Are Retrievals from Long-Term Memory Interruptible?

Michael D. Byrne
byrne@acm.org
Department of Psychology
Rice University
Houston, TX 77251

Abstract

Many simple performance parameters about human memory are not well-understood. One such parameter is how the cognitive system handles interruption at a relatively low level. This research is an attempt to determine if simple, well-practiced retrievals from long-term memory can be interrupted by a higher-priority task. An experimental paradigm referred to as a "reverse PRP" paradigm is introduced, and the results of one experiment in this paradigm reported. The results suggest that retrievals can indeed be interrupted, but that there is an interruption cost.

Introduction

There are numerous situations in which people are interrupted in doing simple tasks by higher priority tasks and must drop what they are working on the new task. In most situations, this is merely an inconvenience. However, in high-performance tasks such as air traffic control, even a small delay in responding to the interrupting task can have more serious consequences. In many cases, the interruption may place demands on perceptual or motor performance, but in other cases it is a cognitive operation that is interrupted. Generally speaking, cognitive theories have little to say about what should happen in such situations. However, this does not mean that these phenomena cannot be understood in the context of, and do not have implications for, theories of cognition.

ACT-R/PM (Byrne & Anderson, 1998) provides a set of perceptual-motor extensions to the ACT-R cognitive architecture (Anderson & Lebiere, 1998). Communication between central cognition (the ACT-R production system) and the perceptual-motor modules takes two forms: [1] the left-hand, or THEN, side of productions can request activity from the perceptual-motor modules (e.g. shift visual attention, press a key), and [2] perceptual-motor modules deliver results (e.g. representations of percepts) to ACT-R's declarative memory in the form of chunks.

Declarative memory chunks in ACT-R are accessed via retrieval, which is a time-consuming process. That is, retrievals take time, which is part of the process of matching the IF side of productions in ACT-R. Because perceptual-motor modules operate in parallel with the production system, it is possible for one or more of the perceptual-motor modules to change the contents of declarative memory while a retrieval is in progress. The fundamental question this research is attempting to address is what happens in this situation: Do retrievals always complete or can they be interrupted? Rather than attempt to answer this question on theoretical or computational grounds, this research approaches this as an empirical question.

Reverse PRP Paradigm

Consider this simple dual task: two digits appear on a display, and the product of those digit should be spoken aloud. On some trials, the digits are replaced a short time after they appear by a colored block. When the block appears, the task is to make a choice response based on the color of the block as rapidly as possible. Because the delay is short, the appearance of the color block may be interrupting the retrieval of the product of the two digits. Can the single, simple retrieval be interrupted?

This task shares a number of important properties with the psychological refractory period (PRP) paradigm, which is perhaps the simplest dual-task experimental paradigm. The PRP has a long history in psychology (see Pashler, 1994 for a review). In this paradigm, subjects are asked to do two tasks, usually referred to as T1 and T2, in rapid succession. The stimulus for T1 appears, then after some delay (called the stimulus onset asynchronoy or SOA), the stimulus for T2 appears. Subjects are instructed to give T1 maximum priority and the typical results are that responses to T2 are slowed,

