Interrupting Problem Solving: Effects of Interruption Position and Complexity

Helen M. Hodgetts (HodgettsHM@cardiff.ac.uk)

Dylan M. Jones (JonesDM@cardiff.ac.uk)

School of Psychology
Cardiff University, Cardiff CF10 3AT UK

Abstract

The Tower of London problem (Ward & Allport, 1997) – a task that involves the formulation, retention and execution of a series of task goals – is used as a basis for two experiments that assessed the effects of interruption to task performance. When primary task goals were suspended by the requirement to complete a brief, unrelated secondary activity, a time cost was incurred in retrieving these goals relative to uninterrupted solution execution (Experiment 1). This finding is contrary to the predictions of ACT-R's goal stack (Anderson & Lebiere, 1998) by which retrieval of pending goals is instant and error-free. Furthermore, the time cost incurred was greater when the intervening task was more demanding, and when the interruption occurred earlier in the course of problem solving (Experiment 2). Current models of goal memory (Altmann & Trafton, 2002; Anderson & Douglass, 2001) provide a useful basis for the interpretation of these effects, suggesting that goals may suffer decay and interference like other declarative memory elements.

Introduction

How humans cope with interference and task switching – both voluntary and involuntary – has implications for understanding more general aspects of the cognitive system. The current work uses task interruption as a means to investigate how the cognitive system remembers its goals. Future intentions are often thought to have some privileged status in memory above that of non-goal items (e.g., Goscke & Kuhl, 1996) making them more accessible and less easily forgotten. However, although one may conceive that goals may have some kind of heightened activation, it is difficult to imagine that this is true to the extent of reliably error-free and instantaneous goal retrieval.

The classic view of goal memory is that pending goals reside in a stack (e.g., Ernst & Newell, 1969) and are dealt with in a last in, first out manner. For a new (or interrupting) goal to govern behaviour it needs to be pushed down on top of the stack. Once completed, it is 'popped' off leaving the previously suspended goal once again readily available at the top of the stack. Such a stack is central to the ACT-R cognitive architecture (Adaptive Control of Thought – Rational (Anderson & Lebiere, 1998)). In ACT-R, goals remain active even without rehearsal and do not suffer decay or interference like ordinary declarative memory elements. As such, a suspended goal will always be instantly retrievable without time cost and without error.

Two recent computational models derived from the ACT-R framework challenge this idea of a perfect goal memory, and instead have suggested that goals may be subject to the same costs and limitations as other memory items (Altmann & Trafton, 2002; Anderson & Douglass, 2001). Both models apply ACT-R's base-level learning equation to goal memory predicting that a goal's activation will decrease as a power function since it was last sampled. As such, a goal that has been suspended for longer would be less active and therefore more difficult to retrieve.

The more detailed of the two models, the goal-activation model (Altmann & Trafton, 2002), also takes into account the activation level of other competing goals in determining the ease of goal retrieval. If many other goals are currently active then retrieval of a suspended target goal will be subject to greater retroactive interference, therefore making selection more effortful and more prone to error.

Until now, research on interruptions has lacked a sound theoretical basis upon which to base predictions and interpretation. Nevertheless, the negative effect of interruption is evident both in the laboratory and real world settings. For instance, in work environments associated with high risk such as nuclear power plants (Bainbridge, 1984) and the flightdeck (McFarlane & Latorella, 2002), interruptions are cited as a major contributory factor to human error. In the laboratory, studies have examined a range of factors including the effect of interruption complexity (Gillie & Broadbent, 1989), the point of occurrence of the interruption in the primary task (Miyata & Norman, 1986), interruption similarity (Edwards & Gronlund, 1998), and length (Gillie & Broadbent, 1989). However, the disparate methodologies used and the lack of an overarching theoretical framework make it difficult to compare between studies and draw any firm conclusions.

