktop PC in a quiet testing cubicle.

3.2 Results

3.2.1 Effects of Secondary Task Performance on ToL Dependent Measures

Figure 3 (cf. figure 1) shows condition means for the four ToL dependent measures considered in experiment 1 (i.e., proportion of problems solved in the minimum number of moves, excess moves, first move time and mean subsequent move time). In order to evaluate the effect of secondary task interference, one-way repeated-measures analyses of variance were performed on this data, with condition as the independent variable and dependent variables corrected for skewness as in experiment 1. These results revealed significant main effects of condition on all four dependent measures (problem solved in minimum moves: $F(3, 126) = 19.521$, $p < 0.001$, $\eta^2 = 0.317$; total excess moves: $F(3, 126) = 20.320$, $p < 0.001$, $\eta^2 = 0.236$; first move time: $F(3, 126) = 16.950$, $p < 0.001$, $\eta^2 = 0.288$; and mean time per subsequent move: $F(3, 126) = 4.099$, $p < 0.008$, $\eta^2 = 0.089$). Planned comparisons across the secondary task conditions, again assuming a critical *p*-value of 0.008 corrected for multiple comparisons, revealed that:

a) The proportion of ToL problems solved in the minimum number of moves in the set-shifting condition (i.e., when simultaneously performing the plus-minus task) was significantly less than in the memory condition ($t(42) = 3.222$, $p = 0.001$), and significantly less than in the response inhibition condition ($t(42) = 2.521$, $p = 0.008$);

b) The number of excess ToL moves produced during the set-shifting condition was more than during the memory condition, but this effect did not survive correction for multiple comparisons ($t(42) = 2.126$, $p = 0.020$). Similarly the total number of excess ToL moves was greater during the set-shifting condition than during the response

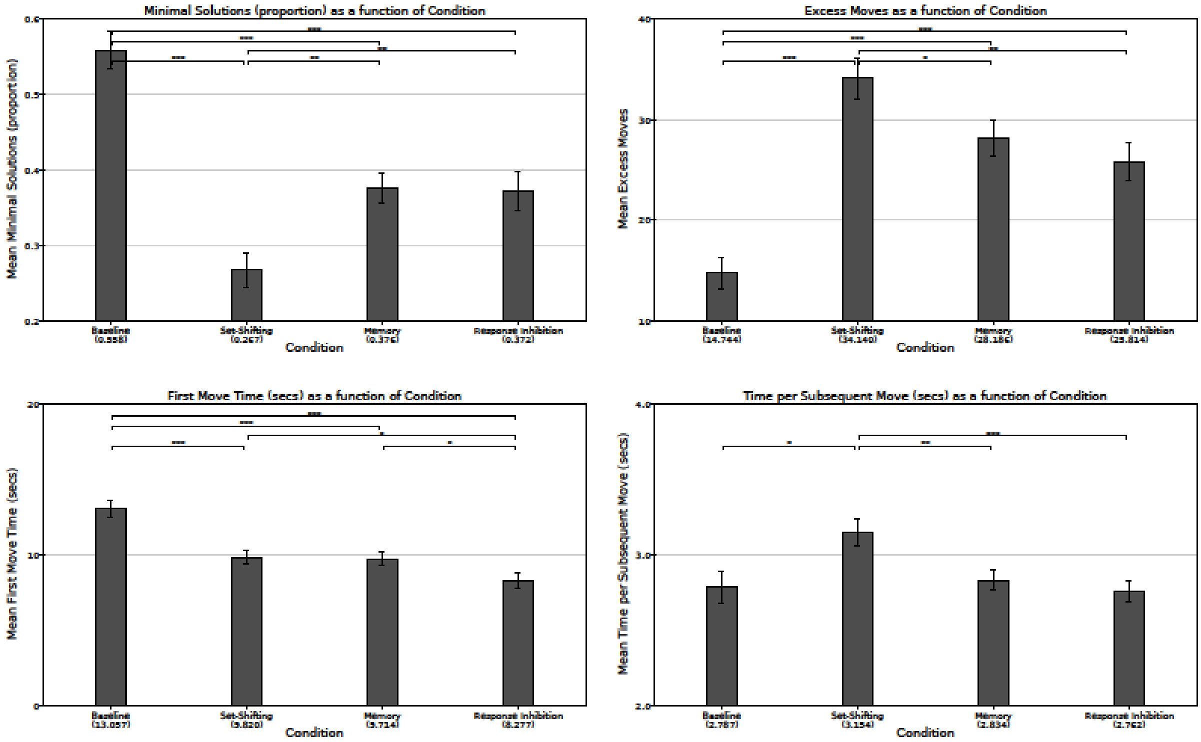

Fig. 3: Bar charts showing mean values of all dependent measures in the four conditions of experiment 2. Error bars indicate one standard error from the mean, corrected for the within-subject comparison. * = significant at p = 0.05 (uncorrected); ** = significant at 0.01 (uncorrected); *** = significant at 0.001 (uncorrected).

