

Interruption of the Tower of London Task: Support for a Goal-Activation Approach

Helen M. Hodgetts
Cardiff University

Dylan M. Jones
Cardiff University and University of Western Australia

Unexpected interruptions introduced during the execution phase of simple Tower of London problems incurred a time cost when the interrupted goal was retrieved, and this cost was exacerbated the longer the goal was suspended. Furthermore, time taken to retrieve goals was greater following a more complex interruption, indicating that processing limitations may be as important as time-based limitations in determining the ease of goal retrieval. Such findings cannot simply be attributed to task-switching costs and are evaluated in relation to current models of goal memory (E. M. Altmann & G. J. Trafton, 2002; J. R. Anderson & S. Douglass, 2001), which provide a useful basis for the investigation and interpretation of interruption effects.

Keywords: interruption, Tower of London, goals, activation, memory

Psychologists' understanding of goal-directed behavior has undoubtedly progressed recently through research on task switching (e.g., Allport & Wylie, 2000; Monsell, 2003) and prospective memory (e.g., Brandimonte, Einstein, & McDaniel, 1996; Einstein et al., 2005). However, another phenomenon that seems to offer an invaluable tool with which to examine both the formulation and execution of goals is task interruption, a common, everyday occurrence that involves both the unexpected suspension, and subsequent resumption, of task goals. Despite its long history (Zeigarnik, 1927), laboratory research on interruptions is sparse, and theory specifically aimed at accounting for interruption behavior is underdeveloped. The present study constitutes one of the first attempts to delineate basic interruption phenomena on which theoretical development can be based. Furthermore, given that task interruption is commonplace in many modern work environments, the results of the research may have some immediate practical implications.

That relatively brief interruption of a task has an impact on the performance of that task is not in doubt. This has been illustrated many times in everyday nonlaboratory settings, with the loss of efficiency being reported in a variety of ways: as increased task performance times (e.g., Paquiot, Eyrolle, & Cellier, 1986), de-

creased accuracy (e.g., Bainbridge, 1984; Flynn et al., 1999), greater stress (Zijlstra, Roe, Leonora, & Krediet, 1999), and self-appraisals of increased workload (Kirmeyer, 1988). Moreover, interruptions are also cited as a significant factor in reported accidents and incidents in work environments associated with high risk, such as nuclear power plants (Griffon-Fouco & Ghertman, 1984) and flight decks (McFarlane & Latorella, 2002). Given the prominence and incidence of interruption-related impairments in efficiency—as well as the theoretical implications of research on such impairments—it is perhaps surprising that laboratory research into interruptions is modest in volume and has reached little consensus. Part of the reason for the lack of consensus may be the disparate methodologies used, but the lack of a specific theoretical framework on which to base predictions may also be partly responsible. In the current study, we assessed how the cognitive system copes with task interruption by drawing from the literature on goal memory and goal retrieval within the well-established Adaptive Control of Thought–Rational (ACT-R) architecture (Anderson & Lebiere, 1998). Using this framework, we addressed two key phenomena highlighted in the literature—interruption length and complexity—and examined how current theories of goal memory may account for the effects of each.

Memory for Goals

Goals are often thought to have some specialized status, being credited with a higher level of activation than that of nongoal items (e.g., Goschke & Kuhl, 1996), a capacity to maintain this level of activation without rehearsal (Anderson, 1993), and a need for active inhibitory processes to remove them from working memory when no longer required (Mayr & Keele, 2000). Such a specialized view of goal memory is exemplified by ACT-R's goal stack, in which goals are stored immune from decay, are retrieved without interference from other goals, and are removed from working memory via a "popping" mechanism so that they can no longer intrude. Although Anderson and colleagues did not specifically emphasize this component of ACT-R, relying on a goal-stack structure to model goal memory may mean that the more specific

Helen M. Hodgetts, School of Psychology, Cardiff University, Cardiff, United Kingdom; Dylan M. Jones, School of Psychology, Cardiff University, and Department of Psychology, University of Western Australia, Perth, Western Australia, Australia.

The research reported here was supported by a postdoctoral fellowship from the United Kingdom's Economic and Social Research Council awarded to Helen M. Hodgetts.

We thank Robert Houghton for writing the initial software for the experiments and Andrew Howes and Robert Hughes for their useful comments on drafts of this article.

Correspondence concerning this article should be addressed to Helen M. Hodgetts or Dylan M. Jones, School of Psychology, Cardiff University, Cardiff CF10 3AT, United Kingdom. E-mail: hodgettshm@cardiff.ac.uk or jonesdm@cardiff.ac.uk

processes involved in the suspension and resumption of goals are overlooked. In the current study, we used the systematic assessment of interruption phenomena to explore in more detail the limitations of goal memory.

The goal stack is perhaps the classic embodiment of goal memory (e.g., Anderson & Lebiere, 1998; Miller, Galanter, & Pribram, 1960); goals are held in a specialized state and are dealt with in a last-in, first-out manner. A new or interrupting goal would be “pushed down” on the top of the stack to become the focus of behavior, whereas upon its completion it would be “popped off,” leaving the previous activity reliably and instantly available by default, once again at the top of the stack. This representation provides a convenient parallel to the processes of goal decomposition in means–ends analysis whereby goals are hierarchically broken down into smaller, more manageable subgoals, with each, upon completion, pointing back to a preceding parent goal. Experience tells us, however, that goal memory is rarely this precise: Action slips occur if the correct goal is not retrieved at the correct time (Reason, 1990), or goals may be forgotten altogether in the absence of a cue appropriate to prime retrieval (Byrne & Bovair, 1997). Although it does not challenge other aspects of the ACT-R architecture, recent research contests ACT-R’s assumption of a goal stack and instead suggests that goal memory may be subject to the same processes that govern declarative memory. We turn now to a consideration of current theories of goal memory (Altmann & Trafton, 1999, 2002; Anderson & Douglass, 2001), which we use as a theoretical basis for the study of task interruption. Although the primary focus of these models is the simulation of behavior in the Tower of Hanoi task, they lend themselves easily to the study of interruption because some of the key elements of the task when performed continuously—the suspension and resumption of goals—are also a primary consequence of interruption.

Evidence that goal memory suffers decay in a way similar to declarative memory traces comes from a study by Anderson and Douglass (2001). Both formulation and storage of goals were found to influence move latencies in an analysis of 15-move Tower of Hanoi problems, but goal retrieval was found to be the most critical factor. Participants’ actions were slower at points during the task when they were to retrieve goals, and latencies were slower still the longer ago that the goal was formulated. Similarly, error data showed that those goals formed longer ago were also more likely forgotten. In declarative memory, the likelihood of a given item being retrieved on a particular cycle will depend on its activation or relevance at that time as determined by two components: history of use (base-level activation) and context (associative activation; Anderson & Schooler, 1991). Using the central principle of base-level activation, and not the goal stack, Anderson and Douglass (2001) developed an ACT-R model that was able to provide an accurate simulation of the experimental data. Because goal memory appeared to show a retention function similar to that of ordinary memory items, Anderson and Douglass (2001) concluded that there is no need for a separate goal stack, either perfect or fallible.

Altmann and Trafton (1999, 2002) also eschewed the idea of the goal stack and argued that the retrieval and maintenance of goals are no different than the retrieval and maintenance of other declarative memory items. At any given time, the most active goal will be the one to govern behavior. Total activation is determined

first by the duration of a goal’s lifetime and the number of times it has been retrieved in that lifetime (base-level activation), but this can also be boosted by spreading activation from environmental cues (associative activation). These two principles are implemented in the model through the constraints of strengthening and priming. A new or interrupting goal must be repeatedly sampled or strengthened within a short period of time in order to build up its activation above that of other competing goals and to overcome proactive interference. For an old goal to be retrieved, it must be primed by associative cues in either the mental or physical environment so as to overcome retroactive interference (see Figure 1). The Altmann and Trafton model makes the prediction that preparation before the onset of an interruption can aid subsequent task resumption, as associative priming clearly depends on efficient associative encoding. Boosting activation of the to-be-suspended goal and encoding cues prior to the interruption will facilitate later retrieval. In line with the goal stack view, it is often true that the most active goal will also be the most recent, but the goal-activation model is also flexible in that goals can be suspended, reinstated, or interleaved.

