Toward a Unified Theory of the Multitasking Continuum: From Concurrent Performance to Task Switching, Interruption, and Resumption

Dario D. Salvucci [1]
[1] Department of Computer Science
Drexel University
3141 Chestnut St.
Philadelphia, PA 19104, USA
salvucci@cs.drexel.edu

Niels A. Taatgen [2,3]
[2] Department of Psychology
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213, USA
taatgen@cmu.edu

Jelmer P. Borst [3]
[3] Dept. of Artificial Intelligence
University of Groningen
Postbus 407
9700 AK Groningen, the Netherlands
jpborst@ai.rug.nl

ABSTRACT
Multitasking in user behavior can be represented along a continuum in terms of the time spent on one task before switching to another. In this paper, we present a theory of behavior along the multitasking continuum, from concurrent tasks with rapid switching to sequential tasks with longer time between switching. Our theory unifies several theoretical effects — the ACT-R cognitive architecture, the threaded cognition theory of concurrent multitasking, and the memory-for-goals theory of interruption and resumption — to better understand and predict multitasking behavior. We outline the theory and discuss how it accounts for numerous phenomena in the recent empirical literature.

Author Keywords
Multitasking, attention, interruption, cognitive architecture.

ACM Classification Keywords
H.5.2 [Information Interfaces and Presentation] User Interfaces – theory and methods; H.1.2 [Models and Principles] User/Machine Systems – human factors, human information processing.

INTRODUCTION
The modern world is a multitasking world — in all kinds of environments and scenarios, people spend a great deal of time engaged in multiple tasks at the same time. For example, a recent study [16] found that employees of an information-technology company spent an average of only 3 minutes per task before switching to another task. In addition, user interfaces have rapidly spread from standard desktop settings into real-world multitasking environments due to the proliferation of mobile computing devices (e.g., the use of cell phones while walking or driving). All together, these trends necessitate a better understanding of the capacity and the limitations of human multitasking performance: Only with such knowledge can we design and build new user interfaces that complement, rather than distract from or hamper, our daily lives.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CHI 2009, April 4–9, 2009, Boston, MA, USA.
Copyright 2009 ACM 978-1-60558-246-7/09/04…$5.00

The Multitasking Continuum

One rough but useful way of characterizing multitasking behavior is in terms of the time spent on one task before switching to another. This span, which we call the *multitasking continuum*, is shown in Figure 1 along with sample task domains at their approximate position along the continuum. For example, on the left-hand side of the continuum, talking while driving involves frequent switching between tasks — with switches perhaps every second, if not more often in normal conversation. On the right-hand side, cooking while reading a book may involve fairly long spans between task switches; for instance, one might start boiling pasta, read for 10 minutes until the pasta is cooked, then strain and prepare the pasta for a meal. Multitasking behavior thus spans time scales at several orders of magnitude, namely, as termed by Newell [31], the levels corresponding to the Cognitive Band (10^{-1}–10^{1} s) and Rational Band (10^{2}–10^{4} s) of human behavior.

The tasks on the left-hand side of the continuum could be characterized as *concurrent multitasking*, in which the tasks are, in essence, performed at the same time. There has been a long and detailed research literature on concurrent multitasking dating back at least to the 1930s [e.g., 42]. Some of the earlier work, which continues today, examines concurrent performance of simple stimulus-response tasks (e.g., in the dual-choice paradigm [32]). At the same time, research has explored concurrent performance in a wide variety of real-world tasks, from piloting [23] to driving [38] to radar operation [43]. This empirical work has been accompanied by theoretical and computational models of concurrent multitasking [e.g., 24, 34, 36] that have aimed to explain empirical phenomena with unifying ideas and frameworks.

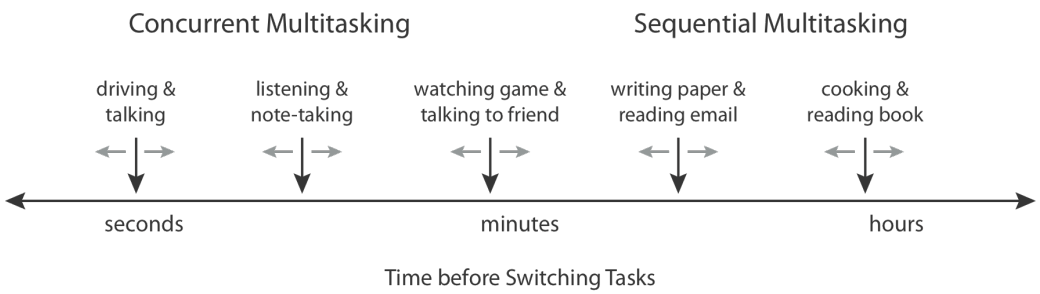

Figure 1: The Multitasking Continuum.