inhibition condition, though again this did not survive correction for multiple comparisons (*t*(42) = 2.431, *p* = 0.010);

c) The mean time to first move was similar in the set-shifting condition and the memory condition (*t*(42) = 0.329, *p* = 0.372), but it was longer in the set-shifting condition than in the response inhibition condition (*t*(42) = 2.265, *p* = 0.014), and in the memory condition than in the response inhibition condition (*t*(42) = 2.221, *p* = 0.016), though in both cases the difference was not significant at the adjusted alpha threshold; and

d) The mean time per move was significantly longer in the set-shifting condition than in both the memory condition (*t*(42) = 2.699, *p* = 0.005) and the response inhibition condition (*t*(42) = 3.520, *p* < 0.001). It was also longer in the memory condition than the response inhibition condition, though not significantly so (t(42) = 1.050, p = 0.150).

Thus, to summarise, performing the plus-minus task while attempting to also perform the ToL led to greater interference on ToL performance than simultaneous performance of either of the other secondary tasks. This was true for three of the four dependent variables: proportion of problems solved in the minimum number of moves, number of excess moves and time per subsequent move. As in experiment 1, only for time to first move was this pattern of interference not observed.

3.2.2 Item Analysis

Throughout, problem by condition analyses of variance revealed main effects of condition and problem, but no interactions (see Table 2). The main effects of problem indicate, as in

Table 2: Analysis of variance results for two-within (condition by problem) factor analyses for each dependent variable of experiment 2. Note the lack of interaction effects in all cases.

Dependent variable	*Effect*	*F statistic*	*Significance*	*Effect Size*
Problems solved in minimum number of moves	Condition	F(3, 126) = 19.521	0.001	0.317
	Problem	F(5, 210) = 6.529	0.001	0.437
	Condition × Problem	F(15, 630) = 1.033	0.418	0.024
Excess moves	Condition	F(3, 126) = 20.836	0.001	0.332
	Problem	F(5, 210) = 21.709	0.001	0.341
	Condition × Problem	F(15, 630) = 1.346	0.169	0.031
First move time	Condition	F(3, 126) = 21.630	0.001	0.340
	Problem	F(5, 210) = 3.839	0.002	0.084
	Condition × Problem	F(15, 630) = 0.867	0.602	0.020
Time per subsequent move	Condition	F(3, 126) = 14.475	0.001	0.256
	Problem	F(5, 210) = 6.596	0.001	0.136
	Condition × Problem	F(15, 630) = 0.396	0.980	0.009

experiment 1, that some problems are more difficult than others, but the lack of interaction again suggests that the secondary tasks did not have differential effects on the different problems.

3.2.3 Analysis of Auditory Task Data

Accuracy on the auditory vocal tasks was poorer than that in experiment 1 (plus-minus: 70.0%; 2-back: 76.3%; categorisation/stop: 63.6%), but significantly greater in the 2-back task than in either of the other tasks (versus plus-minus: $t(42) = 2.629$, $p = 0.006$; versus categorisation/stop: $t(42) = 3.427$, $p = 0.001$). Accuracy on all three tasks was positively correlated (plus-minus and 2-back: $r = +0.621$, $p < 0.001$; plus-minus and categorisation/stop: $r = +0.362$, $p < 0.01$; 2-back and categorisation/stop: $r = +0.236$, $p < 0.05$; $N = 43$) in all cases, implying that participants who were good at one secondary task tended to be good at the other secondary tasks.

During the set-shifting condition, accuracy on the secondary task (the plus-minus task) was negatively correlated with time per subsequent move ($r = -0.210$, $N = 43$, $p < 0.05$), but not with any other dependent measure on the primary task. Similarly, during the memory condition accuracy on the secondary task (the 2-back task) was negatively correlated with time per subsequent move ($r = -0.296$, $N = 43$, $p < 0.05$), but not with any other primary task dependent measure. During the response inhibition condition accuracy on the secondary task (the categorisation/stop task) did not correlate with any primary task dependent measures. Thus, and in contrast to experiment 1, there is some evidence of a possible trade-off between the primary and secondary tasks in two of the experimental conditions. Note however that the significance of the correlational statistics in support of this conclusion do not survive correction for multiple comparisons.