Both models outlined above adopt ACT-R’s base-level learning equation proposing that retrieval of a goal, like retrieval of other memory elements, is dependent on the time course of its activation and decay in memory. Altmann and Trafton (1999, 2002) additionally incorporated the influence of associative activation in goal retrieval and proposed that preparation before the onset of interruption can mitigate the effects of base-level activation decay. We consider now how these models of goal memory may speak to the issues of interruption length and complexity, which as yet remain equivocal in the interruption literature.

Length of Interruption

On the basis of the assumptions of the ACT-R architecture, the empirical expectation is that the duration of the interruption should affect the efficiency with which a task is resumed. ACT-R incorporates a base-level learning equation according to which the activation of a declarative memory item increases with repeated sampling but decays as a power function of delay. Both the model of Anderson and Douglass (2001) and that of Altmann and Trafton (1999, 2002) apply this principle to memory for goals and predict

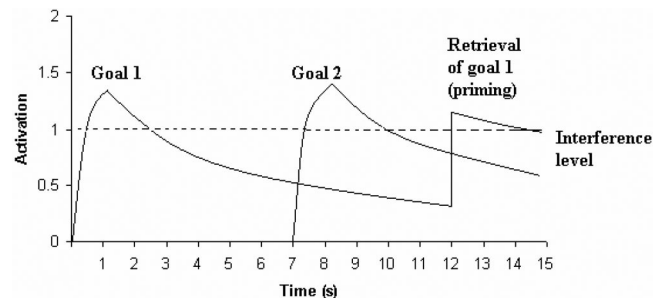


Figure 1. Priming to increase activation of an old goal above that of other competing goals at the interference level. Adapted from Figure 1 (p. 46) and Figure 2 (p. 48) in “Memory for Goals: An Activation-Based Model” by E. M. Altmann and G. J. Trafton, 2002, *Cognitive Science*, 26, 39–83. Copyright 2002 by the Cognitive Science Society. Adapted with permission.

activation level to be inversely related to the duration of the interruption: The more a goal has decayed, the longer it will take to reinstate and the more prone it will be to error. Such a prediction is also supported by classic findings in short-term memory research (e.g., Brown, 1958; Peterson & Peterson, 1959), according to which items retained over a longer interval are more likely to be forgotten. However, research into prospective memory suggests that there is no effect of the length of a retention interval over which an intention must be maintained, whether it is on the order of minutes (e.g., Guynn, McDaniel, & Einstein, 1998) or seconds (e.g., Einstein, McDaniel, Williford, Pagan, & Dismukes, 2003).

Within the interruption literature, the empirical evidence concerning interruption length is similarly inconclusive. An interruption involving the completion of mental arithmetic problems had no effect on performance of a computer-based shopping task irrespective of whether the interruption lasted 30 s or 2.75 min (Gillie & Broadbent, 1989). This finding led Gillie and Broadbent to conclude that the length of the interruption alone could not be an important factor in determining disruption. Similarly, no effect of interruption length was found in a study of various tasks, such as adding and reading comprehension, that included instant-message-style pop-up interruptions (Bailey, Konstan, & Carlis, 2000). In contrast, using a text editing task, Lahlou, Kirsh, Rebotier, Reeves, and Remy (2000) reported that 2-min-long interruptions were more disruptive than those lasting 30 s. Of course, the effects of interruption length may also be a function of the type of task. Given the inconsistent nature of the evidence and the unambiguous prediction of decay from the ACT-R framework, the issue of interruption duration merits further experimental scrutiny.

As a further point, we also sought to investigate the prediction made by the goal-activation model that participants engage in a process of rehearsal to further consolidate a to-be-suspended goal. An earlier version of the model (Altmann & Trafton, 1999) suggested that this rehearsal may be proportional to the length of the anticipated retention interval, and as such this model is of interest in the current work on interruption duration. It proposed that participants strengthen more those goals that will be suspended for longer, so as to build up base-level activation and consequently offset the imminent effects of decay. Although Altmann and Trafton (2002) did not include this feature in the final version of their model (the underlying mechanism only accounted for a small amount of variance in the latency data), it is still perhaps an interesting point to pursue. For example, it seems logical that one may invest more time in consolidating the place reached in a task if one anticipates that the interruption may take some time (e.g., answering a telephone call rather than just responding to a colleague's casual question). Conversely, the idea of differential strengthening of goals was disputed by Anderson and Douglass (2001), who believed that participants do not engage in any strategic rehearsal process before goal suspension and instead choose to pay the price of forgetting at retrieval.

Complexity of the Interruption

Short-term memory research indicates that decay over time is certainly not the only factor to influence memory loss (see Nairne, 2002). The type of intervening activity also plays an important role, with the rate of forgetting increasing with a more complex task of equivalent duration (Kroll & Kellicutt, 1972; Nakajima &

Sato, 1989; Posner & Konick, 1966). Of course, the term *complexity* can be difficult to define adequately, but it could be considered in terms of the number of different components involved in the task and the skill required to coordinate them (Wood, 1986). A complex intervening task may increase interference with the suspended goal because of the number of these subelements involved, or it may act by suppressing rehearsal of the to-be-remembered items during the retention interval.

Anderson and Douglass (2001) would not have predicted an effect of interruption complexity on either of these accounts. They did not believe that participants rehearse suspended goals, and so a situation in which rehearsal is prevented would therefore make no difference to goal maintenance. Also, their model is based solely on a decrease in activation over time. It incorporates no notion of interference from other goals, and so the number of active subgoals would have no bearing on the ease of goal retrieval. Altmann and Trafton's (2002) model, however, takes into account the influence of other goals and not just the activation of the target. This model predicts that an increased number of goals will raise the interference level; as such, a suspended goal will require a greater degree of strengthening to bring it above the interference threshold, thus making it more time-consuming to reinstate.

Interruption complexity was found to be critical within the context of a computer-based shopping task (Gillie & Broadbent, 1989). There was no disruption when the interrupting task involved solving simple mental arithmetic problems, but when these problems were exacerbated by the requirement to first decode letters into digits, a marked effect on primary task performance was observed. In a study of text comprehension, the time taken to read the first sentence after an interruption was greater following both longer intervals (30 s compared with 10 s) and also following a more demanding interruption task (e.g., recalling digits rather than completing simple addition problems; Fischer & Glanzer, 1986). On the other hand, interruption complexity appeared to have no effect during a text editing task in which participants were interrupted for a set interval with either cognitive interruptions (e.g., a logical reasoning question) or social interruptions (e.g., idle conversation; Lahlou et al., 2000). Intuitively, the cognitive interruptions seem more demanding, but they did not increase disruption of the primary task relative to social interruptions. In the related domain of task switching, there seems to be some evidence for a greater cost in switching to a more complex task (Garcia-Ogueta, 1993; Rubinstein, Meyer, & Evans, 2001). However, using a task-switching span procedure, one study found that memory for pending tasks was not compromised by concurrent processing demands (Logan, 2004).

In sum, the current state of the literature is inchoate, with little apparent consensus between studies and only rudimentary theoretical progress. The present series of experiments can therefore be regarded as exploratory, assessing the effects of interruption length and complexity within a specified framework. Despite its exploratory character, the present study approached these factors at a fine-grained level of analysis in order to examine the effects of interruptions at the level of goal activation. The task chosen for this study was the five-disk Tower of London task (Ward & Allport, 1997), which has a number of characteristics that make it well suited for the study of goal-directed behavior generally and the interruption of performance in particular.

Tower of London

Based on the well-known Tower of Hanoi (ToH) problem, the Tower of London (ToL) task was originally developed for use with neuropsychological patients (Shallice, 1982), but it has since been adapted to assess executive functioning in nonclinical populations. In a computerized version of the five-disk ToL task (Ward & Allport, 1997), participants are presented with a starting array of different-colored, equal-sized disks arranged on three pegs, with a goal-state configuration displayed at the top of the screen (see Figure 2). Clicking consecutively on two pegs on the main display will move the top disk on the first selected peg across to its new position on the second peg. The aim is to rearrange the disks on the main display in this manner, one at a time from peg to peg, until they match exactly the configuration shown in the goal-state box. Unlike in the ToH problem, the movement of disks is not governed by size constraints. This makes it the more straightforward of the two problems in terms of task rules and strategies, and we hoped therefore that participant performance might be more uniform. The task involves the formulation, retention, and execution of a planned series of actions, and interruption at any of these stages is likely to disrupt the efficiency with which the goal state is achieved. In the current study, we examined the effect of interruption of the execution phase: What factors affect retrieval of the next move in a planned sequence when execution of this sequence is unexpectedly broken?