At the same time, the tasks on the right-hand side of the continuum could be characterized as *sequential multitasking*, in which a longer time (say, minutes to hours) might be spent on one task before switching to another. As for concurrent multitasking, there has been a rich history of literature on what we are calling sequential multitasking, broadly in the areas of task switching, interruption, and resumption. Basic psychological research [see 30] has primarily focused on the "switch cost" (in time) encountered when switching between tasks (using tasks of short duration but enforcing sequential multitasking through the experimental paradigm itself). More complex studies have examined analogous measures for real-world human-computer interaction tasks [e.g., 1, 9, 12, 13, 22]. Again, like for concurrent multitasking, researchers have developed conceptual and computational models [e.g., 2, 14, 28] that attempt to make explicit the sources of these switch costs and how they might be mitigated in different scenarios and environments.

Considering that concurrent and sequential multitasking represent different ranges on the same continuum, there has been surprisingly little cross-fertilization between the research in the two areas. In part, this separation between the two areas has evolved for a good reason: Each area has focused on distinct characteristics of behavior that are interesting and warrant detailed study in their own right. Nevertheless, because of their overlap on the multitasking continuum, we strive for a unified theory of human multitasking that is able to account for both concurrent and sequential multitasking — that is, provide an account of human behavior at all points along the multitasking continuum.

Toward a Unified Theory of Multitasking

Our goal in this paper is to outline a unified theory of human multitasking that spans both concurrent and sequential multitasking. The theory incorporates three core components: the *ACT-R cognitive architecture* [5, 6], which provides a theory and computational framework for human processing resources and their limitations; *threaded cognition theory* [36], which provides an account of concurrent performance for two or more arbitrary tasks; and *memory-for-goals theory* [2], which provides an account of task interruption and resumption based on activation and recall of task goals in declarative memory. We use these components as the building blocks of our theory, showing how each contributes to the larger view of multitasking and how the unification of these components helps to account for a wide range of behavior across the multitasking continuum.

BEGINNINGS: CONCURRENT MULTITASKING

We begin by examining concurrent multitasking, which we define as the execution of two or more tasks at the same time. Our account of concurrent multitasking employs the two component theories of ACT-R and threaded cognition. We now briefly describe this theory, and then outline how the theory accounts for basic phenomena related to concurrent multitasking for both simple laboratory and complex real-world tasks.

A Theory of Concurrent Multitasking

The ACT-R architecture [5] holds as a central assumption that the best way to understand cognition at the functional level is to consider it as a set of relatively independent but interacting modules (or resources [44]). Some of the modules deal with perception and motor activities, like vision, audition, manual control, and speech. More characteristic for ACT-R, though, is a set of central cognitive modules. The *declarative memory module* serves as a memory for factual knowledge, which also includes episodic knowledge, and task instructions. The *goal module* represents the current goal of the system, and can keep track of progress and other state information. The *problem representation module* (sometimes called the *imaginal module*) holds partial representations needed by the task — for example, an intermediate expression during algebraic equation solving. Finally, the *procedural module* connects all other modules together, using knowledge in the form of condition-action production rules to control the flow of information among modules. These production rules match the state of information represented in other modules and map these onto actions that are to be executed by the various modules. For example, the visual module might read "1+2," after which the procedural module queries the declarative memory module for the 1+2 addition fact. When the declarative module produces the answer, the procedural module would send this answer to the manual module, which would then type "3" on the keyboard.

The key aspect of ACT-R for multitasking is that all modules can operate in parallel, but each module can serve only one task at a time. Threaded cognition theory [36] builds on the ACT-R architecture by allowing multiple active tasks, or *threads*, to execute simultaneously across the ACT-R processing modules. The different threads share available modules through a greedy/polite policy: If a module is not in use, any of the threads can claim it (greed); however, once a thread is done with a module, it releases it immediately (politeness). These policies emerge in a straightforward way from the module/buffer structure of the ACT-R resources. The periods of time that threads use a particular module are usually relatively short, typically less than a second, or only 50 ms in the case of the procedural module — the time needed to fire a single production rule.

The theory posits that each of the modules can potentially act as a source of interference in multitasking. For our purposes here, it is critical to note three sources of interference that affect multitasking performance. First, the declarative module can lead to interference when two tasks both use memory for standard memory retrievals or for retrievals of task instructions [39]. Second, the problem representation module can lead to interference when two or more tasks require complex representations that must be swapped out for each task switch. Third, the procedural module, as the central cognitive bottleneck of the system, can lead to interference when two tasks attempt to start using the same module at the same time.

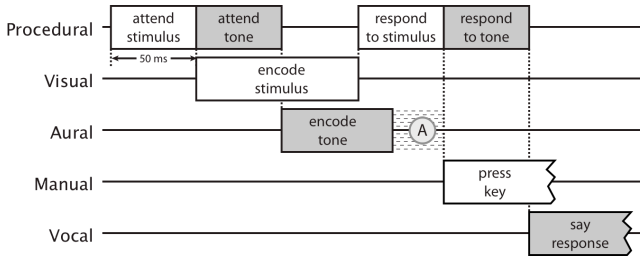

Figure 2: Module processing in a dual-choice paradigm: visual stimulus with manual response (white boxes) and aural stimulus with vocal response (gray boxes).