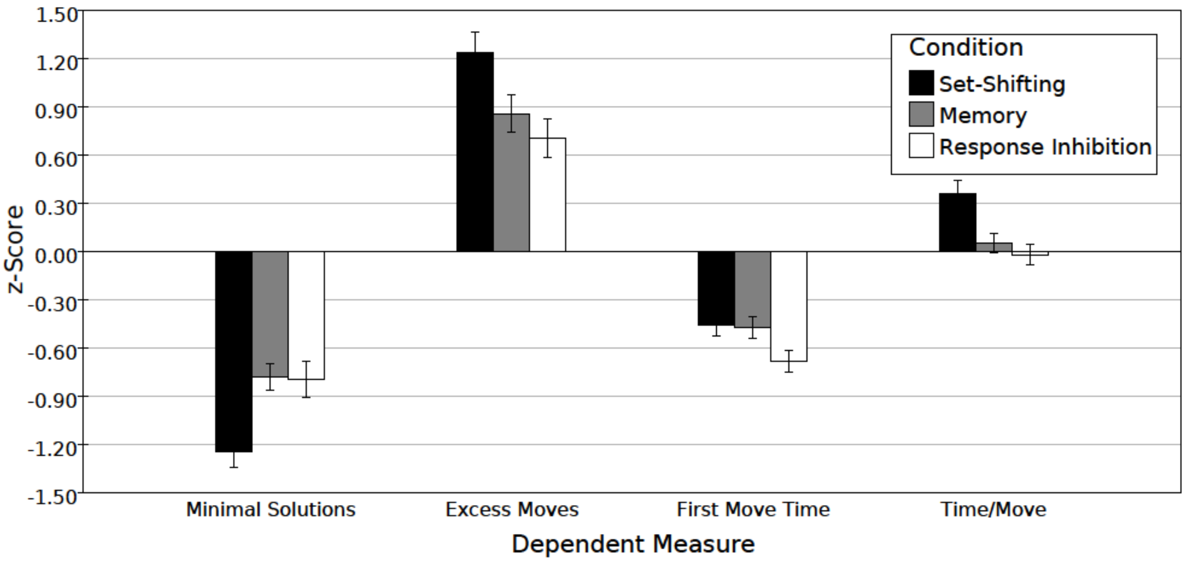

Fig. 4: Interference effects of the three secondary tasks on the four dependent measures. As in figure 2, interference effects are calculated as the difference between the baseline mean and the mean in the experimental condition, converted to a z score by dividing by the standard deviation of the dependent measure in the baseline condition. Error bars represent one standard error of the dependent variable, corrected for within-subject comparisons.

3.3 Discussion

The results indicate that, on three out of the four dependent measures, simultaneously performing the plus-minus task results in greater interference on ToL performance than simultaneously performing either of the other secondary tasks. On the fourth dependent measure, time to first move, the plus-minus task and the two-back task led to equivalent levels of interference (see Figure 4). This pattern of performance mirrors that observed in experiment 1: on all dependent measures except time to first move, more disruption to ToL performance is caused by concurrent performance of a set-shifting task than by concurrent performance of a working memory maintenance/monitoring task or of a response inhibition task.

In considering the results of experiment 1, we proposed five hypotheses that might account for the observed effects. The results of experiment 2 speak against four of these. First, it was suggested that the results of experiment 1 might be due to secondary task difficulty. If this were the case the findings reported here would be of little theoretical interest. However, in experiment 2 the same pattern of performance occurs with different secondary tasks that are held to tap the same three putative executive functions as in experiment 1. Moreover, in experiment 2 the secondary task with poorest performance was the categorisation/stop task (i.e., the response inhibition task), and not the plus-minus task (i.e., the set-shifting task). These results therefore further support the position argued for in the discussion of experiment 1, namely that secondary task difficulty is not behind the observed effects.

The second possibility was that participants might spontaneously adopt different time-sharing strategies for the primary and secondary tasks due to differences in response timing of secondary tasks. In experiment 2, trials in all three secondary tasks were of equal duration (2.7 seconds). The fact that the same pattern of primary task performance arises in both experiments therefore undermines any argument for differences in primary task performance as being due to an artefact of differences in the timing of the secondary tasks.

The third alternative considered was that the secondary tasks used in experiment 1 might not be process pure, and so differences in interference might be due to some subprocess(es) beyond the assumed executive requirements of each secondary task. This approach to an explanation does not hold given the differences between the secondary tasks employed in experiments 1 and 2. That is, while one might argue that the secondary tasks used in experiment 2 are also not process pure, the differences between the secondary tasks in the two experiments undermines any explanation of the effects due to specific (unintended) characteristics of the secondary tasks. The constant across the two sets of secondary tasks is the executive functions held to be involved.

The fourth alternative explanation was that executive processes may not be fractionable across tasks during dual task performance, and in particular that if response inhibition works to inhibit the production of a response globally then concurrent performance of two response inhibition tasks might facilitate performance. Again (as in experiment 1), this is possible but it does not account for the lack of interaction between problem and condition in the item analysis, or for poorer performance in the set-shifting condition than in the memory condition.