The current work uses the 5-disc Tower of London (ToL; Ward & Allport, 1997) as a primary task because it allows for a fine-grained analysis of performance at the level of individual move times. Based on the wellknown Tower of Hanoi (ToH) problem, discs are to be moved one at a time from peg to peg until those on the main display exactly match a given goal state. Unlike the ToH, discs are all equal in size and so any can be placed on top of any other.

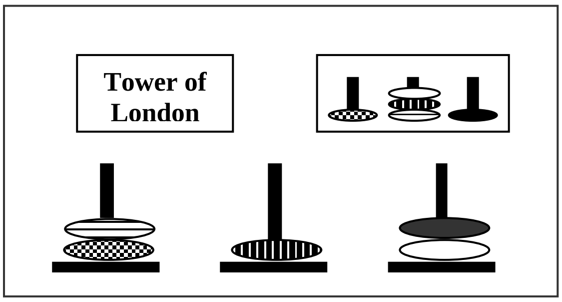

Figure 1: The 5-disc Tower of London task

Interruption during the execution phase of ToL problems involved completion of a mood checklist: Participants were presented with a list of six mood statements describing a mood continuum, e.g., *extremely happy, fairly happy, slightly happy, slightly sad, fairly sad, extremely sad,* of which they were to select the statement that most applied to their mood at that time.

Experiment 1

Participants were interrupted on 6 out of 25 trials, as they completed their third move on six-move ToL problems. It was predicted that time taken to execute the fourth move in the solution sequence following interruption would be longer than time to make the same move on uninterrupted control trials, reflecting the time needed to retrieve a suspended goal. Also, interrupted trials were expected to be more prone to error. A cost of goal retrieval in terms of time or error would support the proposals of Altmann and Trafton (2002) and Anderson and Douglass (2001) that goal memory is not perfect, and instead may be subject to the same limitations as other declarative memory items.

As additional controls, six three-move trials were also included. The starting state of these trials corresponded to the point at which the interruption would have occurred in the interrupted six-move trials. This arrangement allows a test of the hypothesis that participants would take longer to make the first move on a newly presented three-move trial, than to make the fourth move in a previously planned sequence following an interruption. This would indicate that participants were engaging in some form of goal recovery by reinstating previously activated intentions, and not simply starting the planning process anew.

Method

Participants Fifteen students at Cardiff University received course credit for their participation.

Apparatus and Materials The task was conducted on a personal computer using a ToL program written in *Visual Basic 6.0*. The main display comprised five different coloured discs arranged on three pegs. To move a disc, participants were to click on the corresponding button below each peg – first on the one holding the chosen disc, and then on the peg to which it was to be moved. The goal state was displayed in a box in the top right hand corner of the screen and participants were notified via a pop-up box when they had completed the problem. Clicking a button labelled *OK* initiated the next trial and displayed a different start and goal state.

Upon completion of the third move on interruption trials, the screen was blanked apart from a 6 x 4 cm mood checklist box, positioned in the centre. This listed six statements, one below the other, describing a mood continuum. Clicking with the mouse highlighted a statement and recorded this response. Participants then clicked a button labelled *Continue* which returned them to the ToL task at the exact point at which it was left. Typically, the interruption took around 5 s to complete. The program recorded resumption time and number of moves made on each trial.

Design A within-participants design was used. There were 25 trials in total, six of which contained an interruption. These were each matched to a control trial (no interruption) which was essentially the same problem but with the colours of the discs changed. Also, each interruption trial was matched to a three-move problem (starting at the point at which the participant would resume the primary task in an interruption trial). These three equivalent problems were always located at least five trials apart. The remaining seven problems in the experiment were filler trials and were not analysed.

Procedure Participants read a standardised instruction sheet explaining the ToL task and how the discs were to be moved. They were told that all problems could be solved within six moves or fewer and that they should first thoroughly plan this solution path before beginning to move any of the discs. The instructions warned that they would be asked to assess their mood at certain points during the task, but that they should continue with the main part of experiment as quickly as possible afterwards. Two ToL practice trials were given to gain familiarity with the task. The experiment typically lasted half an hour.