As an example, Figure 2 shows the expected module processing in a simple dual-choice paradigm: In the visual-manual task (white boxes), the user responds to a circle on the left, center, or right by pressing an associated finger; in the aural-vocal task (gray boxes), the user responds to a tonal pitch by saying "one"-"three." We can see in this depiction how processing is spread across modules, allowing both task threads to interleave with each other. We also see how, as typical in ACT-R processing, each step comprises a procedural step (production-rule firing) of 50 ms and subsequent processing in another module whose timing depends on the nature of the process (e.g., visual encoding versus motor response). In this case, there are no perceptual-motor module conflicts; however, there is a procedural conflict at point A in the figure: The first task uses the procedural module to initiate its response, and although the second task's encoding finishes, the second task cannot initiate a response until the procedural module is free. Such procedural interference can cause small but observable effects in high-performance tasks [e.g., 40]. In situations for which there is no such procedural interference, the theory predicts perfect time-sharing, a phenomenon that has been observed in the empirical literature [e.g., 37].

Accounts of Concurrent Multitasking Phenomena

Threaded cognition combined with ACT-R has been shown to account for concurrent multitasking behavior in a variety of domains, including:

- *Dual Choice*. Cognitive psychologists have long been interested in the dual-choice paradigm [see 32] in which people perform two simple choice-reaction tasks simultaneously or with a slight delay between the choice tasks. The theory has accounted for a number of results in the empirical literature for this paradigm, most notably the effects of differing perceptual and response modalities and the learning effects that occur over time [36].
- *Pursuit Tracking*. Another basic task is that of pursuit tracking [27], in which a user moves a cursor to track a moving target. Like previous efforts in the CHI literature [e.g., 25], the theory has accounted for the effects of impaired tracking performance during execution of a concurrent secondary choice task [36].
- *Driving and Driver Distraction*. A more realistic domain is that of driver distraction, where a driver performs a secondary task (e.g., dialing a cell phone) while controlling a vehicle. The theory has accounted for a range of secondary-task effects on driver performance [33, 36], such as differing effects of input modalities on both steering and braking performance.

Because theoretical accounts of these domains have been published elsewhere as cited above, we do not dwell on the details of these accounts. Instead, we are most interested in integrating the theory of concurrent multitasking with other recent efforts and generalizing the theory to sequential multitasking, as described in the next section.

GENERALIZING TO SEQUENTIAL MULTITASKING

Given this theory of concurrent multitasking, we now build on and extend this theory to account for behavior in sequential multitasking. We rely on two additional components for our theory, namely the memory-for-goals theory as well as recent work on problem representation in the context of ACT-R and threaded cognition. As for concurrent multitasking, we begin with a description of our theory and continue with a description of how this theory accounts for various empirical phenomena found in the literature.

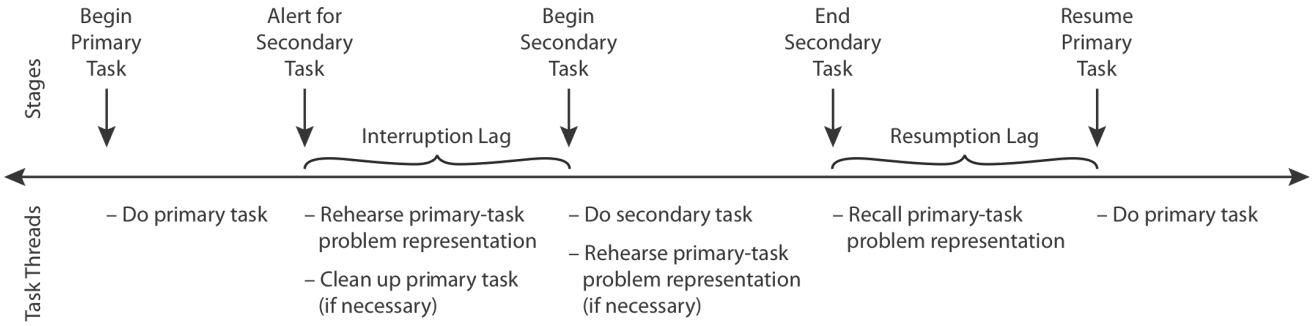

Figure 3: The stages of interruption and resumption (adapted from [43]) and the task threads associated with each stage.

A Threaded Theory of Sequential Multitasking

Altmann and Trafton [2] have posited a theory of "memory-for-goals" that represents task goals in terms of general memory mechanisms of activation and associative priming. The core idea of the theory is that when people initiate a new task goal, the goal must be strengthened in memory to the point where its activation rises above old goals, making it the primary source of goal-directed attention. At the same time, these goal memories decay according to the memory mechanisms incorporated into ACT-R; thus, an interrupted task goal that has fallen out of active use requires more time to recall and resume. This idea seems to be inconsistent with threaded cognition, which assumes that goals can be active in parallel and do not suffer from decay.