We are therefore left with the hypothesis with which we began, namely that set-shifting is a critical executive function for ToL performance, and that the impaired performance when completing ToL problems while also completing the plus/minus task is attributable to the shared set-shifting requirements of both tasks.

4 General Discussion

Two dual-task experiments have shown that performance on the Tower of London task is affected more strongly when the task is paired with a secondary task that requires the putative executive function of set-shifting than a secondary task that requires the putative executive functions of either memory monitoring/updating or response inhibition. To be clear, concurrent performance of a memory monitoring/updating task does impair ToL performance (as one would expect given that planning requires working memory), but less so than concurrent performance of a set-shifting task. This is consistent with views in which online planning or problem-solving requires shifting between different strategies or different subgoals (e.g., Bull et al., 2004), but possibly surprising given that one might assume that both memory monitoring/updating and response inhibition might play a substantial role in planning tasks (e.g., Welsh et al., 1999; Miyake et al., 2000).

What might be the possible role of set shifting (as opposed to memory updating and/or response inhibition) in solving ToL problems? Consider problem A of experiment 1 (see Figure 5 in appendix 1). Initially, none of the three balls are in their desired positions. To solve the problem one might first ask: Which ball should be moved first (cf. Gilhooly et al., 1999)? Thus, one immediately considers three possible subgoals (move red to target position, move green to target position, or move blue to target position). In order to solve the problem in as few moves as possible one must flexibly shift between these three subgoals to evaluate which is most easily or directly achieved (rather than focussing on, for example, the red ball first without considering the other balls). In this way, successful performance on the ToL requires one to switch between evaluating different subgoals. More generally, successful problem solving often requires exploring multiple possible solutions (or even multiple possible solution strategies) and hence shifting between possible solutions or strategies.

Consistent with this view, in a related study using the same methodology but a different primary task (the Wisconsin Card Sorting Task – WCST – instead of the ToL), Cooper et al.

(2012) found that simultaneous performance of a set-shifting task (the digit-switching task) led to substantial interference on several WCST measures. Most notably, participants took significantly more trials to attain the first category when the WCST was paired with a set-shifting task than when it was paired with either a memory monitoring/updating task (the 2-back task) or a response inhibition task (the go/no-go task). While the WCST is commonly used to assess executive dysfunction, it is generally regarded as a test of cognitive flexibility rather than a test of planning or problem-solving. Good performance requires that subjects switch between different stimulus-response mappings in response to negative feedback. Studies of WCST performance frequently fail to report the number of trials needed to attain the first category – the primary dependent variable of interest is typically the number of perseverative errors after the first category has been attained – but in those that do it is generally elevated in patients with frontal brain damage (see Heaton, 1981, for a review). Given the logic employed in this study, the similarity in interference pattern (of set-shifting on ToL and WCST) raises the possibility of a common subprocess shared by the ToL, rule attainment within WCST and simple set-shifting tasks. One possibility is that both the ToL and initial performance on the WCST require exploration of different hypotheses and, consequently, shifting between alternative hypotheses. This shifting between hypotheses, we suggest, is impaired by concurrent performance of a secondary shifting task.

The role of set-shifting in ToL as argued here also supports a novel explanation for the role of left frontal regions in ToL performance (cf. Shallice, 1982; Glosser & Goodglass, 1990; Morris et al., 1997). Left DLPFC has frequently been associated with the executive function of set-shifting (see Shallice et al., 2008, for a review). We have argued that this function is critical for successful solution of ToL problems. This argument, together with the brain-based studies of set-shifting, is entirely consistent with the apparent involvement of left prefrontal cortex in the solution of ToL problems, but two other related hypotheses concerning the involvement of left PFC cannot be ruled out. First, imaging studies of reasoning (e.g., Christoff et al., 2001; Green et al., 2006) have argued that PFC, and in particular left PFC, is involved in the manipulation of relational information. Christoff et al. (2001) explicitly relate this to “self-generated information”. Successful solution of ToL problems clearly requires the generation and manipulation of such information, but this hypothesis says nothing about the nature of the informational manipulation. Second, left lateral PFC has been argued to be critical also to the generation of multiple hypotheses (Reverberi et al., 2005). Hypothesis generation is therefore another possible explanation for the involvement of left lateral PFC in ToL performance, as successful solution typically requires generating multiple hypotheses concerning moves required to reach the solution. Note however that the generation of multiple hypotheses requires that one must switch (i.e., engage and disengage) between those hypotheses. Switching thus seems a more primitive operation than hypothesis generation.