Results and Discussion

The number of trials in which errors were made was recorded but there appeared to be little difference between interruption and controls: For data pooled across all 15 participants, out of a possible 90 trials in each condition, 37 interruption and 35 control trials were not completed in the minimum number of moves. Fewer errors were made in the easier three-move trials: Only five of the 90 trials were not completed in the minimum number of moves. Overall, the error rate was quite low, so as a dependent measure it may not be sensitive enough to reveal any effects. However, the fact that participants did sometimes make errors suggests that the retrieval of previously formulated goals is not always reliable.

Time data were also recorded, giving a measure across conditions of time taken to move at a point when the memory load amounted to three subsequent moves (either following an interruption, as part of a continuous sequence, or as a novel problem). Time to make the fourth move in control trials was 4.07 s (sd = 1.50), in interruption trials was 6.97 (sd = 2.10), and at the start of a newly presented 3-move trial was 9.04 (sd = 2.23). A repeated measures ANOVA, $F(2, 28) = 31.87$, $MSE = 2.94$, $p < .01$, and pairwise comparisons showed the difference between each condition to be statistically significant. Data were also analysed separately for perfect trials. Individual data points were pooled across participants and those for which each of the three equivalent problems were solved in the minimum number of moves were selected. The same difference between conditions was found, $F(2, 78) = 62.12$, $MSE = 6.09$, $p < .01$.

Participants took longer to move when retrieving an old goal following interruption, compared to making the equivalent move in the uninterrupted control condition. This difference is likely to be more than just an effect of task switching. A typical cost in the task switching literature is in the region of a few hundred milliseconds (e.g., Rogers & Monsell, 1995), whilst the cost of interruption compared to control moves in the current study was about three seconds. This difference is long enough to demonstrate the additional cognitive requirements of goal recovery. Furthermore, latencies were longer still to make a move at the start of a newly presented three-move trial, since solution execution is not supported by an existing memory trace. The observed time difference is most likely accounted for by the need for additional cognitive processing in the three-move condition, such as the formulation of new goals as opposed to the reactivation of old ones. This suggests that some residual knowledge survives the interruption and participants are actually retrieving old goals rather than simply planning anew.

It has been noted in previous studies that participants become better at dealing with interruptions with practice (Trafton, Altmann, Brock, & Mintz, 2003). The resumption time data in the current experiment were therefore analysed to test if participants recovered quicker from those interruptions occurring later in the experiment. However, a repeated measures ANOVA conducted across the six interruption positions showed this not to be the case, $F(5, 70) = .49$, $MSE = 13.64$, $p > .05$.

Experiment 2

Experiment 2 built on the previous findings by investigating those factors that may accentuate the cost of interruption. Specifically, we tested whether the complexity of the interrupting task and the point in the primary task at which the interruption occurs may influence the ease of goal retrieval.

One might expect that the more complex the interrupting task, the more difficult it will be to resume the suspended goal. ‘Complexity’ can be considered in terms of the number of subgoals involved in the task and the difficulty in executing and coordinating each of these elements (Byrne & Bovair, 1997). If a complex task comprises more components or more subgoals, then it may follow that this would cause greater interference at the point of goal retrieval. The goal-activation model (Altmann & Trafton, 2002), after all predicts that selection of a target goal will be more difficult amongst a greater number of active distractors. The existing literature with regard to the effect of interruption complexity has found both greater disruption (Gillie & Broadbent, 1989) and null effects (Lahlou, Reeves, Rebotier, & Remy, 2000).

Intuitively, given that complex tasks contain many disparate elements of cognitive activity, the ease with which a task is resumed following interruption may be dependent upon the point in the task at which it occurs. The goal-activation model would predict that at points of high memory load, more items or intentions are currently active, meaning that more goals are competing to govern behaviour. Therefore, following an interruption it may be more difficult to select the appropriate goal or specific item to reinstate, increasing resumption times thereby.