Recent work by Borst and Taatgen [10, 11] sheds some light on how threaded cognition and memory-for-goals can be reconciled. A new development in the ACT-R theory is a separation of the goal into a control representation and a problem representation [4]. The goal concept in threaded cognition refers to the control representation, which directs the procedural module to production rules relevant to the current goal. In contrast, the goal concept in memory-for-goals refers to the problem representation that stores working information relevant to the current goal. Borst and Taatgen investigated whether this distinction holds under multitasking conditions, and found that the need for multiple problem representations thus causes interference in multitasking in the form of slower performance and a higher rate of errors [10, 11]. In essence, if multiple tasks require problem representations, they cannot be maintained at the same time, and have to be swapped out when there is a switch between tasks. The old problem representation will be stored in declarative memory, from where it can be retrieved later on when switching back to the first task. Retrieving a representation from memory and restoring it takes time and is not always successful; errors can be prevented by rehearsing the problem representation, but this will slow down performance even more.

The memory-for-goals theory, along with the idea of distinct problem representations, meshes nicely with threaded cognition in providing a fuller theory of sequential multitasking where various task threads are active during the stages of interruption and resumption. Figure 3 shows the stages of interruption (derived from [43]) along with the task threads corresponding to each stage. During performance of the primary task, the primary task is the only task thread. At the time when an alert interrupts the primary task, users initiate two distinct threads: The first begins rehearsing the primary-task problem representation to strengthen its memory activation, and the second "cleans up" the primary task if necessary before interruption (e.g., while a telephone is ringing, the user may finish the sentence being typed before answering). During the secondary task, the user is performing the secondary task as well as continuing to rehearse the problem representation, if necessary (i.e., if rehearsal during the interruption lag was insufficient). Finally, after completion of the secondary task, the user recalls the primary-task problem representation during the resumption lag and finally resumes the primary task.

A Computational Model of Interruption and Resumption

We have developed an ACT-R model to demonstrate how the processing outlined in Figure 3 can be realized in a computational cognitive architecture. The model includes procedural rules that implement the processes of problem-representation creation, rehearsal before and during task interruption, and retrieval during task resumption. It also includes two generic models of primary- and secondary-task behavior that simply fire a single rule repeatedly, simulating at a high level the procedural usage of these tasks. While we plan to implement further models of specific tasks in future work, these generic models allow us to explore the interruption-relevant processes without committing to specific primary and secondary tasks, and to focus on breadth of theoretical coverage over a wide range of empirical phenomena.

The model operates as follows. After performing the primary task for a short time, the model sees a visual warning pop up on the screen (as in [43]), subsequently stops execution of the primary task and begins mentally rehearsing the problem representation. In ACT-R, the rehearsal process simply comprises alternating use of the procedural module to initiate a memory retrieval and the declarative memory module to perform the retrieval, as

illustrated in Figure 4.[1] When the secondary (interrupting) task appears on-screen, the model begins performing this task and can continue rehearsal if desired (to be discussed shortly). Finally, when the secondary task ends and the primary task reappears, the model retrieves the original primary-task problem representation and then resumes performing the primary task.

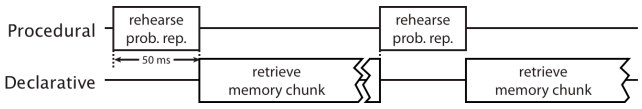

Figure 4: Module processing for the rehearsal thread, comprising repeated retrievals of the problem representation.

It is important to emphasize the relationship here between the memory-for-goals and threaded cognition theories. In a sense, threaded cognition facilitates the lower-level cognitive routines necessary in realizing the memory-based processes postulated by memory-for-goals. Specifically, memory-for-goals postulates that memory strengthening and retrieval is the critical process for interruption and resumption, and thus implies that there are task threads that manage this process. During both stages for which this thread is active (the interruption-lag stage and the secondary-task stage), the rehearsal thread is interleaved with execution of another thread. Threaded cognition thus helps to account for how the rehearsal thread and any other existing threads balance processing across modules, and how and when these threads may experience module conflicts that result in effects on multitasking performance.

Accounts of Sequential Multitasking Phenomena

As we did for concurrent multitasking, we now review how our overall theory can account for a variety of empirical phenomena reported for sequential multitasking, specifically the literature concerning the interruption and resumption of user tasks. As mentioned, our treatment here emphasizes breadth of coverage over a wider range of reported empirical results using conceptual explanations as well as illustrative simulation results from the developed ACT-R computational model.

Effects of Interruption Lag

One important aspect of task interruption involves the interruption lag and its impact on the subsequent resumption lag. As mentioned, the interruption lag represents the time between an alert (or warning) of an imminent interruption and the actual start of the interrupting task, whereas the resumption lag represents the time between the end of the interrupting task and the re-initiation of the original task (typically measured in terms of some task action). Trafton et al. [43] compared the behavior of users in a "Warning" condition, including an interruption lag of 8 seconds, with that of users in an "Immediate" condition with no interruption lag. They found that users in the Warning condition resumed the original task significantly faster than those in the Immediate condition — that is, the time period of interruption lag allowed users to speed up the resumption of the original task following the interruption.