In the current series of experiments, we used only simple six-move ToL problems to establish basic effects of interruption. All the critical trials followed the same structure such that the first and the fourth moves were the beginning of two-move sequences to place disks in their target locations, whereas the remaining moves all placed disks directly in their goal positions (see Appendix A for the ToL trials used). Interruption occurred in the middle of a trial before the fourth move, so that task resumption always required an initial indirect move (i.e., one that did not place a disk directly in its target location). This was true both when the problem was solved in the minimum number of moves and when more moves were taken.

The ToL method used in the current experiment included an additional constraint: No more than three disks could be held on one peg at one time. A variation of the ToL task such as this one has not been used before, but we hoped that an additional restriction on moves in what is a relatively unconstrained task would further emphasize the need for thorough planning at the start of a trial. More important, this constraint restricted moves such that for the critical six-move problems, only one optimal solution path was

possible. A direct comparison was then possible for perfect trials, as latencies would reflect exactly the same move in each case.

The interrupting task involved completing a mood checklist that consisted of a list of six descriptions along a mood continuum (e.g., “extremely happy,” “fairly happy,” “slightly happy,” “slightly sad,” “fairly sad,” and “extremely sad”). Participants were to read the list and click with the mouse on the description that most applied to their mood at that point. This provided a brief, undemanding interruption that would be unlikely to have a high processing demand. We began by examining whether interruption of the execution phase of the ToL task would incur costs to the primary task in terms of resumption time and error and whether these costs would be exacerbated by a longer interruption interval.

Experiment 1

Participants were interrupted on 8 of 25 trials. We predicted that the time taken to make the next move in the sequence following an interruption would be greater than the time taken to make the equivalent fourth move in a control trial without interruption. Also, we expected that participants would more frequently complete the problems in the minimum number of moves on those trials that were uninterrupted. Furthermore, on the basis of the decay function incorporated into the models of Altmann and Trafton (1999, 2002) and Anderson and Douglass (2001), we expected that increasing the length of an interruption would decrease activation and hence render the reinstatement of the suspended goal more time-consuming and prone to error. We tested this expectation by using interruptions of two lengths, 6 s and 18 s in duration. Six seconds was chosen on the basis of pilot work as a suitable time for completion of one checklist. The 18-s interruption consisted of three 6-s checklists that were displayed in succession. Length and load are potentially confounded here, but arguably load increase is not appreciable.¹ An interruption of three times the length was judged to be different enough to produce effects of length, should there be any, but not too long for the goal to decay completely.

In addition, we tested one of the predictions that flows from Altmann and Trafton’s (1999) earlier model, namely, that goals are rehearsed in proportion to the length of time for which they will be suspended. On half of the interruption trials, participants were told the duration of the interruption (one checklist or three checklists) at its onset, and on the other half of the trials, this information was not specified. At the onset of the interruption, the main display—although still in view—was frozen, and a button labeled “mood list” appeared on the screen, which participants were to press to take them to the mood checklist task. As the main display was frozen, participants could effectively rehearse or strengthen the to-be-suspended goal for as long as they wished before pressing the button to complete the mood task. According to Altmann and Trafton’s (1999) model, a goal that will be suspended for longer will require a greater degree of strengthening to boost its base-level activation and mitigate the effects of decay. This view is contrary to that of Anderson and Douglass (2001), who predicted

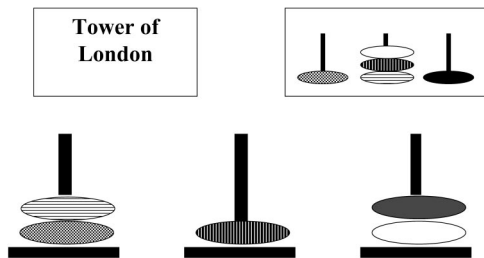


Figure 2. The five-disk Tower of London problem.

¹ An unpublished experiment from our laboratory using the same ToL methodology (25 trials, 6 interruptions) found no difference in resumption times when one, two, or three checklists were presented during a 12-s interval, $F(2, 44) = 0.62$.

that participants would not engage in any rehearsal process and would instead choose to pay the cost of forgetting at retrieval. We aimed to test the effects of interruption length and prior knowledge of interruption length on both preparation and retrieval times as well as any effect that the former may have had on the latter.

Method

Participants. Participants were 20 undergraduate students at Cardiff University who received course credit for their participation.

Apparatus, materials, and procedure. The task was carried out on a personal computer using a ToL computer program written in Visual Basic 6.0. Two sets of ToL disks and pegs were displayed: In the larger of the two sets, the disks were under the control of the participant; the smaller set showed the goal state. The main display panel (33×25 cm) showed a ToL starting configuration with five different-colored movable disks arranged on three pegs. In the top right-hand corner of the screen was a 9×5 cm box with a white background on which appeared a diagram of the goal state, the configuration of which remained fixed throughout a particular trial. The initial state of the disks and the goal state were different for each trial.

To encourage forward planning, we included a specific planning stage at the beginning of each problem. As a new trial began, the word *PLAN* appeared in the top left-hand corner of the screen, and the current display was frozen. After 8 s, this word was replaced by the word *SOLVE*, and participants were then able to move the disks. Eight seconds was selected reasonably arbitrarily as a planning time, but it seemed long enough to dissuade participants from simply executing moves instantly without having planned them at all but not so long that faster planners would become bored. Participants could spend longer than the obligatory 8 s to plan if they so wished, but the main purpose of freezing the display for the 8 s was to reinforce the need for planning at the beginning of each trial. The need for planning was emphasized so that the problems could be solved in the minimum of six moves and so that solution execution could be continuous once the first disk was moved.

After the planning time, disks on the larger display could be moved one at a time by clicking with the mouse first on the peg holding the chosen disk (so that the words *FROM HERE* would appear below it) and then on the peg to which the disk was to be moved (which caused the words *TO HERE* to be displayed). The disk then moved to its allotted position. The program was governed by the additional constraint that no more than three disks could be held on a peg at one time; attempting to put more than three disks on a peg activated a warning sound (a beep) to make the participant aware that the move was not legitimate. A pop-up box appeared upon the completion of each problem, notifying participants how many moves had been made in that trial. This notification did not inform the participant whether this number of moves was over the minimum number possible, but instructions at the beginning of the experiment made it clear that all problems could be solved in six moves or fewer. Clicking in the box on a button labeled *OK* initiated the next trial.

Upon completion of the third move on interruption trials, a small gray button (2×4 cm) appeared centrally at the bottom of the screen; the participant was to click on this button to proceed to the mood task. When the button was on the screen, no disks could be moved even though the main ToL display remained clearly in view. On half of the interruption trials, the button was simply labeled "mood," whereas on the other half of the trials it also indicated the duration of the forthcoming interruption, that is, "mood (1 list)" or "mood (3 lists)." Clicking on this button took the participant to the interrupting task. The ToL display was replaced with a 6×4 cm mood checklist box positioned in the center of an otherwise blank screen. This box contained a list of six descriptions, one below the other in a 10-point Arial font, describing a mood continuum, for example, "extremely alert," "fairly alert," "slightly alert," "slightly tired," "fairly tired," and "extremely tired." Clicking with the mouse highlighted a statement and

recorded this response. Interruptions were either 6 s in duration (short), consisting of just one mood checklist, or 18 s in duration (long), consisting of three mood checklists each displayed sequentially for 6 s. No mood checklist was repeated during the course of the experiment. After each 6-s interval, the checklist would either change automatically to the next list or automatically return the participant to the primary task. The program recorded time to resume the primary task as well as the time between the appearance of the button on the screen and when it was pressed (preparation time).

Participants were given standardized instructions at the start of the experiment and two practice trials to gain familiarity with the task. The second practice trial also gave an example of the long interruption condition, thereby demonstrating the length of time available for the completion of each checklist before it changed automatically. Additional practice trials were given if the participant thought them necessary. The whole experiment typically lasted half an hour.

Design. A 2×2 repeated measures design incorporated the length of the interruption (6 s or 18 s) and also whether the participant knew this interruption duration before its onset. There were eight interruptions in total, two of each of the four possible types. Interruptions always occurred on the same trials across participants (Trials 4 and 10; Trials 7 and 19; Trials 10 and 21; and Trials 15 and 25), but the lengths of the interruptions were counterbalanced in a partial Latin square design. Each interruption trial was matched to a control (no-interruption) trial (Trials 6, 9, 13, 16, 17, 20, 22, and 23), which consisted of essentially the same problem requiring the same solution path but with the disk colors changed. The remaining nine problems in the experiment were filler trials, which required either four, five, or six moves to solution. The dependent measures were preparation time, time taken to make the fourth move, and the number of trials not solved in the minimum number of moves.