A second issue of potential concern relates to possible differences between planning and problem-solving. As in the study of Phillips et al. (1999; see also Phillips et al., 2001), our participants were not told the minimum number of moves needed to solve each problem. While this may reduce the requirement for planning prior to the first move, it does not eliminate on-line planning – participants generally do not move balls at random but must select moves that somehow transform the current state towards the final goal state. Consistent with Gilhooly et al. (1999), we suggest that participants focus on different subgoals (e.g., first get the red ball in place) during the course of each trial, and plan for these subgoals, both prior to their first move and during the course of the trial.

Consider finally a third substantive issue. Both experiments failed to find an interaction between problem and secondary task on any of the four dependent variables. That is, all secondary tasks led to similar levels of interference on all problems. This supports the conclusion that while problems may vary in difficulty, the differences in the problems considered here do not reflect a critical dimension when it comes to differentially engaging specific executive functions. Kaller et al. (2011b) came to a similar conclusion in motivating their imaging study of performance on a variant of the ToL. They argued that the critical dimensions were not the number of moves required or the presence of superficially tempting moves per se, but subgoal order ambiguity and the need for intermediate moves. (See also Newman & Pittman, 2007.) As discussed briefly in the introduction, Klahr and Robinson (1981) noted with the Tower of Hanoi (ToH) that different goal states impose different constraints on the order in which subgoals must be achieved. So-called “tower” states, where all disks are on a single peg, can only be produced by first achieving the subgoal of positioning the lowest disk, then the subgoal of positioning the middle disk, and finally the subgoal of positioning the top-most disk. The order in which the three subgoals must be achieved is therefore unambiguous. This is not the case for “flat” target states, where one disk is on each peg, and where the order in which subgoals must be achieved so as to minimise total moves is ambiguous. The tower/flat distinction applies equally to the ToL as it does to the ToH (Waldau, 1999; Kaller et al., 2008; Kaller et al., 2011a), and as noted in the introduction, Kaller et al. (2011b) found that problems with ambiguous subgoal ordering engage left lateral PFC relatively more than those without (while problems that require intermediate moves activated right lateral PFC more than those that did not). Ambiguous subgoal ordering again implies the need for a switching function: successful solution of such problems will require considering different orderings of the subgoals, and hence switching between subgoals when planning moves. If this is the case, then we should expect greater interference from a switching secondary task not on problems where superficially tempting moves should be resisted, but on ambiguous subgoal problems. It is not possible to assess this in the current study, as all problems had partially or totally ambiguous subgoal ordering. More generally, however, this perspective may explain why different studies have found different executive requirements of the ToL and the ToH – the executive requirements appear to be a function of the specific problems considered, rather than intrinsic properties of the specific apparatus.

The inability to assess whether a switching secondary task produces greater interference on problems with ambiguous subgoal order than on problems with superficially tempting moves is one limitation of the current study. Another concerns the specific dual-task methodology employed. As noted, comparison of performance on the baseline condition with performance on the dual-task conditions was not interpretable because of a potentially confounding order effect. These limitations may be addressed through the use of alternative problems (e.g., similar to Kaller et al., 2011b) and alternative experimental designs (e.g., by reassessing baseline performance multiple times throughout the experiment, or by measuring performance on both primary and secondary tasks and calculating dual task decrements for each as described by, for example, Miyake et al., 2000).

Regardless of these limitations, one way in which the role of shifting in the performance of the ToL (or indeed the ToH) might be specified more precisely is through the development of a fully functioning computational model of the task and how it is performed given different initial and goal states. While such models have been developed (e.g., for the ToL, Baughman & Cooper, 2007; Dehaene & Changeux, 1997; Newman et al., 2003), and while the model of Baughman and Cooper (2007) explicitly incorporates a mechanisms of response inhibition, existing models fail to consider the switching requirements of the task. At the same time, while the various executive functions might be viewed as atomic building blocks out of which complex task performance is constructed, the set of executive functions does not necessarily uniquely identify an algorithmic or processing level account of task behaviour, nor does it lead to a unique characterisation of task performance. Individual differences in strategy will presumably require different combinations of executive functions to support those strategies. The results presented here, however, suggest that the majority of participants, when attempting the Tower of London task, adopt an approach that involves switching focus between different strategies or subgoals.

Acknowledgements

We are grateful to Eddy J. Davelaar and Karolina Wutke for assistance with some aspects of the design of the study reported here, and to Karolina Wutke and Toni Saint for assistance with the data collection. We are also grateful to Christoph P. Kaller and an anonymous reviewer for highly constructive feedback on an earlier version of this manuscript. This work was partly funded by a Faculty Research Grant from the Faculty of Science, Birkbeck, University of London, to Richard P. Cooper. All experimental work was approved by the departmental ethics committee prior to initiation and carried out in accordance with the declaration of Helsinki. Verity Marsh is now at Queen Elizabeth Hospital, Lewisham and Greenwich NHS Trust. Please address correspondence to Richard P. Cooper (R.Cooper@bbk.ac.uk).