A prediction of user behavior in the Warning condition arises directly from memory-for-goals and problem representations: Given time before the interrupting task, a user can strengthen the problem-representation memory and thus, upon resumption, more easily recall it and resume the task. The Immediate condition, though, requires more thought. Under one interpretation of the memory-for-goals theory, one might expect that users never rehearse the problem representation in that condition because there is no interruption lag during which to rehearse. However, Trafton et al. found instances of verbal utterances representing rehearsal concurrent with the interrupting task: "Preparation was deemed important enough to be attempted concurrently with the secondary task" (p. 593). In such a case, threaded cognition would allow rehearsal (like these utterances) to be performed concurrently with the interrupting task, interleaved between cognitive steps related to the interrupting task.

To illustrate the predictions of our theory, we ran the ACT-R model in the same conditions as those in the Trafton et al. study (but with generic task models as described earlier). For both the Warning and Immediate conditions, we ran two simulations, one of which rehearsed the problem representation only during the warning (interruption lag) and the other which rehearsed during both the warning and the interruption. The time needed to retrieve the primary-task problem representation after interruption is shown in Figure 5. When the model rehearsed only during the interruption lag, it exhibited large differences in retrieval time: The Warning condition allowed the model to rehearse for 8 seconds and thus retrieval was reasonably fast (116 ms), but the Immediate condition allowed for no rehearsal and thus retrieval failed after 1 second (thereby forcing a re-encoding of the problem representation by re-analyzing the screen). This first model matches best to Trafton et al.'s results in the early stages of practice, where they found large differences in resumption lag between conditions. When the model rehearsed during the interruption itself, retrievals were fast in both conditions (approximately 20 ms). This second model matches best to Trafton et al.'s results in the later stages of practice, where both conditions show roughly similar performance. We discuss these issues further in a later section on changing effects with learning over time.

[1] An alternative model could use subvocalization for rehearsal; while we focus on declarative memory rehearsal here, it is possible that subvocal rehearsal could facilitate multitasking in cases where the interrupting task itself makes heavy use of declarative memory.

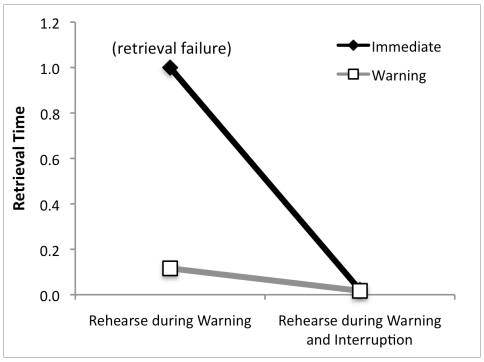

Figure 5: Time to retrieve the primary-task problem representation in the two Trafton et al. [43] conditions with two possible models of rehearsal.

Effects of Interruption Timing and Mental Workload

Another aspect of task interruption highlighted in several research efforts [e.g., 1, 13, 9, 19, 20] involves the timing of interruptions and how they relate to mental workload. Adamczyk and Bailey [1] have examined this issue in terms of the task model hierarchies inherent in the interrupted task: When viewing a task as a hierarchy of higher- to lower-level subtasks, an interruption coinciding with lower-level subtasks should be more disruptive. They tested this prediction with three different primary tasks involving document editing, media summarization, and web browsing, and an interrupting task that involved reading and responding to a brief newswire listing. Their predicted "best points" for interruption (i.e., at the higher-level task boundaries) produced better results across several measures (annoyance, frustration, etc.) than the predicted "worst points" for interruption. Along similar lines, Iqbal et al. [20; see also 9] created a formal GOMS model of route-planning and document-editing tasks and made similar predictions based on model subtask boundaries. They too found that interruptions at higher-level subtask boundaries are associated with decreased mental workload and thus serve as better points for task interruption.

In our theory, mental workload could be associated with a number of concepts, but most relevant to our purposes is the interpretation of workload as the complexity of problem representation. Specifically, an ACT-R model of a complex task (represented in a similar subtask hierarchy as mentioned above) would maintain problem representation at different levels of the hierarchy, and therefore would accumulate a fuller representation at the lower levels of the hierarchy. Because problem representation needs to be retrieved from memory under our theory, additional memory chunks for representation introduce more latency and more potential for error (i.e., failure to recall). For example, as illustrated in our model simulations in the previous section, retrievals normally require tens to hundreds of milliseconds per "chunk" of information, and if they are not properly rehearsed, the retrieval can fail (as in Trafton et al.'s Immediate condition with no rehearsal during interruption) and force re-encoding of the problem representation. In addition, the complexity of the memory chunks themselves may affect timing of recall: Altmann and Trafton [3] have recently posited that problem representations are retrieved from memory incrementally, one component at a time, leading to larger recall times for more complex representations. Both aspects of the memory system make interruption at higher-level subtasks more desirable, since they minimize problem representation and thus decrease resumption time and potential for resumption errors.

Effects of Interruption Task Type

Studies have also examined how different types of interrupting tasks may have differential effects on how interruption impacts primary-task performance. As one extreme, Monk et al. [29] compared two experimental conditions: one with pursuit tracking as the interrupting task, and the other with a "no-task interruption" in which the user was presented with a blank screen and waited for the primary task to resume. They found that users were much faster resuming the primary task after the no-task interruption than after the tracking-task interruption; the effect size of roughly .5 seconds was quite large given the overall resumption lag of roughly 1-1.5 seconds.