Results and Discussion

We designed the procedures used in these experiments to affect the speed, but not the accuracy, with which participants solved the ToL tasks. As shown in Appendix B, accuracy on the ToL task was high in all conditions in all of these experiments, and there were no reliable effects of the independent variables. In the interest of brevity, the accuracy data are not discussed further.

Preparation time, that is, the time between the mood button's appearing on the screen and its being pressed, was recorded for each interruption condition (see Table 1). A 2 (interruption duration) $\times 2$ (knowledge of interruption duration) repeated measures analysis of variance (ANOVA) showed that prior knowledge of the length of the forthcoming interruption had no effect on preparation times, $F(1, 19) = 1.03$, $MSE = 0.45$, $p = .32$. Also, there was no main effect of interruption length, $F(1, 19) = 1.62$, $MSE = 0.57$, $p = .22$, and no significant interaction, $F(1, 19) = 0.01$. Thus we obtained no evidence that participants rehearsed more those goals that they expected to be suspended longer, a finding that supports the prediction of the Anderson and Douglass (2001) model. Fur-

Table 1
Preparation Time (in Seconds) for Each Interruption Condition in Experiment 1

Interruption condition	<i>M</i>	<i>SE</i>
Short, duration known	1.81	0.18
Short, duration unknown	1.70	0.16
Long, duration known	1.64	0.14
Long, duration unknown	1.45	0.11

ther research would be required to determine the conditions in which rehearsal may occur, perhaps when the costs of forgetting are higher (e.g., a task environment that does not support goal reconstruction). However, this is not explored any further in the current series of experiments, and we now turn instead to look at the postinterruption phase.

Times taken to make the fourth move were recorded (see Table 2). In the interruption condition, this was the time from the offset of interruption until the execution of the next move, and in the control condition, this was the time from completion of the third move until completion of the fourth move. As expected, the time taken to make the fourth move in interruption trials was considerably greater than the time taken to make the equivalent move in the control trials. Resumption times in the interruption conditions were compared using a 2 (interruption duration) \times 2 (knowledge of interruption duration) repeated measures ANOVA. Prior knowledge of interruption duration was not expected to have any effect on resumption time given that there was no evidence that any differential strengthening of goals had taken place during the preparation time. This expectation was confirmed, $F(1, 19) = 0.67$. There was, however, a significant main effect of interruption length, $F(1, 19) = 24.49$, $MSE = 1.11$, $\eta_p^2 = .56$, $p < .01$; participants were significantly slower to resume the primary task when it had been suspended for longer (18 s rather than 6 s). There was no interaction between prior knowledge and interruption length, $F(1, 19) = 0.16$.

Time data were also analyzed separately for perfect trials. It is reasonable to presume that task resumption after the third move in a trial in which participants took six moves to solution may be different from task resumption after the third move in a trial in which seven or more moves were taken. Although the move on which participants were required to resume was always indirect, for nonperfect trials this move would not always have been identical in this regard. Perfect and nonperfect trials were distributed evenly across conditions (there was no difference between conditions in terms of primary task error), but the data were nevertheless analyzed for potential differences. Only 5 participants completed all eight interruption trials in the minimum number of moves, and their data are provided for comparison. There was a significant main effect of interruption length, $F(1, 4) = 10.46$, $MSE = 0.93$, $\eta_p^2 = .72$, $p < .05$; no main effect of prior knowledge of interruption length, $F(1, 4) = 0.68$; and no interaction, $F(1, 4) = 4.75$, $p = .10$. The resumption times from perfect trials are therefore accurately represented by the data from the whole set.

The difference between interruption and control conditions in terms of fourth-move latencies demonstrates that even a brief, undemanding, and unrelated task disrupts solution execution in ToL problems. Furthermore, the cost of goal retrieval was found to

be greater following a longer retention interval, a finding that fits well with the models of Altmann and Trafton (2002) and Anderson and Douglass (2001), both of which apply ACT-R's base-level learning equation to goal memory. In accordance with these models, a goal increases in activation when sampled repeatedly but decays when it is no longer actively involved in the governance of behavior.

It has been noted in previous studies that with practice, participants become better at dealing with interruptions (Hess & Detweiler, 1994; Trafton, Altmann, Brock, & Mintz, 2003). We therefore analyzed the resumption time data in the current experiment to test whether participants would recover more quickly from interruptions occurring later in the experiment than from those occurring earlier in the experiment. Resumption time data were pooled across interruption conditions and analyzed according to position of occurrence. A one-way repeated measures ANOVA showed that the point in the experiment at which the interruption occurred had no significant effect, $F(7, 133) = 1.65$, $MSE = 3.87$, $p = .13$. Participants took no longer to resume the task after the first interruption than after any of the other seven. Furthermore, in each of the eight interruption positions, resumption of the primary task was slower following the longer interruption than following the shorter interruption.

In sum, the findings of Experiment 1 demonstrate that a goal suspended for longer is more time-consuming to retrieve, a finding that seems compatible with the predictions of current theories of goal memory (Altmann & Trafton, 2002; Anderson & Douglass, 2001). Perhaps the null effect of interruption length obtained in some classic interruption studies (e.g., Gillie & Broadbent, 1989) was due to the global dependent measures used, such as postinterruption completion time, which may have masked any effects on goal retrieval by compensatory strategies later on.

Experiment 2

Retrospective memory research has shown that the length of time over which an item must be retained is unlikely to be the only factor to increase forgetting; what happens during that time is also critical (see Greene, 1992). In both the model of Altmann and Trafton (2002) and that of Anderson and Douglass (2001), the loss of information from memory is detailed only as a function of decay over time, and the influence of factors such as the complexity or similarity of the intervening task is less clear. In terms of interruption complexity, predictions can be based on the more detailed of the two models, the goal-activation model, because this also takes into consideration activation of the target in relation to other goals and not just the decrease in activation of that goal since it was last sampled. The model predicts that a greater number of distractor goals will raise the interference level and therefore make sampling of the target goal more difficult. Because "complexity" can be considered in terms of "number of actions to perform, number of subgoals, difficulty in executing individual steps, amount of information to be managed, and so forth" (Byrne & Bovair, 1997, p. 46), one would expect that a greater number of subgoals presented by a more complex task would give rise to greater retroactive interference at the point of retrieval.

To examine this possibility, in Experiment 2 we interrupted participants with either a mood checklist or a verbal reasoning task of the type "A follows B — AB," which required a true or false

Table 2
Time to Make Fourth Move (in Seconds) in Experiment 1

Interruption condition	<i>M</i>	<i>SE</i>
No interruption	2.65	0.19
Short, duration known	5.01	0.25
Short, duration unknown	4.83	0.35
Long, duration known	6.26	0.34
Long, duration unknown	5.91	0.33

decision (Baddeley, 1968). Both tasks involved reading and clicking with the mouse on the most appropriate statement, yet it seems reasonable to assume that the verbal reasoning task was more demanding. The mood statements can be scanned for gist and a response can be decided fairly arbitrarily because there is no right or wrong answer. The reasoning task, on the other hand, is based on grammatical transformation and is thought to involve “higher mental processes” because it correlates with verbal intelligence (Baddeley, 1968). A single reasoning statement involves several elements, including the ordering of A and B in the sentence, the ordering of A and B in the pairing, and whether the sentence is positive or negative, is active or passive, or involves the term *follow* or the term *precede*. These elements are all interdependent in determining the validity of the proposition, and each must be processed, retained, and verified so as to reach a true or false decision. We predicted that reactivation of a suspended goal would take longer upon returning to the primary task when the interruption was a verbal reasoning problem (involving a number of subelements) than when the interrupting task was a mood checklist (with a relatively low processing demand).

Method

Participants. Twenty-two participants took part in the study in exchange for course credit. All were undergraduate psychology students at Cardiff University and had not participated in Experiment 1.