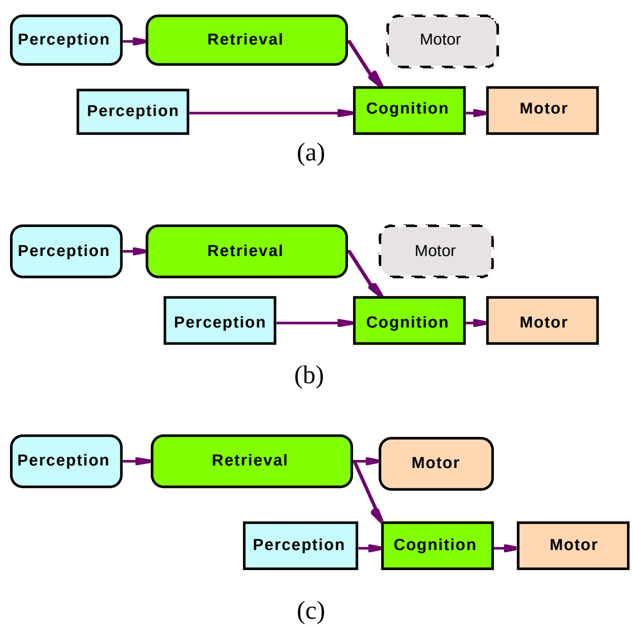
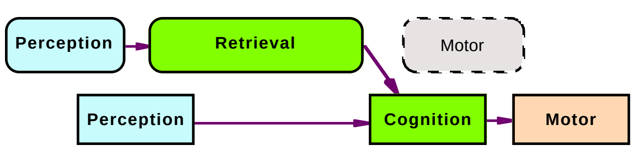


Figure 1. Predictions of the ballistic retrieval hypothesis

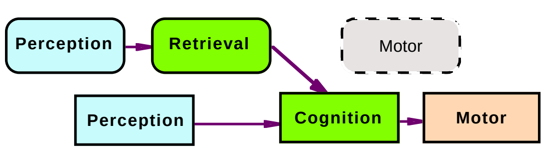
and more so at shorter SOAs. Results of such experiments have been taken as evidence that central cognition is effectively serial (again, see Pashler 1994 for a review).

The basic experimental paradigm used in this research inverts the priority instruction given to the subjects. That is, subjects are instructed to give T2 maximum priority; when the T2 stimulus onsets, subjects are to immediately give that stimulus highest priority. If T1 involves retrieval from declarative memory, the interruptibility of of that retrieval will have a large impact on response time for T2. If the T1 retrieval is not interruptible (this will be termed “ballistic”), then, assuming serial cognition, cognitive processing of the T2 stimulus will be forced to wait for the completion of the retrieval and will thus be slowed. In particular, it should be slowed more at shorter SOAs. This situation is depicted in Figure 1. In Figure 1 and the following figures, time moves from left to right, and each stage of processing is represented by a box. Arrows represent dependencies. T1 stages are the upper set of boxes, T2 stages the lower set.

Panel (a) of Figure 1 shows the situation at short SOAs, which will result in a long T2 response time. Cognition for T2 must wait for the T1 retrieval to complete, which causes an elevated T2 response time. As SOA increases, T2 response time should decrease (Figure 1, panel b) until at long enough SOAs T2 should no longer be slowed at all (Figure 1, panel c). The slope of T2 response time as a function of SOA should thus be -1 until the “long enough” SOA is reached and the slope drops to zero. At this point, the response time for T2 should be the same as when T2 is not an interrupting task, that is, the single task time.



(a)



(b)

Figure 2. Difficulty effect under the ballistic retrieval hypothesis

A secondary prediction made by the ballistic retrieval hypothesis is that the duration of the T1 retrieval should directly impact the T2 response time. If processing for T2 must wait for the completion of T1 retrieval, extending the duration of that retrieval (e.g. by making the retrieval more difficult) should directly impact T2 response time. If processing for T2 must wait for the completion of the T1 retrieval, extending the duration of that retrieval should result in a time cost for T2 identical in size to the increase in retrieval difficulty . This is depicted in Figure 2: panel (a) depicts a short T1 retrieval, panel (b) depicts a long T1 retrieval.

If, on the other hand, retrievals are interruptible, T2 response should be insensitive to the state of the T1 retrieval. That is, there should be no effect of either SOA or T1 difficulty. This situation is depicted in Figure 3.