References

Albert, D., & Steinberg, L. (2011). Age differences in strategic planning as indexed by the Tower of London. *Child Development*, *82*, 1501-1517.

Allport, A. D., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.), *Attention and Performance 15: Conscious and Nonconscious Information Processing* (pp. 421–452). Cambridge, MA: The MIT Press.

Asato, M. R., Sweeney, J. A., & Luna, B. (2006). Cognitive processes in the development of TOL performance. *Neuropsychologia*, *44*(12), 2259-2269.

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *49*(1), 5-28.

Baughman, F. D. and Cooper, R. P. (2007): Inhibition and young children’s performance on the Tower of London task. *Cognitive Systems Research*, *8*, 216-226.

Berg, W. K., & Byrd, D. L. (2002). The Tower of London spatial problem-solving task: Enhancing clinical and research implementation. *Journal of Clinical and Experimental Neuropsychology*, *24*(5), 586-604.

Bull, R., Espy, K. A., & Senn, T. E. (2004). A comparison of performance on the Towers of London and Hanoi in young children. *Journal of Child Psychology and Psychiatry, 45*, 743-754.

Cheetham, J. M., Rahm, B., Kaller, C. P., & Unterrainer, J. M. (2012). Visuospatial over verbal demands in predicting Tower of London planning tasks. *British Journal of Psychology*, *103*(1), 98-116.

Christoff, K., Pranhakaran, V., Dorfman, J., Shao, Z., Kroger, J.K., Holyoak, K.J., & Gabrieli, J.D.E. (2001). Rostrolateral prefrontal cortex involvement in relation integration during reasoning. *Neuroimage*, *14*, 1136-1149.

Cockburn, J. (1995). Performance on the Tower of London test after severe head injury. *Journal of the International Neuropsychological Society*, *1*(06), 537–544.

Cooper, R.P., Wutke, K., & Davelaar, E. (2012). Differential contributions of set-shifting and monitoring to dual-task interference. *Quarterly Journal of Experimental Psychology, 65*(3), 587-612.

Coxon, J. P., Stinear, C. M., & Byblow, W. D. (2009). Stop and go: The neural basis of selective movement prevention. *Journal of Cognitive Neuroscience*, *21*(6), 1193-1203.

Dehaene, S. & Changeux, J.-P. (1997). A hierarchical neural network for planning behaviour. *Proceedings of the National Academy of Sciences, USA, 94,* 13293-13298.

Donders, J. & Larsen, T. (2012). Clinical utility of the Tower of London – Drexel University, Second Edition (TOL^{DX}) after adolescent traumatic brain injury. *Developmental Neuropsychology*, *37*, 333-342.

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 & reasoning*, *5*(4), 339-361.

Glosser, G., & Goodglass, H. (1990). Disorders in executive control functions among aphasic and other brain-damaged patients. *Journal of Clinical and Experimental Neuropsychology*, *12*(4), 485-501.

Green, A.E., Fugelsand, J.A., Kraemer, D.J.M., Shamosh, N.A., & Dunbar, K.N. (2006). Frontopolar cortex mediates abstract integration in analogy. *Brain Research*, *1096*, 125-137.

Hartman, E., Houwen, S., Scherder, E., & Visscher, C. (2010). On the relationship between motor performance and executive functioning in children with intellectual disabilities. *Journal of Intellectual Disability Research*, *54*, 468-477.

Heaton, R. K. (1981). *A Manual for the Wisconsin Card Sorting Test*. Odessa FL: Psychological Assessment Resources.

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology,* Whole No. 89.

Kaller, C. P., Rahm, B., Spreer, J., Mader, I., & Unterrainer, J. M. (2008). Thinking around the corner: The development of planning abilities. *Brain and Cognition*, *67*(3), 360-370.

Kaller, C. P., Rahm, B., Köstering, L., & Unterrainer, J. M. (2011a). Reviewing the impact of problem structure on planning: a software tool for analyzing tower tasks. *Behavioural brain research*, *216*(1), 1-8.

Kaller, C. P., Rahm, B., Spreer, J., Weiller, C., & Unterrainer, J. M. (2011b). Dissociable contributions of left and right dorsolateral prefrontal cortex in planning. *Cerebral Cortex*, *21*(2), 307–317.

Klahr, D., & Robinson, M. (1981). Formal assessment of problem-solving and planning processes in preshool children. *Cognitive Psychology*, *13*, 113-148.

Köstering, L., McKinlay, A., Stahl, C., & Kaller, C. P. (2012). Differential patterns of planning impairments in Parkinson's disease and sub-clinical signs of dementia? A latent-class model-based approach. *PloS one*, *7*(6), e38855.

Logan, G. (1994). On the ability to inhibit thought and action. In Dagenbach, D. & Carr, T.H., (Eds.) *Inhibitory Processes in Attention, Memory and Language*, pp. 189–239, Academic Press.