Apparatus, materials, and procedure. The same ToL program was used as previously except that now the onset of interruption was under the control of the computer, occurring automatically on completion of the third move. There were six interruptions, involving completion of either one mood checklist (“happy–sad,” “angry–calm,” or “bored–interested”) or one verbal reasoning task requiring verification of the order of two subsequent letters: “A precedes B — BA” (false), “A does not precede B — BA” (true), or “A is preceded by B — AB” (false). The interruption display was very similar for both: The screen went blank at the point of interruption, and the relevant information appeared in a box in the center of the screen in a 12-point Arial font. The participant was to highlight either the appropriate mood statement or *True* or *False* for the reasoning task. Both tasks were self-paced, and a button labeled *Continue* and positioned below the other task items served both to end the recording of the time spent on the interruption and to return the participant to the primary task. Participants completed two ToL practice trials and were also given examples of both a mood task and a reasoning task that were separate from the ToL problems.

Design. A repeated measures design was used in which each participant completed six interruption trials and six matched control trials. Type of interruption was counterbalanced so that for half of the participants, reasoning tasks occurred on Trials 4, 7, and 15 and mood checklists on Trials 12, 19, and 25; for the other half of the participants, the mood checklists occurred on the earlier trials and the reasoning tasks on the later trials.

Results and Discussion

One participant whose resumption time data were more than three standard deviations from the mean was excluded from the analyses, and the move time data for the remaining participants are shown in Table 3. A repeated measures ANOVA revealed a significant effect of condition, $F(2, 40) = 25.84$, $MSE = 1.10$, $\eta_p^2 = .56$, $p < .01$, such that participants took less time to make the fourth move following the mood checklist than following the reasoning task and took still less time in the uninterrupted control

Table 3
Time Taken (in Seconds) to Make the Fourth Move in Experiment 2

Condition	<i>M</i>	<i>SE</i>
Control	2.74	0.21
Mood list	4.09	0.23
Reasoning task	5.05	0.31

condition. Paired comparisons revealed significant differences between each of the three conditions. In terms of practice effects, there was no significant effect on resumption times of the order in which the six interruptions occurred, $F(5, 100) = 1.45$, $MSE = 5.75$, $p = .21$.

The short verbal reasoning tasks were chosen so as to be roughly equated with the mood checklist in terms of time needed for completion. Time on each task, however, was under the control of the participant and was not imposed by the computer program. This was the case for a number of reasons. We thought that a set time limit would pressure participants too much on the more complex task, which in turn might impair their problem-solving abilities. This could result in the participant either not reaching an answer in the given time or hurriedly trying to supply any response before the task automatically finished. Setting the interval at too long a duration, on the other hand, could result in a period of unfilled time during the interruption. Unfilled time of course, by definition, has very low processing demands and so may confound the results in this experiment concerning interruption complexity. For these reasons, interruption length was similar in the two conditions but not strictly controlled. In light of this, and in light of the significant effect of interruption length obtained in the preceding experiment, we analyzed whether the effect of complexity was independent of an effect of interruption length. First, there was a slight trend for participants to spend more time on the reasoning task ($M = 5.32$ s, $SD = 1.55$) than on the mood checklist ($M = 4.77$ s, $SD = 0.93$), but this difference was not statistically significant, $t(20) = 1.29$, $p = .21$. Also, there was no significant correlation between the time spent on the interruption and the time needed to resume the ToL task following both the mood checklist and the verbal reasoning task, according to a Pearson’s test, $r(42) = .18$, $p = .25$. It could therefore be presumed that the difference in resumption times obtained between conditions is the result of differences in the complexity of the two tasks.

The models of Anderson and Douglass (2001) and Altmann and Trafton (2002) do not explicitly account for the current findings because activation decrease in these models is solely a time-based function. The effect of complexity, however, suggests that the type of interpolating activity as well as the length of that activity can influence primary task performance. This has already been demonstrated for ordinary declarative memory elements by way of distractor task paradigms (e.g., Kroll & Kellicutt, 1972; Nakajima & Sato, 1989), as well as with task interruption specifically (Fischer & Glanzer, 1986). If the Anderson and Douglass (2001) and Altmann and Trafton (2002) models are to maintain that memory for goals is equivalent to memory for other declarative memory elements, then further elaboration of the models may be necessary to account for the observed effect of intervening task type.

Anderson and Douglass (2001) did not incorporate any notion of interference into their model: Although goals are subject to decay and thus incur a time cost to retrieve, they are still reliably retrieved in a last-in, first-out manner. The presence or absence of other distracting goals therefore has no impact on the process of goal retrieval; the only factor is the passage of time since the target goal was last sampled. In the Altmann and Trafton (2002) model, on the other hand, there is no special ordering, and goal selection is dependent on a goal's activation at that time. Old goals are held at an interference threshold, described as a "mental clutter" or, in formal terms, as the mean activation of the most active distractor. If activation of a target goal exceeds this threshold, then it is more likely to be sampled than any other, although activation is subject to noise, and so selection on a given cycle is still not guaranteed. If more goals are activated during the interruption, then the interference level will be higher at resumption, creating greater retroactive interference for the target goal. Or, if goals are simply stored in declarative memory, then the more declarative information activated during the interruption would also create retroactive interference. The current results seem compatible with this interpretation.

However, there are also other possible explanations for the effect of the complexity of the interrupting task. It may be that the more complex task is simply better at suppressing rehearsal. Rehearsal of the suspended goal during an undemanding retention interval would refresh base-level activation and aid subsequent retrieval; suppression of this rehearsal during a more complex task would mean that a greater degree of strengthening would be required at resumption to overcome decay. Such an explanation could be accommodated within the goal-activation model, which makes reference to the strategic strengthening of goals, but would be incompatible with the Anderson and Douglass (2001) model, which eschews the idea of goal rehearsal.

Another interpretation of the complexity effect is in terms of task-switching costs rather than in terms of goal maintenance. That is, it may be more effortful to shift from the reasoning task than from the mood checklist because of the need to disengage a more complex task set, regardless of whether task-relevant goals are to be retained during the interruption. The task-switching literature is ambiguous with regard to effects of complexity: Studies examining whether there is a greater switch cost in changing to a more complex task have found evidence both in support (Garcia-Ogueta, 1993; Rubinstein et al., 2001) and to the contrary (Allport, Styles, & Hsieh, 1994). The current findings therefore warranted further experimental scrutiny, and in the following experiment, we examined whether the observed effect of interruption complexity was one of task switching or goal maintenance.

Experiment 3a

Experiment 3a marked a shift to a slightly different methodology. To be confident that we were actually assessing the effects of goal maintenance—that participants were in fact retaining and retrieving goals following within-trial interruptions rather than simply reconstructing—we removed the goal-state picture from the screen following the planning stage. Participants therefore had no option but to execute their moves from memory, because replanning upon returning to the ToL display was not a viable option. Slightly different interruptions were used in this experi-

ment, with simple and complex mental arithmetic sums (e.g., $2 + 3$ or $17 + 36$) replacing the mood and reasoning tasks, respectively. This allowed us to further substantiate and replicate the effect by using a different set of tasks that varied in complexity; furthermore, the choice of mental arithmetic allowed us to measure performance on the interruption. Because both the simple and complex interruptions involved arithmetic sums, they were more comparable than the tasks used previously, and the distinction in complexity was less arbitrary. ACT-R makes the prediction that the answer to common, well-practiced sums can simply be retrieved from memory; on the other hand, large and infrequent number pairings require calculation, a process that imposes greater demands on working memory.

In this experiment, we aimed to test whether the effect of complexity was due to a more demanding intervening task disrupting the maintenance of pending goals or to the possibility that changing from the "mental set" of a complex task takes longer than switching from a simple task. To this end, secondary tasks (both simple and complex) were positioned either in the middle of solution execution (requiring suspension and resumption of primary task goals, as before) or between completed ToL tasks (creating a task switch but requiring no goal maintenance). We tested whether the more complex interruption might not only increase resumption times within a trial but also increase self-selected planning times as the participant switched to a new trial.

Method

Participants. Twenty-four undergraduate psychology students participated in the study in exchange for course credit.

Apparatus, materials, and procedure. The ToL program was modified slightly in this experiment so that the goal state was no longer visible during solution execution. The absence of the goal state would make trials slightly more difficult, as participants would no longer be able to support solution execution with online planning and verification. In view of this change, five-move rather than six-move problems were used. These were the same as the trials used in the previous experiments except that the starting states were advanced by one move. Interruption thus occurred after the second move, so that, as before, there were still three moves to completion following interruption.

Participants were unable to move any disks during an initial planning phase but clicked on a button labeled *Ready* once they had planned their solution sequence. This served both to activate the program and to remove the goal-state picture from the screen. Given that planning time was a dependent variable in this experiment, no minimum planning time was enforced as it had been in the two previous experiments. When participants felt that the goal state had been achieved, they clicked on another button labeled *Complete*. This revealed the goal state and triggered a pop-up box that explicitly informed them whether they had been correct. Clicking on *OK* initiated the next ToL trial.