Previous research indicates that the point at which an interruption occurs may be critical in determining disruption. One study which required participants to program a video cassette recorder reported that resumption lags were shortest when the interruption occurred at the start of a new subtask, before the new activity began (Monk, Boehm-Davis, & Trafton, 2002). Logically, it could be expected that interruptions occurring between subtasks should be less disruptive: Memory load associated with the primary task is lower after completion of a subgoal, so there are fewer goals competing for ascendancy.

Method

Participants Twenty-four participants at Cardiff University received course credit for their participation.

Apparatus and Materials The same ToL program was used as in Experiment 1, but with a few alterations. To be sure that participants are actually retrieving previously planned goals and not simply replanning after the interruption, a 'screen' appeared over the main display of discs so that moves were to be executed from memory. At the start of each trial, no discs could be moved until the participant clicked a button labeled *Ready* to indicate that they had finished planning. This activated the program so that the discs could be moved, although a screen appeared to conceal the actual location of discs on pegs. Discs were moved in the same manner as before by clicking on the buttons below the pegs; a 'clunk' sound was played each time a disc arrived at its new location. When participants felt that the goal state had been achieved, they clicked a button labeled *Complete* which revealed the location of the discs on pegs and informed them whether or not they were correct. Since solution execution would be done completely from memory and without the support of any online planning, only four-move problems were used for this experiment. They were essentially the same problems used in Experiment 1, but the starting state was advanced by two moves.

Interruptions occurred either before making the first move (triggered as the participant clicked the *Ready* button), or before the third move (initiated on completion on move 2). The simple interruption was a mood checklist as before, and the complex one was a verbal reasoning task of the type "A follows B – AB" [false] (Baddeley, 1968). The display for these was very similar. Interruption duration was under the control of the computer: After 6.75 s, the participant was returned automatically to the ToL display.

Design A 2 (position: first or third move) x 3 (interruption type: no interruption, simple, complex) repeated measures design was used. There were eight interruptions in total (two of each of four types), and each of these were matched to a control trial.

Procedure The procedure was as in Experiment 1 but with participants also being given an example each of the mood and reasoning tasks, as well as two ToL practice trials**.**

Results and Discussion

Two participants whose resumption time data exceeded 3 standard deviations from the mean were removed from analyses. Error data were collected in accordance with condition. For data pooled across the remaining 22 participants, out of a total of 44 trials in each condition, the number not completed in the minimum of six moves were as follows: simple, first move = 16; complex, first move = 22; simple, third move = 14; complex, third move = 14. Although there is a trend for more errors to be made when interrupted at the start of problem execution, these differences were not significant according to a chi square goodness of fit test, χ^2 (3) = 2.61, p > .05. Out of 176 matched control trials, 51 were not completed in the minimum which proportionally, is comparable to the error rate in the experimental trials.

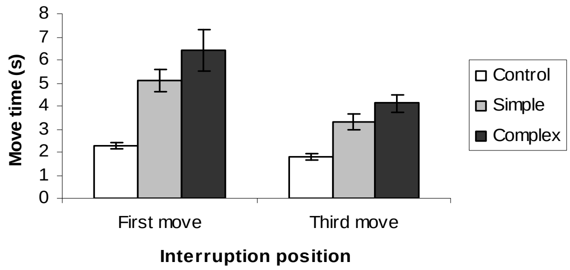

Figure 2: Move time (s) according to interruption position and type.

Move times were recorded (Figure 2) and subjected to a 2 (position) x 3 (interruption condition) repeated measures ANOVA. There was a main effect of position such that times to make the first move were significantly longer than time to make the fourth move, $F(1, 21) = 20.78$, $MSE = 3.69$, $p < .01$. There was also a main effect of interruption type, such that move times were quicker following a simple than a complex interruption, and were quicker still when solution execution was uninterrupted, $F(2, 42) = 27.65$, $MSE = 4.34$, $p < .01$. The interaction between position and complexity did not quite reach significance, $F(2, 42) = 2.44$, $MSE = 3.96$, $p = .10$, although it appears that the differences between conditions in terms of interruption type may be more marked before the first than the third move.