Luciana, M., Collins, P. F., Olson, E. A., & Schissel, A. M. (2009). Tower of London performance in healthy adolescents: The development of planning skills and associations with self-reported inattention and impulsivity. *Developmental Neuropsychology*, *34*(4), 461-475.

Marzocchi, G. M., Oosterlaan, J., Zuddas, A., Cavolina, P., Geurts, H., Redigolo, D., ... & Sergeant, J. A. (2008). Contrasting deficits on executive functions between ADHD and reading disabled children. *Journal of Child Psychology and Psychiatry, 49*(5), 543-552.

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1423-1442.

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A. & Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134-140.

Morris, R. G., Miotto, E. C., Feigenbaum, J. D., Bullock, P., & Polkey, C. E. (1997). The effect of goal-subgoal conflict on planning ability after frontal- and temporal-lobe lesions in humans. *Neuropsychologia, 35,* 1147-1157.

Newman, S. D., Carpenter, P. A., Varma, S., & Just, M. A. (2003). Frontal and parietal participation in problem solving in the Tower of London: fMRI and computational modeling of planning and high-level perception. *Neuropsychologia*, *41*(12), 1668-1682.

Newman, S. D., & Pittman, G. (2007). The Tower of London: A study of the effect of problem structure on planning. *Journal of Clinical and Experimental Neuropsychology*, *29*(3), 333-342.

O’Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18*(2), 283–328.

Phillips, L. H., Wynn, V., Gilhooly, K. J., Della Sala, S., & Logie, R. H. (1999). The role of memory in the Tower of London task. *Memory*, *7*(2), 209–231.

Phillips, L. H., Wynn, V. E., McPherson, S., & Gilhooly, K. J. (2001). Mental planning and the Tower of London task. *The Quarterly Journal of Experimental Psychology Section A*, *54*(2), 579–597.

Rainville, C., Lepage, E., Gauthier, W., Kergoal, M-J., & Belleville, S. (2012). Executive function deficits in persons with mild cognitive impairment: A study with a Tower of London task. *Journal of Clinical and Experimental Neuropsychology, 34,* 306-324.

Raye, C., Johnson, M., Mitchell, K., Greene, E., & Johnson, M. (2007). Refreshing: A minimal executive function. *Cortex*, *43*(1), 135-145.

Reverberi, C., D’Agostini, S., Skrap, M., & Shallice, T. (2005). Generation and recognition of abstract rules in different frontal lobe subgroups. *Neuropsychologia*, *43*(13), 1924-1937.

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

Shallice, T., Stuss, D.T., Picton, T.W., Alexander, M.P., & Gillingham, S. (2008). Mapping task switching in frontal cortex through neuropsychological group studies. *Frontiers in Neuroscience*, *2*, 79-85.

Sullivan, J. R., Riccio, C. A., & Castillo, C. L. (2009). Concurrent validity of the tower tasks as measures of executive function in adults: A meta-analysis. *Applied Neuropsychology*, *16*, 62-75.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, *7*(2), 268-288.

Towse, J. N. & Neil, D. (1998). Analyzing human random generation behaviour: A review of methods used and a computer program for describing performance. *Behaviour Research, Instruments and Computers*, *30*(4), 583–591.

Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task switching: Interplay of reconfiguration and interference control. *Psychological Bulletin, 136(4),* 601-626.

Waldau, R. (1999). A developmental study of problem solving in children aged 3–6: Development of planning strategies. *Unpublished BSc Thesis*, University of London.

Welsh, M. C., Satterlee-Cartmell, T., & Stine, M. (1999). Towers of Hanoi and London: Contribution of Working Memory and Inhibition to Performance. *Brain and Cognition*, *41*(2), 231–242.

Wilson, M. (1988). MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, *20*(1), 6-10.

Zook, N. A., Davalos, D. B., DeLosh, E. L., & Davis, H. P. (2004). Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks. *Brain and cognition*, *56*(3), 286-292.