Under our account, the no-task interruption would offer a significant advantage over a tracking-task interruption because of the added time for memory rehearsal of the problem representation. During the no-task interruption, the user can repeatedly rehearse the representation such that it is immediately recalled when the primary task resumes. In contrast, an intensive interrupting task such as a pursuit tracking task would offer less time for rehearsal and thus hamper the recall of problem representation at the time of resumption. For tasks not on either extreme, the effects of interruption task type would be largely dependent on their interference with the procedural and memory processes necessary for memory rehearsal — the greater the use of these resources during the interruption, the larger the predicted resumption lag due to suppression of rehearsal.

To explore this issue, we ran our model with the same 5-second interruption as in the Monk et al. study. For a no-task interruption, the model performs 24 rehearsals during the interruption and retrieves the problem representation after interruption in 47 ms. For an interruption that includes a secondary task (using our general task model), the model performs only 14 rehearsals and retrieves the final representation in 71 ms. Of course, the latter case assumes that the person maximizes the number of rehearsals with no regard for performance on the secondary tracking task. If the person reduces the number of rehearsals to perform better on tracking, the retrieval time increases substantially; for example, limiting to 5 rehearsals results in a retrieval time of 326 ms, and limiting to 2 rehearsals results in a retrieval time of 707 ms.

Besides the effects of procedural and memory processes, another important property of the interrupting task is whether it requires the use of a problem representation. Some tasks do not require any problem representation (e.g., a pursuit tracking task), namely in situations where no information needs to be carried over between interruption and resumption. As Borst and Taatgen [10, 11] have shown that only one representation can be maintained concurrently, it follows that if the interrupting task does not need a problem representation, the representation of the main task can be maintained. This should result in less interference, as the problem representation does not have to be swapped out, and no memory retrievals are necessary. Indeed, comparing a combined driving and navigation task with and without the need for problem representations, they found that if both tasks need a problem representation there is more interference than when only one or none of the tasks needs a representation [10]. This effect was also found in a multitasking experiment where participants had to interleave solving subtraction problems and entering text, again showing increased interference when both tasks required a problem representation [11].

On a related note, another aspect of task type that has been studied is the relevance of the interrupting task to the original (interrupted) task. Cutrell, Czerwinski, and Horvitz [12] found that interruptions related to the primary task (in their case, instant messages relevant to a web search task) resulted in shorter times both processing the interruption and resuming the original task as compared to interruptions unrelated to the primary task. Under our account, this effect is closely associated with the processes inherent in the user's memory processing: As dictated by the ACT-R architecture, problem representation related to the current task spreads activation to related information during memory retrieval, and thus retrieval of task-relevant information occurs more quickly and with less chance of error (misretrieval or failure to retrieve). Thus, the theory allows the current task context to dictate, in at least one way, how easily relevant or irrelevant interruptions can be processed.

Effects of Primary-Task Visibility

Yet another factor that can affect resumption lag is the visibility of the primary task during interruption. In a recent study, Iqbal and Horvitz [22] analyzed how visual cues may affect resumption of an interrupted task. Specifically, they found that suspended application windows that were less than 25% visible during interruption took longer to return to than application windows that were more than 75% visible. This effect held for two types of interrupting tasks, namely for both email and instant-messaging alerts.

The effects of an interrupted task being visible or obscured could be accounted for in at least two ways under our theory. First, when an interrupted task remains visible, the rehearsal processes that mentally refresh the problem representation need not be completely based on declarative memory — they may also include repeated visual encoding of the visible problem state, reinforcing the representation in the same manner as memory retrieval. (If the problem representation must be reconstructed through visual encoding processes as part of resumption, as it might for complex tasks, this reinforcement would also speed up the reconstruction process.) Second, even if a person were not explicitly re-encoding pieces of the problem representation, any visual encoding of the interrupted task may spread activation to the full problem representation (as discussed in the previous section for task-relevant interruptions) and, in turn, thus facilitate recall of the primary-task problem representation and resumption of the primary task.

Effects on Interrupting-Task Performance

The phenomena outlined above emphasize the effects of interruption on the primary (interrupted) task. However, in some situations we may also be interested in potential effects on the *interrupting* task — that is, whether there would be performance differences when comparing its use as a secondary, interrupting task as opposed to a stand-alone, primary task. Salvucci and Beltowska [35] performed a study in which users memorized lists of numbers and, as an interrupting task, drove a computer driving simulator while rehearsing the memorized list. In addition, they collected data in a driving-only condition in which there was no list to memorize. The results showed a small but significant effect of memory rehearsal on two different measures of driver performance (lateral steering and brake reaction). The results are particularly relevant to our theory in that they show how memory rehearsal — the critical task thread needed during interruption — can affect performance of the interrupting task.