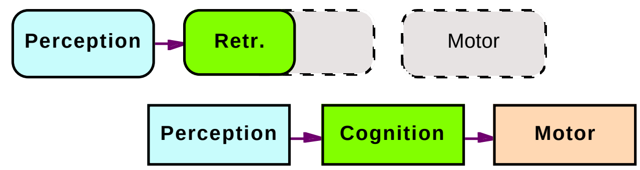


Figure 3. Interruptible retrieval hypothesis

There is a potential complication, which is interruption cost. The shift from T1 to T2 may have a cognitive cost. If such a cost exists, and it is fixed, then the T2 response time in the interruption situation should be elevated when compared to the T2 response time when T2 is performed in isolation (the single-task case). This should hold regardless of T1 difficulty or SOA. Figure 4 represents the situation in which retrievals are interruptible but with an interruption cost.

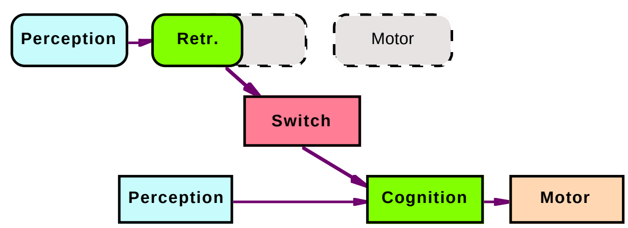


Figure 4. Interruptible retrieval with switch cost

To summarize, the ballistic retrieval hypothesis predicts that T2 response time should have a -1 slope as a function of SOA, and that adding difficulty to retrieval should generate a parallel function of SOA, with the distance between the RT functions equivalent to the single-task difficulty effect of T1. On the other hand, interruptible retrieval hypothesis predicts that T2 should be insensitive to either SOA or T1 retrieval difficulty.

Methods

Participants

The participants were 39 Rice University undergraduates who participated for credit in a psychology class.

Stimuli and Design

There were three kinds of trials: multiplication only, color identification, and interruption. Single-digit multiplication was used as T1 in this paradigm. Participants saw two digits presented visually (e.g. “6 8”) and responded with the product of the two digits vocally (e.g. “forty-eight”). Retrieval difficulty was manipulated by varying the size of the digits used. This manipulation has been shown to be effective in previous work (Byrne & Anderson, 1999). “Easy” retrieval used the digits from 1 to 4, while “hard” retrieval used the digits 6 through 9. Squares (e.g. “7 7”) were not used.

A simple color identification task served as T2 in this paradigm. This was a choice reaction time task with two alternatives. A rectangular block of color appeared on the display. If the color block was blue, participants pressed one key on the keyboard; if the block was red, another key was pressed.

For interruption trials, the color block appeared and covered the digits on the screen. The SOA was the time between the onset of the digits and the onset of the color block, measured in milliseconds. SOAs of 200, 375, 550, and 725 ms were used. Participants were instructed that when an interruption occurred, they were to respond to T2 as rapidly as possible and that completion of T1 was not necessary. These instructions were given to maximize the priority given to T2; participants should have no reason to continue with T1 and thus should switch to T2 as rapidly as possible.

The design was also blocked, each block consisted of five sets of 40 trials. One set in each block consisted of only color identification trials, to provide an estimate of single-task response time. The remaining four sets were a mixture of multiplication-only trials and interruption trials, with interruptions occurring 20% of the time. Thus, for interruption trials, there were three factors, all within-subjects: block, from one to three, four levels of SOA, and two levels of difficulty. Which trials contained interruptions and the order of sets within a block were randomized.

Procedures

Participants were first trained on the color identification task until they performed two consecutive sets of 40 trials with 95% or better accuracy. Participants were then given 40 trials of practice with multiplication-only trials, followed by a 40-trial set of multiplication trials, 20% of which contained interruptions.

Apparatus

Stimulus presentation and data collection were done on Apple iMac personal computers. Vocal responses were timed with an Apple PlainTalk microphone by monitoring the microphone level and stopping the timer when a threshold level of input was exceeded. Keypress responses were timed by actively polling the state of the keyboard. Both measures should be accurate to approximately 5 ms.