Appendix A: ToL Problems Used in Experiments 1 and 2

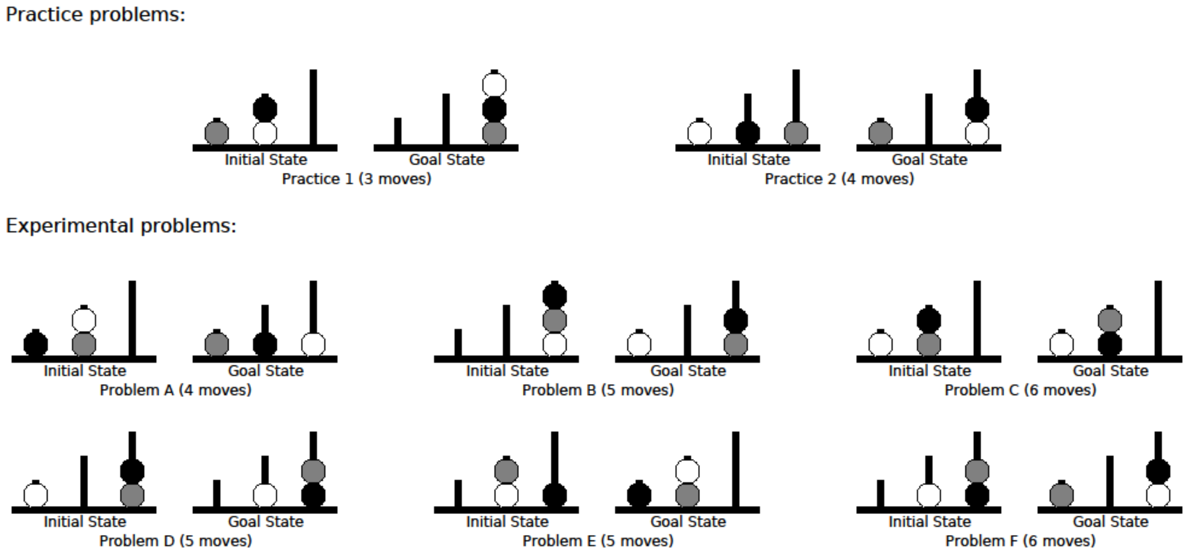

Fig. 5: The 2 practice and 6 experimental Tower of London problems used in experiment 1, using greyscale for ball colour. The experimental problems differ in numerous respects. Problem A has a fully ambiguous goal state, while problems B to F have partially ambiguous goal states. For the first move, there are two options for problems A to D, but three options for problem E and four options for problem F. Problem C requires moving the white ball away from its initial position, even though that initial position is also the goal position. Problems E and F require avoiding the temptation, on the initial move, to place a ball in its goal position. (See Kaller et al., 2011a, for public domain software that analyses Tower problems on these and more dimensions.) The mapping of shades of grey to colours of the balls as presented to participants (red, green and blue) was randomised on each block so that participants did not become familiar with the problems.

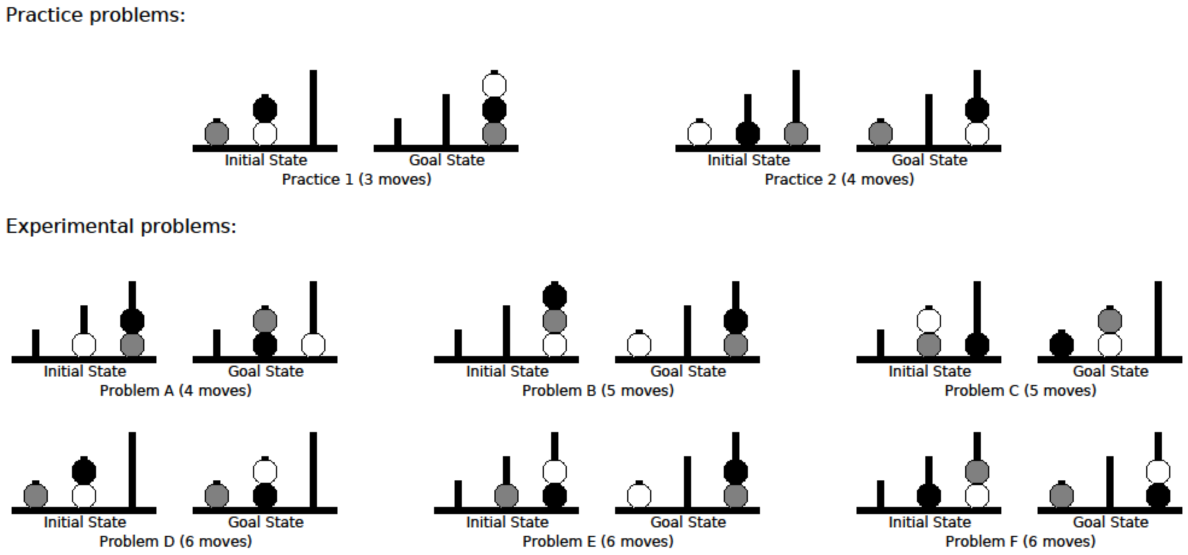

Fig. 6: The 2 practice and 6 experimental Tower of London problems used in experiment 2, using greyscale for ball colour. All problems have partially ambiguous goal states, but the problems differ in various other respects, including the number of options available for the initial move and whether one of those options involves moving a ball to its goal position. As in experiment 1, mapping of shades of grey to colours of the balls as presented to participants (red, green and blue) was randomised on each block so that participants did not become familiar with the problems.