Our theory can account for such effects on interrupting-task performance as follows. Because rehearsal of the problem representation can occur concurrently with the interrupting task, our theory suggests that, in general, an interrupting task will experience decreased performance due to the additional cognitive processing needed for rehearsal. The extent of interference depends on what resources the interrupting task needs. Memory rehearsal in the ACT-R theory is not a particularly intensive process for the central procedural resource: It involves an occasional procedural step of 50 ms followed by a long declarative retrieval of perhaps 200-500 ms, as shown in Figure 4. If the secondary task involves heavy use of declarative memory (e.g., for instruction following), we can expect more significant effects on performance (although it is possible that, if memory rehearsal can be performed by subvocalization, these effects could be alleviated). On the other hand, even for interrupting tasks with no memory-resource conflicts, the procedural step of 50 ms every 200-500 ms may produce observable effects for tasks that require intensive, rapid sequences of action. This is the case with the example of memory rehearsal while driving above: This procedural step produces a small but significant effect on driver performance when interleaved with the rapidly iterating procedural steps of the driving task.

Difficulty	Category	Example
1 (least)	Motor movements	Move mouse towards a menu item or select a menu item
2	Routine content generation	Enter a new filename or select a transition for a video clip
3	Comprehension or store information	Read text or comments, retrieve a route segment's distance and fare information and commit it to memory
4	Recall information	Recall a route segment's distance and fare information
5	Creative content generation	Edit document text or edit video clips
6 (most)	Mathematical reasoning	Add distance or fare information

Table 1: Levels of subtask difficulty, from Iqbal and Bailey [20].

Changing Effects over Time

Compared to the many studies of the effects interruption on task performance, there have been relatively few studies of the changing effects over time as might come with learning or practice. In their study discussed earlier, Trafton et al. [43] included an analysis of how users in the Warning (8-second interruption lag) and Immediate (no interruption lag) changed their performance with practice. They found that users were able to improve their ability to resume the task — specifically, reducing the resumption lag — over time, but only in the Immediate condition. In fact, by the third of three 20-minute sessions, the resumption lag in the Immediate condition was reduced to a value very close to that in the Warning condition (roughly 4.3 seconds vs. 3.7 seconds respectively). Thus, it seems that even with no interruption lag during which to prepare for interruption, users can adapt and improve with repeated interruptions.

An account of this learning effect arises from a prediction of the underlying ACT-R architecture, namely that with practice, task performance becomes more proceduralized through a process called *production compilation* [39], which reduces the demands on declarative and procedural resources. As a result, threaded cognition would be better able to interleave rehearsal with the interrupting task, thus predicting a smaller resumption lag over time. The simulation results in Figure 5 demonstrate how, when additional rehearsal occurs during interruption, retrieval times in the Warning and Immediate condition begin to equalize and show the same learning effect observed in the Trafton et al. study. Indeed, in addition to reducing the resumption lag over time, Trafton et al.'s users became more proficient on the interrupting task over the time, supporting the notion that improvements in the individual task skills relate to overall improvements with respect to interruption and resumption.

Self-Interruption

While the above phenomena are primarily concerned with external interruptions generated by an outside source (human or otherwise), people also have the ability to self-interrupt — that is, to stop themselves while performing one task in order to switch to a potentially more critical task. Self-interruption can occur after a relatively short time, such as a user deciding to quickly check email for a few seconds before returning to writing a paper. Self-interruption can also occur over longer intervals, such as a user deciding to work on a manuscript for one hour before responding to email.

Our theory can account for at least some aspects of self-interruption in the sense of setting an internal clock to return to a task. A recent extension to the ACT-R theory posits a computational theory of psychological time [41] by which a person estimates the time spent on a task. Using such a mechanism, a user could set this internal clock to run for a set period of time, and when the clock reaches the desired time, switch to another task. This theory of time estimation applies to estimates on the order of several seconds to several minutes. Unfortunately, our theory currently says little about task switching that may involve higher-order planning and scheduling over longer periods of time.

Relation to Model-Based Approaches to Interruption

In an interesting approach to exploring the various aspects of interruption, Iqbal and Bailey [21] identified several factors that may affect interruption and ran a multiple-regression analysis to determine the most influential factors. They focused on three factors in particular: *level*, *task difficulty*, and *information carry-over*. The first factor, *level*, is that discussed earlier in the context of mental workload: Whether the interruption occurred between higher- or lower-level task boundaries. The second factor, task difficulty, was posited in terms of a six-way characterization of the difficulty of the next subtask, derived from their earlier studies and depicted in Table 1. The third factor, *information carry-over*, represented (in four levels) the amount of mental information that needed to be carried over from one task to another. Iqbal and Bailey found that these three factors accounted, all together, for 26% of the variance in cost of interruption, where cost was measured by resumption lag.