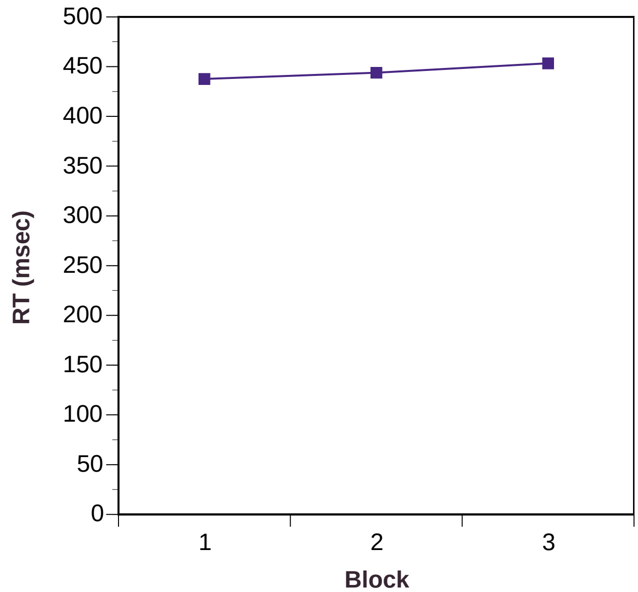


Figure 5. Color identification response time as a function of block

Results

Due to the excellent power of the repeated-measures design and the large number of subjects and trials, an alpha level of 0.01 will be used for all statistical tests.

The color identification task is fairly simple and participants were forced to practice to a relatively stringent criterion, so performance was expected to be rapid but there was still the possibility that subjects may have been speeding up with practice. Figure 5 presents single-task color identification response time as a function of block. Clearly, there was no practice-related speedup in this case, in fact, the absolute response times actually went down slightly with practice, though this is probably coincidental. Overall, the effect of block was not reliable, $F(2, 70) = 1.83$, $p = 0.17$. The lack of learning on this task suggests that performance on this task is limited primarily by fixed architectural properties such as perceptual-motor limitations; the cognitive demands of this task are fairly minimal.

Multiplication-only trials demonstrated a much more complex pattern. Response time for multiplication-only trials is shown in Figure 6. As expected, there was an effect of difficulty, $F(1, 35) = 81.74$, $p < 0.001$ with hard problems clearly slower than easy ones, on average, about 350 ms slower. There was also a main effect of block, $F(2, 70) = 10.30$, $p < 0.001$,[1] and a block by difficulty interaction, $F(2, 70) = 12.14$, $p < 0.001$, both primarily a function of improvement on hard problems. If retrievals are ballistic, all of these effects should show up in T2 response time in the interruption trials, since T2 cognition should be forced to wait for the completion of the retrieval.

[1] To control for sphericity problems, repeated-measures factors with more than two levels were adjusted with either Huynh- Feldt epsilon or Greenhouse- Geisser epsilon where appropriate.

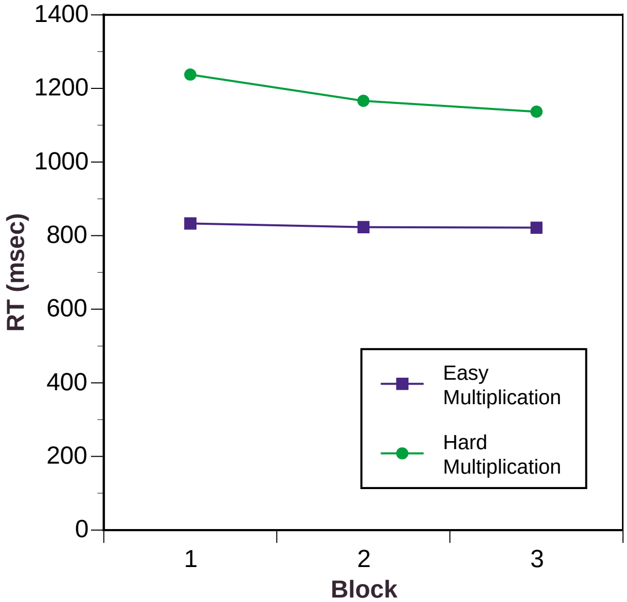


Figure 6. Multiplication-only response time for easy and hard problems as a function of block