Interruption was initiated either upon completion of the second move in a sequence or by clicking the *OK* button at the end of a completed trial. Interruptions comprised either four simple sums that never exceeded nine (e.g., $2 + 3$, $4 + 1$) or one complex sum that always required a carry over a tens boundary (e.g., $27 + 38$). Pilot work showed that these two tasks took roughly the same amount of time to complete. Clicking a button labeled *Continue* presented further sums in the simple condition and served to return the participant to the primary task.

Participants read instructions and completed three practice trials. Upon completion of the first and second practice trials, participants attempted examples of the simple and complex interruption conditions, respectively. They were told to complete each of the problems as quickly and accurately

as possible. If they wanted to take a brief pause at any point in the experiment, they were told to do so only immediately after completing a ToL problem when the pop-up *OK* box was on the screen, never once a new trial had commenced or following one of the secondary tasks.

Design. A 2×2 repeated measures design was used in which the secondary task was either simple or complex mental arithmetic that occurred either within or between ToL trials. Participants completed two of each of the four interruption trials as well as eight matched control trials and nine filler problems.

Results and Discussion

Move times were recorded with respect to secondary task type and position (see Figure 3) and subjected to a 2 (position: within or between trials) \times 3 (secondary task: complex, simple, or control) repeated measures ANOVA. There was a main effect of interruption position, $F(1, 23) = 107.16$, $MSE = 26.97$, $\eta_p^2 = .82$, $p < .001$, such that move times were greater between trials (planning time) than within a trial (time to execute the third move). This result was to be expected, as it takes longer to plan new goals than to retrieve old ones. There was also a main effect of secondary task type, $F(2, 46) = 8.09$, $MSE = 7.47$, $\eta_p^2 = .26$, $p < .001$, such that move times were greater following a simple secondary task than no secondary task, and greater still when the intervening activity was a complex mental arithmetic sum. It is interesting that there was also a significant interaction, $F(2, 46) = 7.43$, $MSE = 7.32$, $\eta_p^2 = .24$, $p < .01$. Pairwise comparisons showed a significant difference between each level of secondary task type for interruptions occurring within trials, but there was no significant difference between any of these conditions when the secondary task occurred between trials.

The complexity effect for within-trial interruptions was replicated using this new ToL paradigm, which adds to the generality of the results. For between-trial interruptions, switching from the more complex task to a new ToL problem did not cause a significant impairment in participant planning speed or efficiency, indicating that secondary task complexity does not affect the ease of switching between completed tasks. The finding that there is little or no cost to performance when the interruption arises between tasks is in line with existing research reporting that interruptions are least disruptive when they occur between completion of one goal and formulation of the next (Miyata & Norman, 1986; Monk, Boehm-Davis, & Trafton, 2002).

As a further point, we investigated whether the requirement to maintain goals over the course of the interruption might have

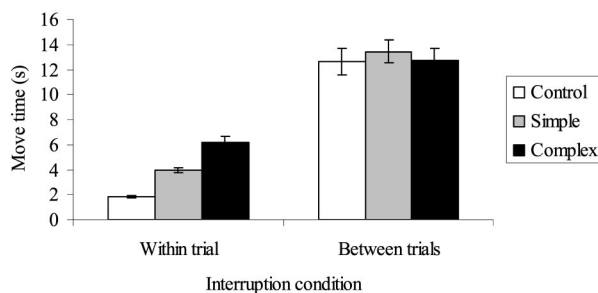


Figure 3. Time taken (in seconds) to make the fourth move in Experiment 3a.

compromised performance on the arithmetic task relative to between-trial interruptions when there was no associated memory load. Error on the arithmetic tasks was overall quite low. As expected, accuracy was best for simple sums, with only one incorrect answer reported across all participants in each of the between- and within-trial conditions. For complex sums, the 24 participants made a total of 4 errors between them on between-trial tasks and 10 errors within trials. These differences were not statistically significant according to a 2×2 chi-square analysis, $\chi^2(1) = 0.375$. Also analyzed was the time taken to complete the interrupting arithmetic tasks. A 2×2 repeated measures ANOVA showed that there was no main effect of complexity, $F(1, 23) = 0.66$, which indicates that the simple and complex tasks were well-matched in duration. There was, however, a main effect of interruption position, $F(1, 23) = 17.79$, $MSE = 2.52$, $\eta_p^2 = .44$, $p < .01$, such that participants took significantly longer to complete the arithmetic tasks, both simple and complex, when they occurred within trials rather than between trials. This finding quite clearly shows that in this ToL paradigm, participants were retaining goals across the interruption and that this goal maintenance imposed an overhead cost on performance of the secondary task. The interaction between interruption position and complexity was not statistically significant, $F(1, 23) = 0.23$.

Move times were also analyzed for practice effects. Given the large difference in move times between the first move and the third move, these were analyzed separately (although whether the interruption was simple or complex was combined across interruption positions). There was a significant practice effect in terms of between-trial interruptions, $F(3, 69) = 15.15$, $MSE = 16.89$, $\eta_p^2 = .40$, $p < .01$, such that planning times on the first and second trials were significantly longer than on any of the subsequent trials. Participants presumably became more efficient at planning as they gained familiarity with the task. There was also a significant practice effect for interruptions occurring within trials, $F(3, 69) = 3.99$, $MSE = 10.95$, $\eta_p^2 = .15$, $p < .01$, with resumption times following the first interruption being significantly greater than those after either the second or the third interruptions. This is surprising given that practice effects were not obtained in the previous two experiments, but it could perhaps be that the slightly different methodology (arithmetic interruptions, goal state not visible) caused the first interruption to be particularly disruptive.

Having established that the effect of complexity—at least in this memory-based ToL paradigm—is one of goal maintenance and not one of task switching, we were able to use the results of the current experiment to infer more about the source of this effect. The methodology of Experiment 3a lends itself easily to a test of the hypothesis that it is the number of intervening goals that increases interference and renders reinstatement of the suspended goal more time-consuming. In the simple interruption condition, participants were to complete four sums, compared with just one calculation in the complex condition. Although the simple interruption involved the completion of more individual tasks, it was still the complex task (comprising fewer, yet interdependent, subgoals) that resulted in longer resumption times. It would seem, then, that an explanation simply in terms of the number of activated subgoals is not sufficient; more important are the processing demands imposed by each.

Experiment 3b

In Experiment 3b, we sought to further substantiate the effect of complexity by using interruption intervals that were strictly controlled in terms of interruption length. Interruptions of 9 s in duration were judged to be suitable on the basis of the time taken to complete the complex calculation in Experiment 3a. As well as the simple and complex arithmetic tasks used before, this experiment included an additional “unfilled” interruption condition that involved the suspension and resumption of primary task goals (the ToL display was blanked for the duration), but there were no intervening goals to be achieved during that time (no secondary task).

Method

Participants. Participants were 27 undergraduate psychology students at Cardiff University who participated in exchange for course credit.

Apparatus, materials, and procedure. The same ToL program was used as in Experiment 3a, involving five-move problems and a goal state that was concealed throughout the execution phase. Interruptions always occurred after completion of the second move and were each 9 s in duration. There were three types of interruption: one complex arithmetic sum, a series of four simple sums, or no secondary task (the screen was blanked and displayed *Please wait*). The arithmetic interruptions featured a *Continue* button below the numerical display. On simple trials, clicking this button would serially display new sums throughout the 9-s period. On complex trials, for which there was only one calculation, clicking the *Continue* button would change it to read *Loading. . .* so that the participant realized that his or her answer had been registered, but the ToL display did not return until 9 s was reached. In each case, the computer ended the interruption after the 9-s interval.

Participants completed at least two practice trials and were also given examples of both the simple and complex arithmetic tasks (unlike in the experiment proper, these practice tasks had no time limit). Participants were not warned of the 9-s time limit to complete the arithmetic tasks because we thought that imposing time pressure might lead them to adopt less processing-intensive strategies that might attenuate any effects.

Design. A repeated measures design was used in which each participant completed six control and six interruption trials (two of each interruption type), the order of which was counterbalanced.