The effect of interruption position was in accordance with predictions made based on the goal-activation model. Immediately after the planning stage, a number of goals relating to the execution of the current problem are activated. This makes selection of the correct goal following interruption more difficult, as there are many active distractors. Towards the end of solution execution however, there are fewer active goals pertaining to the current task so resumption of the suspended goal at this stage will be comparatively less effortful.

The effect of complexity was again as might be predicted based on the goal-activation model. The verbal reasoning task contains a number of elements such as the ordering of the letters and the precise wording of the statement, each of which must be retained and coordinated in order to make a true/false judgment. The number of subgoals activated during this task may increase retroactive interference at the point of task resumption. The suspended goal will require a greater degree of strengthening in order to build up activation to a level above that of other competing goals. Resumption time therefore is increased relative to the simple, mood checklist condition for which fewer subgoals are activated during the interruption.

General Discussion

The experiments show that even brief and relatively undemanding interruptions incur a time cost to primary task performance, demonstrating that retrieval of goals is not as instantaneous a process as ACT-R's goal stack suggests. The cost is exacerbated by both a more complex secondary task and one that occurs earlier in the course of problem solving when demands on working memory are high. The results are as would be predicted by the goal-activation model based on the idea of increased interference. The greater number of subgoals activated during a more complex interruption create more active distractors at the point of task resumption. Similarly, selection of the target goal is impeded if the interruption occurs earlier in the task because of a greater number of active goals, this time those pertaining to the primary task. Resumption times are increased as sampling of the suspended goal becomes more effortful.

These experiments show that goal memory is not perfect and suggest that the goal-activation model may provide a fruitful basis for future research into interruption effects. Given the ubiquity of interruptions in office environments and the associated costs to performance, such research may have practical benefits in the domain of human-computer interaction, for example, establishing the points at which computer-initiated notifications may least disrupt the user.

Acknowledgments

Helen Hodgetts is funded by an Economic and Social Research Council postdoctoral fellowship.

References

Altmann, E. M., & Trafton, G. J. (2002). Memory for goals: An activation-based model. *Cognitive Science, 26,* 39-83.

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.

Baddeley, A. D. (1968). A three-minute reasoning test based on grammatical transformation. *Psychonomic Science, 10,* 341-2.

Bainbridge, L. (1984). Analysis of verbal protocols from a process control task. In E. Edwards and F.P. Lees (Eds.), *The human operator in process control*. London: Taylor and Francis.

Byrne, M. D., & Bovair, S. (1997). A working memory model of a common procedural error. *Cognitive Science, 21,* 31-61.

Edwards, M.B., & Gronlund, S.D. (1998). Task interruption and its effects on memory. *Memory, 6,* 665-687.

Ernst, G. W., & Newell, A. (1969). *GPS: A case study in generality in problem solving.* New York: Academic Press.

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. Einstein, & M. McDaniel (Eds.), *Prospective memory: Theory and applications*. Hillsdale, NJ: Erlbaum.

Lahlou, S., Kirsh, D., Rebotier, T., Reeves, C., & Remy, M. Interruptions in the workplace. *Available: http://adrenaline.ucsd.edu/edf/Experiment1.htm.*

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.

Miyata, Y., & Norman, D. (1986). Psychological issues in support of multiple activities. In D. A. Norman and S. W. Draper (Eds.), *User Centred Systems Design: New Perspectives on Human-Computer Interaction*. Hillsdale: Lawrence Erlbaum Associates.

Monk, C., Boehm-Davis, D., & Trafton, J. G. (2002). The attentional costs of interrupting task performance at various stages. In the *Proceedings of the 2002 Human Factors and Ergonomics Society Annual Meeting.* Santa Monica, CA: HFES.

Rogers, R. D., & Monsell, S. (1995). The cost of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124,* 207-231.

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, 50A,* 49-78.