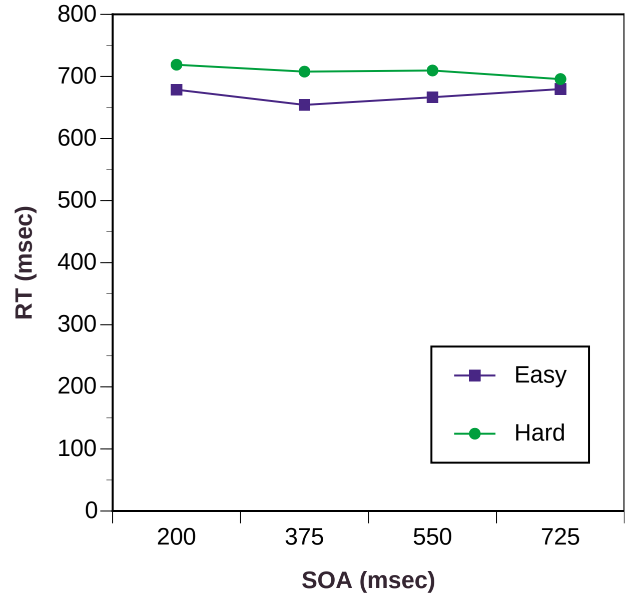


Figure 7. T2 interruption response time as a function of SOA for hard and easy T1 retrievals

The interruptible retrieval hypothesis, given the lack of practice effects on color identification in isolation, should predict no effect of difficulty or block on T2 interruption performance.

The data of primary interest, of course, are the data for the interruption trials. These data, as a function of SOA, are presented in Figure 7. The results are generally consistent with the interruptibility hypothesis. Most importantly, there was no effect of SOA, $F(3, 105) = 0.90$, $p = 0.40$. There is clearly a potential problem of accepting the null hypothesis here. However, the prediction made by the ballistic hypothesis is specific: there should be a -1 slope with SOA. This can be tested with a linear contrast on SOA, which was not reliable, $t(35) = -0.51$, $p = 0.61$. A -1 slope would be a large effect in this context, and power to detect a large effect in this situation was estimated to be 0.99 (see Cohen, 1988 for details on this procedure). Thus, accepting the null hypothesis in this case is statistically justifiable.

All other effects and interactions were also not reliable, save one: the effect of T1 difficulty (the difference between the easy and hard conditions) was reliable, $F(1, 35) = 8.29$, $p = 0.007$. The absolute magnitude of this difference is small, however, at just under 40 ms. The two difficulty effects, one in multiplication-only trials, and one in interruption trials, is presented for each block in Figure 8. These effects are obviously different, and indeed a repeated-measures ANOVA on the difficulty reveals a very reliable effect of multiplication-only vs. interruption, $F(1, 35) = 65.46$, $p < 0.001$. This suggests that while the difficulty effect did manifest itself in the T2 response time, this effect is probably not due to retrieval difficulty in T1, since that difficulty effect was roughly nine times larger.

There was also a reliable effect of block, $F(2, 70) = 6.73$, $p = 0.004$, and an interaction, $F(2, 70) = 10.68$, $p < 0.001$ on the difficulty effect. This seems to be driven primarily by the previously-mentioned improvement in "hard" multiplication problems over time, which results in a reduction in difficulty effect for the multiplication-only trials; in contrast, the small