Results and Discussion

Move times were collected with respect to interruption condition (see Table 4). A one-way repeated measures ANOVA revealed a significant effect of condition, $F(3, 78) = 45.98$, $MSE = 2.82$, $\eta_p^2 = .64$, $p < .01$. Pairwise comparisons demonstrated that each condition was significantly different from each of the others. Resumption times were shorter following the simple task than following the complex task, a result that replicates the previous finding but with the potential confound of interruption length

eliminated. Task resumption was significantly quicker in the unfilled interruption condition than in either of the arithmetic conditions, a finding that was as expected because an unfilled break imposes no complex processing demands and participants would be free to rehearse primary task goals should they so choose. Although this unfilled condition presented no interfering subgoals during the interruption period, resumption times in this condition were still longer than the time taken to make the third move in uninterrupted control trials. Task resumption times were analyzed for practice effects, but there was no significant difference over the course of the six interruptions, $F(5, 130) = 1.35$, $MSE = 8.16$, $p = .25$.

The effect of complexity replicates that obtained in the previous experiment, and we now consider explanations for this finding. Experiment 3a cast doubt on two suggestions—that the effect is one of task switching, or that it is due to the number of subgoals activated during the interruption increasing interference. Another possibility is that a complex task is more successful at suppressing rehearsal, so that base-level activation cannot be maintained as efficiently; however, the current experiment suggests that perhaps participants are not motivated to rehearse at all, regardless of the level of difficulty of the intervening task. In the unfilled interruption condition, participants were free to rehearse primary task goals as they wished, yet still there was a cost of retrieval of around 1 s relative to the control trials. If the principal factor governing disruption in these complexity experiments was the opportunity to refresh base-level activation during the interruption interval, then one would expect that in a condition in which this rehearsal was freely permitted, there would be relatively little or no cost at goal retrieval.

One further possibility is that the effect of complexity may not be due to problems maintaining base-level activation; rather, the influence of associative activation may be of importance. In ACT-R, a goal's total activation comprises both base-level and associative activation, a type of attentional activation that maintains those memory nodes relevant to the current task in an accessible state. Unlike base-level activation, associative activation is limited in ACT-R so that the more items that are currently relevant, the less the activation each one receives (e.g., Anderson, Reder, & Lebiere, 1996). It could perhaps be expected that a complex interrupting task with high processing demands might require more associative activation to support that task, therefore leaving less associative activation available for the maintenance of the suspended goal. During a simple task, the suspended goal may be retained at a higher level of activation because the overall demands on working memory at this time are lower. This limitation on activation should perhaps not be considered in terms of a restriction on the number of items per se that can be “stored” at one time; instead, it may reflect a limitation on the degree to which it is possible to process and coordinate multiple task goals according to their varying cognitive demands.

Whereas the Anderson and Douglass (2001) model uses only base-level activation to determine memory for goals, the Altmann and Trafton (2002) model also incorporates the influence of associative activation through its role in priming. By this token, Altmann and Trafton's (2002) model may potentially be able to integrate the observed effects of interruption length and complexity by elaborating on the role of this important second source of activation in the model. Associative activation is, of course, of

Table 4
Time Taken (s) to Make the Fourth Move in Experiment 3b

Condition	<i>M</i>	<i>SE</i>
Control	1.99	0.11
No task	2.84	0.17
Simple arithmetic	5.17	0.31
Complex arithmetic	6.80	0.54

most importance in this model at the point of goal retrieval through priming, but perhaps if sufficient resources are available, associative activation that persists from before the goal is suspended may also help maintenance of that goal throughout the interruption.

General Discussion

The current series of experiments demonstrated the disruptive effects of interruption on problem solving in the ToL task. In each of the four experiments, midtask interruptions incurred a cost in terms of time taken to make the next move in a solution sequence, and this cost was accentuated by both a longer intervening activity and a more complex intervening activity. The effect of interruption complexity was obtained in both the display-based and memory-based versions of the ToL task, adding to the robustness of the finding. Generally speaking, interruption produced little effect on the incidence of error in the ToL task, but this may have been due to the fact that the trials were not too complex and that there was thus little variation in performance. In terms of the effect of practice, the results are ambiguous. Only one of the experiments (Experiment 3a) found that task resumption times were longer for interruptions at the start of the experiment than for those occurring later. The issue of practice effects would therefore require further investigation with more of a controlled manipulation, perhaps by using participants who were well-practiced at the ToL task so as to disambiguate the effects of practice recovering from interruptions and the effects of ToL expertise.

Despite the broad relevance of task interruption to experimental psychology, as well as to commercial and human factors domains, empirical research on the topic has been relatively thin and has lacked coherence. The experiments reported here illustrate how some coherence might be found by using a controlled task environment, and they suggest a specific theoretical framework from which to continue establishing some of the basic effects of interruptions. We believe we have demonstrated the suitability of the ToL task as a primary task in the study of interruptions. It allows for a fine-grained analysis of performance at the level of individual moves, therefore enabling an analysis of the basic cognitive cost of interruptions uncontaminated by compensatory strategies. Problem solving is segmented into a sequence of short individual moves that can be timed precisely and interrupted and resumed at any point. This level of precision stands in contrast to that of previous studies in which more general dependent variables have been measured, such as overall task time or postinterruption task time (e.g., Gillie & Broadbent, 1989).

The experiments reported in the current study suggest that the maintenance and retrieval of goals are not the straightforward and reliable operations that the goal stack implies. Although the finding that goal memory does have its limitations is perhaps to some extent obvious, the idea that goals are constantly active and readily available is an assumption that has been easy to rely on in cognitive psychology and has only recently been explicitly challenged. The current work provides empirical data to discount this assumption and suggests that theories of cognition should look to specify in more detail those processes involved in the suspension and resumption of goals.

The models of Anderson and Douglass (2001) and Altmann and Trafton (2002) posit that goal memory follows a retention function similar to that of other declarative memory elements, a proposal

that is supported by the experiments in the current study. The findings of Experiment 1 fit well with the base-level learning equation of ACT-R that is incorporated into these models. Activation increases rapidly with use but decreases as a power function since the goal was last sampled. However, although the Anderson and Douglass (2001) model can easily accommodate the findings on interruption length, it is found wanting with respect to the present findings on interruption complexity. A simple decay-based function was adequate for modeling the suspension and resumption of goals in the ToH, but the suspension and resumption of goals as a result of an unrelated interrupting activity are less simple. Interruptions are heterogeneous; time is not the only dimension along which they can vary, so the model is only applicable up to a certain point. This finding of interruption complexity is not readily explained by the goal-activation model either, but with further elaboration it is possible that the model might have the mechanisms to do so. Altmann and Trafton's (2002) goal-activation model goes slightly further by additionally incorporating the influence of associative activation and also the idea of interference from other old goals, although these features are not explicitly linked to the effects of cognitive load during an interruption. Further elaboration of these features is therefore needed to better understand the effect of intervening task complexity on goal retrieval, but the goal-activation model seems to provide a fruitful basis for this further exploration.

To summarize, the implications of the current study are threefold. First, we have highlighted the usefulness of the ToL task as a primary task. It enables the assessment of interruption effects at a fine-grained level in a task environment that limits extraneous variables. Second, we have shown that even relatively short and undemanding interruptions can have a large and reliable disruptive effect on resumption time and that interruption duration and complexity exacerbate the effect. Third, the plausibility of the goal-activation model (Altmann & Trafton, 2002) as a framework for the study of interruption effects has been demonstrated, although further elaboration of the model—specifically in terms of the role of associative activation—may be needed to account for the effects of interruption complexity. Nonetheless, the goal-activation model, coupled with the present study's findings, provides a useful springboard for future research on interruptions.

References

- Allport, A., & Wylie, G. (2000). Task-switching, stimulus-response binding, and negative priming. In S. Monsell & J. S. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes* (pp. 35–70). Cambridge, MA: MIT Press.
- Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Switching intentional set: Exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and unconscious information processing* (pp. 421–452). Cambridge, MA: MIT Press.
- Altmann, E. M., & Trafton, G. J. (1999). Memory for goals: An architectural perspective. In *Proceedings of the twenty-first annual meeting of the Cognitive Science Society* (pp. 19–24). Hillsdale, NJ: Erlbaum.
- Altmann, E. M., & Trafton, G. J. (2002). Memory for goals: An activation-based model. *Cognitive Science*, 26, 39–83.
- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R., & Douglass, S. (2001). Tower of Hanoi: Evidence for the cost of goal retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1331–1346.

- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.
- Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, 30, 221–256.
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396–408.
- Baddeley, A. D. (1968). A three-minute reasoning test based on grammatical transformation. *Psychonomic Science*, 10, 341–342.
- Bailey, B. P., Konstan, J. A., & Carlis, J. V. (2000, October 8–11). *Measuring the effects of interruptions on task performance in the user interface*. Paper presented at the 2000 IEEE International Conference on Systems, Man, and Cybernetics, Nashville, TN (Vol. 2, pp. 757–762).
- Bainbridge, L. (1984). Analysis of verbal protocols from a process control task. In E. Edwards & F. P. Lees (Eds.), *The human operator in process control* (pp. 146–158). London: Taylor & Francis.
- Brandimonte, M., Einstein, G. O., & McDaniel, M. A. (Eds.). (1996). *Prospective memory: Theory and applications*. Mahwah, NJ: Erlbaum.
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 10, 12–21.
- Byrne, M. D., & Bovair, S. (1997). A working memory model of a common procedural error. *Cognitive Science*, 21, 31–61.
- Einstein, G. O., McDaniel, M. A., Thomas, R., Mayfield, S., Shank, H., Morrisette, N., & Breneiser, J. (2005). Multiple processes in prospective memory retrieval: Factors determining monitoring versus spontaneous retrieval. *Journal of Experimental Psychology: General*, 134, 327–342.
- Einstein, G. O., McDaniel, M. A., Williford, C. L., Pagan, J. L., & Dismukes, R. K. (2003). Forgetting of intentions in demanding situations is rapid. *Journal of Experimental Psychology: Applied*, 9, 147–162.
- Fischer, B., & Glanzer, M. (1986). Short-term storage and the processing of cohesion during reading. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 38A, 431–460.
- Flynn, E. A., Barker, K. N., Gibson, J. T., Pearson, R. E., Berger, B. A., & Smith, L. A. (1999). Impact of interruptions and distractions on dispensing errors in an ambulatory care pharmacy. *American Journal of Health-System Pharmacy*, 56, 1319–1325.
- Garcia-Ogueta, M. I. (1993). Internal attentional switching: Effects of predictability, complexity and practice. *Acta Psychologica*, 83, 13–32.
- Gillie, T., & Broadbent, D. (1989). What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research*, 50, 243–250.
- Goschke, T., & Kuhl, J. (1996). Remembering what to do: Explicit and implicit memory for intentions. In M. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.), *Prospective memory: Theory and applications* (pp. 53–91). Hillsdale, NJ: Erlbaum.
- Greene, R. L. (1992). *Human memory: Paradigms and paradoxes*. Hillsdale, NJ: Erlbaum.
- Griffon-Fouco, M., & Ghertman, F. (1984). Recueil de données sur les facteurs humains à l'Electricite de France. In *Operational safety of nuclear power plants* (pp. 157–172). Vienna: International Atomic Energy Agency.
- Gynn, M. J., McDaniel, M. A., & Einstein, G. O. (1998). Prospective memory: When reminders fail. *Memory & Cognition*, 26, 287–298.
- Hess, S. M., & Detweiler, M. C. (1994). Training to reduce the disruptive effects of interruptions. In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (Vol. 2, pp. 1173–1177). Santa Monica, CA: Human Factors and Ergonomics Society.
- Kirmeyer, S. L. (1988). Coping with competing demands: Interruption and the Type A pattern. *Journal of Applied Psychology*, 73, 621–629.
- Kroll, N. E. A., & Kellicutt, M. H. (1972). Short-term recall as a function of covert rehearsal and of intervening task. *Journal of Verbal Learning and Verbal Behavior*, 11, 196–204.
- Lahlou, S., Kirsh, D., Rebotier, T., Reeves, C., & Remy, M. (2000). Experimental study of the effect of interruption on office work. Experiment 1: Types of interruptions. Retrieved December 10, 2000, from <http://adrenaline.ucsd.edu/edf/Experiment1.htm>
- Logan, G. D. (2004). Working memory, task switching, and executive control in the task span procedure. *Journal of Experimental Psychology: General*, 133, 218–236.
- Mayr, U., & Keele, S. W. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology: General*, 129, 4–26.
- McFarlane, D. C., & Latorella, K. A. (2002). The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction*, 17, 1–61.
- Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart, & Winston.
- Miyata, Y., & Norman, D. (1986). Psychological issues in support of multiple activities. In D. A. Norman & S. W. Draper (Eds.), *User centered systems design: New perspectives on human-computer interaction* (pp. 265–284). Hillsdale, NJ: Erlbaum.
- Monk, C., Boehm-Davis, D., & Trafton, J. G. (2002). The attentional costs of interrupting task performance at various stages. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 1824–1828). Santa Monica, CA: Human Factors and Ergonomics Society.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7, 134–140.
- Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, 53, 53–81.
- Nakajima, Y., & Sato, K. (1989). Distractor difficulty and the long-term recency effect. *American Journal of Psychology*, 102, 511–521.
- Paquot, J., Eyrolle, H., & Cellier, J. (1986). Conséquences de l'interférence entre les tâches de saisie sur la durée des traitements et l'occurrence des erreurs. In R. Patesson (Ed.), *L'homme et l'écran* (pp. 199–209). Brussels, Belgium: Editions de l'Université de Bruxelles.
- Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58, 193–198.
- Posner, M. I., & Konick, A. W. (1966). On the role of interference in short-term retention. *Journal of Experimental Psychology*, 72, 221–231.
- Reason, J. T. (1990). *Human error*. Cambridge, England: Cambridge University Press.
- Rubinstein, J. S., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 763–797.
- Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B*, 298, 199–209.
- Trafton, J. G., Altmann, E. M., Brock, D. P., & Mintz, F. E. (2003). Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies*, 58, 583–603.
- Ward, G., & Allport, A. (1997). Planning and problem-solving using the five-disc Tower of London task. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 50A, 49–78.
- Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, 37, 60–82.
- Zeigarnik, B. (1927). Das Behalten erledigter und unerledigter Handlungen [The retention of completed and uncompleted actions]. *Psychologische Forschung*, 9, 1–85.
- Zijlstra, F. R. H., Roe, R. A., Leonora, A. B., & Krediet, I. (1999). Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational and Organizational Psychology*, 72, 163–185.

Appendix A

Tower of London Problems

No.	Start	Goal
1.	<R, G> <P> <B, Y>	<G> <Y, P, R>
2.	<R> <P, B, Y> <G>	<R> <Y, G, P>
3.	<B, Y> <> <R, P, G>	<Y> <G, P> <R, B>
4.	<Y, B> <G, P> <R>	 <G> <P, R, Y>
5.	<R> <Y> <P, G, B>	<Y> <R, G, P>
6.	<G> <R, P> <Y, B>	<P, G, Y> <R>
7.	<R, P, Y> <G> 	<G, Y> <P> <B, R>
8.	<Y> <G, B, R> <P>	<Y, G> <R, P>

Note. Pegs are indicated by angular brackets, and discs within pegs are ordered so that the leftmost is on top. R = red; Y = yellow; G = green; B = blue; P = pink.

Appendix B

Accuracy Data

Trials that were not solved in the minimum number of moves were classed as errors. Errors in each condition, pooled across participants, were subjected to a chi-square goodness-of-fit test and assessed at the $p < .05$ level. In Experiment 1, out of a possible 40 trials in each interruption condition across participants, the numbers of errors were as follows: short, duration known = 7; short, duration unknown = 8; long, duration known = 5; long, duration unknown = 8; $\chi^2(3) = 0.85$, *ns*. Twenty-three of 160 control trials were classed as errors. In Experiment 2, out of a possible 63 trials in each interruption condition, 24 of the mood trials and 33 of the reasoning trials were not completed in the minimum number of moves, $\chi^2(1) = 1.42$, *ns*. Sixty-one of the 126 matched control trials were classed as errors. In Experiment 3a, out of 48 trials in each condition, the numbers of trials not completed in the minimum of five moves were as

follows: within trial, simple = 7; within trial, complex = 8; between trials, simple = 3; between trials, complex = 6. The differences were nonsignificant, $\chi^2(3) = 2.34$. Twenty-five of 192 control trials were classed as errors. In Experiment 3b, 54 trials were completed in each interruption condition, and the numbers of error were as follows: unfilled interruption = 7; simple arithmetic = 8; complex arithmetic = 13; $\chi^2(2) = 2.22$, *ns*. The matched control trials produced a similar proportion of errors (22 out of 162).

Received February 12, 2003

Revision received August 2, 2005

Accepted August 5, 2005 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://watson.apa.org/notify/> and you will be notified by e-mail when issues of interest to you become available!