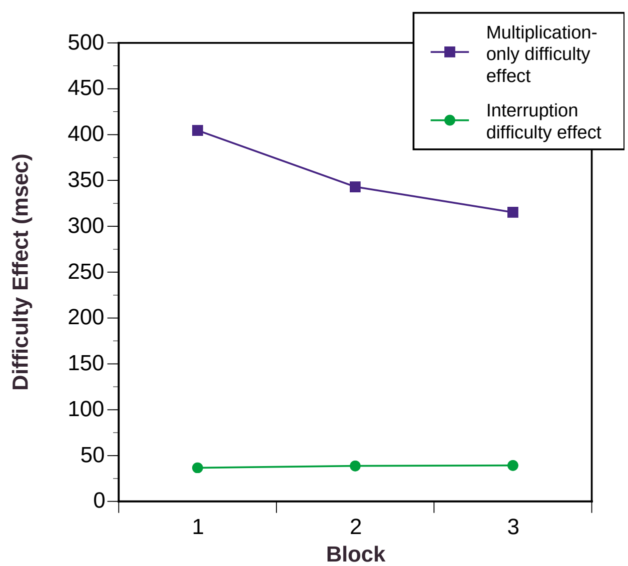


Figure 8. Difficulty effect as a function of block for multiplication-only trials and interruption trials

difficulty effect seen in T2 interruptions is fairly stable over blocks.

Of course, the lack of SOA effect may be due to subjects adopting a strategy of delaying response to T1 until they could be confident that an interruption would not occur. The multiplication-only response times are fairly rapid, suggesting this is unlikely, but there is a more direct test. Subjects often responded to T1 even when the interruption occurred, but they did so more often for long SOAs than for short SOAs and more often for easy problems than hard problems. This is shown in Figure 9. Effects of block, SOA, and their interaction were reliable, [for SOA, $F(3, 105) = 108.70$, $p < 0.001$; for difficulty, $F(1, 35) = 152.24$, $p < 0.001$; for the interaction, $F(3, 105) = 11.42$, $p < 0.001$] but there were no reliable effects or interactions involving block. This sensitivity to SOA and difficulty suggests that participants did indeed attempt to respond as rapidly as possible to T1 and did not uniformly postpone T1 in anticipation of an interruption.

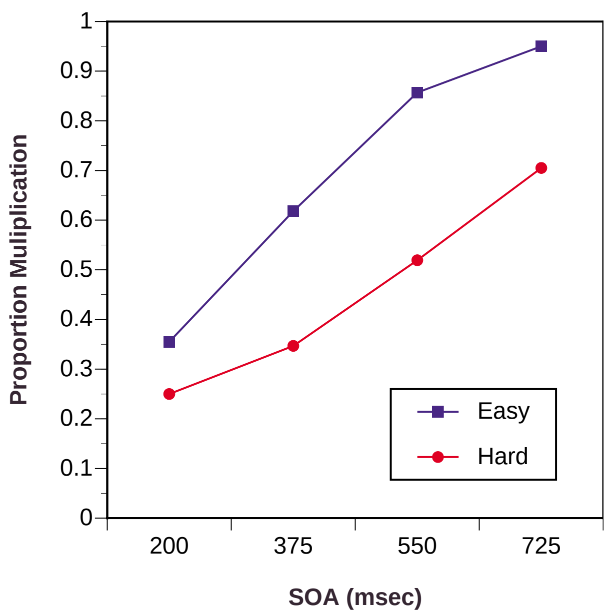


Figure 9. Proportion of interruption trials on which a T1 response was emitted

The final effect to consider is the interruption cost, that is, the difference between color identification response time when it was in isolation vs. when it was the interrupting task. Figure 10 presents the results. Clearly, there was an interruption cost, $F(1, 35) = 235.58$, $p < 0.001$. The absolute magnitude of this difference is large relative to the single-task color identification response time. Single-task response time for color identification averaged just under 450 ms, but with interruptions it was close to 700 ms, a 250 ms penalty.

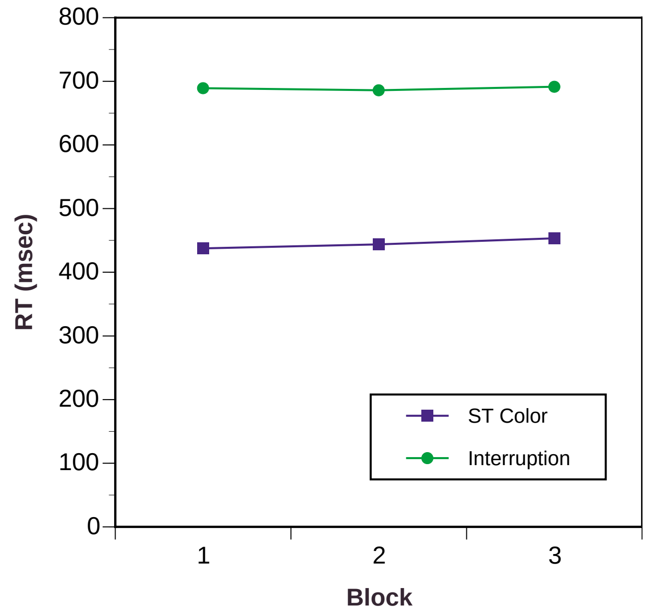


Figure 10. Color identification response time as a function of block when it was in isolation (single-task or ST) vs. as the interrupting T2

There is no real evidence that this cost was reduced with practice as there was no reliable effect of block, $F(2, 70) = 0.80$, $p = 0.45$, or an interaction of task condition and block, $F(2, 70) = 0.81$, $p = 0.45$.

Discussion

These data are clearly more consistent with the interruptibility hypothesis. The lack of SOA effect on T2 response time is most telling. However, the data are not entirely equivocal. There was a reliable effect of T1 retrieval difficulty on T2 response time, though this effect was small and clearly of a different magnitude than the difficulty effect present in T1. The source of this effect is unclear. One possible explanation is that perceptual processing of T1 is more difficult for larger digits but this is purely speculative.

For the purposes of setting architectural policy in ACT-R/PM, these results certainly suggest that retrievals should be interruptible. However, whether retrievals should always be interrupted by any change in declarative memory or whether they should only be interrupted under certain conditions is unclear. In this experiment, the retrieval is interrupted by a higher-priority change that is both presented foveally and displaces the T1 stimulus in the visual array. These conditions at least appear to favor interruption. The frequency of interruption in this experiment, 20%, may also play a role.

At a more general level, the interruption cost itself is quite intriguing. The source of this cost is not clear, though something of its nature was revealed; it appears not to change with practice (blocks) and appears not to be affected by SOA. Whether this cost is sensitive to factors such as interruption frequency, modality match with the T1 stimulus, and T2 difficulty, is unknown. Follow-up research certainly appears appropriate.

However, in some sense, the change from T2 to T1 processing can be thought of as a task-switch (e.g. Rogers & Monsell, 1995). While a great deal is known about task-switching (Altmann & Gray, 1999 provides an excellent account), it is not clear whether or not this is a special case of task-switching phenomenon. In traditional task-switching experiments, one type of task follows the completion of another, but the two tasks do not temporally overlap, that is, one does not interrupt the other. The ramifications of this difference in experimental paradigm are not entirely clear; the interruption cost may be related to the cost associated with task-switching or it may be an independent effect. Again, further research will be required to better understand the interruption cost.

Acknowledgements

I would like to thank Michael Fleetwood and Bryan Blauvelt for their assistance in data collection, and to Erik Altmann for comments on an earlier draft.

References

Altmann, E. M, & Gray, W. D. (1999). The anatomy of serial attention: An integrated model of set shifting and maintenance. Manuscript submitted for publication.

Anderson, J. R., & Lebiere, C. (Eds.). (1998). *The atomic components of thought.* Hillsdale, NJ: Erlbaum.

Byrne, M. D. & Anderson, J. R. (1998). Perception and Action. In J. R. Anderson & C. Lebiere (Eds.) *The atomic components of thought.* Mahwah, NJ: Lawrence Erlbaum.

Byrne, M. D., & Anderson, J. R. (1999). Serial modules in parallel: The psychological refractory period and perfect time-sharing. Manuscript submitted for publication.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220-244.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124(2),* 207-231.