Under our theory, as we discussed earlier, the *level* factor corresponds directly to the size and complexity of the problem representation in the task. The *information carry-over* factor is also closely related here: As additional information needs to be saved during the interruption, rehearsal of this information becomes more difficult in terms of time and potential for error. The *difficulty* factor also falls

out nicely from our theory. The least difficult subtask (1), namely that involving motor movements, makes the least demands on procedural and memory resources and thus interferes least with memory rehearsal. The levels with medium difficulty (2-4) require some amount of memory processing, but not in an intense way, whereas the levels with highest difficulty (5-6) require not only memory processing but also heavy cognitive and procedural processing and higher-level reasoning. The only discrepancy of concern in our account would be the fact that comprehension (3) is only of medium difficulty; at this time, the underlying ACT-R theory does not posit a separate language processor, but it is possible that comprehension here utilizes resources separate from main memory and thus results in lower interference [see 26]. Nevertheless, the core ideas of memory-for-goals and memory rehearsal as a critical process during interruption nicely account for the three primary factors in the multiple-regression analysis.

GENERAL DISCUSSION

While the multitasking continuum has typically been separated into the two separate research areas of concurrent and sequential multitasking, this paper aims to show how a single unified theory can begin to account for behavior along the entire continuum. The ACT-R cognitive architecture provides the specification of resources and a framework for formalizing behavior in terms of procedural steps. The memory-for-goals theory builds on ACT-R memory theory to posit how task goal memories are encoded, rehearsed, and recalled. The work on problem representation shows that the stored state of each task goal is critical for understanding what information must be stored and recalled during interruption. Finally, threaded cognition theory provides a framework for interleaving multiple tasks, thus integrating memory-based representation rehearsal processes with the performance of the primary and secondary tasks.

As mentioned, we have focused in this paper on breadth of explanation, using an over-arching theoretical framework to account for as broad a range of phenomena as possible. We have also presented a computational model that simulates how the core processes of problem representation rehearsal and retrieval would behave, using generic single-rule task models as a stand-in for full-fledged models of the primary and secondary tasks. Future efforts could build on this work by instantiating these models in detail for various combinations of complex tasks. Recent efforts using GOMS models to explore various aspects of task interruption [e.g., 9, 20, 21] offer promising building blocks for the development of these models. Regardless of the particular modeling framework, our work suggests that memory rehearsal is a critical process during interruption and that any attempt to understand user behavior during interruption and resumption requires some consideration of how rehearsal can be interleaved with both the primary and secondary tasks.

In addition, while our focus to this point has been on scientific understanding of the underpinnings of multitasking behavior, a logical step in a different but complementary direction could involve bootstrapping this work into real-time systems that utilize the underlying psychological theories. For example, a number of researchers have been exploring real-time systems that help to understand and manage user interruptions [14, 15, 17, 18]. Such systems typically incorporate mathematical and/or computational models of behavior and interruptibility at various points of interaction. Likewise, we hope that the theory can soon be utilized in this way by representing user state and predicting interruptibility. In fact, ACT-R has already been employed in a similar way in the context of model-tracing algorithms embedded into intelligent tutoring systems [8]; one might imagine using a similar methodology for real-time interruption management by, for instance, checking the complexity and activation of current problem representations and translating such values into a rigorous measure of interruptibility.

ACKNOWLEDGMENTS

This work was sponsored by grants from the Office of Naval Research (#N00014-03-1-0036 and #N00014-08-10541), the National Science Foundation (#IIS-0426674), and the Air Force Office of Scientific Research (#FA95500710359).

REFERENCES

1. Adamczyk, P.D. and Bailey, B.P. (2004). If not now when? The effects of interruptions at different moments within task execution. *Proc. CHI 2004*, 271-278.
2. Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science*, *26*, 39-83.
3. Altmann, E. M., & Trafton, J. G. (2007). Timecourse of recovery from interruption: Data and a model. *Psychological Bulletin & Review*, *14*, 1079-1084.
4. Anderson, J. R. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science, 29*, 313-341.
5. Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford: Oxford University Press.
6. Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*, 1036-1060.
7. Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, *38*, 341-380.
8. Anderson, J. R., Corbett, A. T., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, *4*, 167-207.
9. Bailey, B. P., & Iqbal, S. T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Human-Computer Interaction*, *14,* 1-28.
10. Borst, J. P., & Taatgen, N. A. (2007). The Costs of Multitasking in threaded cognition. *Proceedings of the Eighth International Conference on Cognitive Modelling*, 133-138.

11. Borst, J. P., Taatgen, N. A., & Van Rijn, H. (submitted). Problem representations in multitasking: An additional cognitive bottleneck.
12. Cutrell, E. B., Czerwinski, M., & Horvitz, E. (2000). Effects of instant messaging interruptions on computing tasks. *Proc. CHI 2000*, *Extended Abstracts*, 99-100.
13. Czerwinski, M., Cutrell, E. and Horvitz, E. (2000). Instant messaging: Effects of relevance and timing. *People and Computers XIV: Proc. HCI 2000*, 71-76.
14. Fogarty, J., Hudson, S.E. and Lai, J. (2004). Examining the robustness of sensor-based statistical models of human interruptibility. *Proc. CHI 2004*, 207-214.
15. Fogarty, J., Ko, A.J., Aung, H.H., Golden, E., Tang, K.P. and Hudson, S.E. (2005). Examining task engagement in sensor-based statistical models of human interruptibility. *Proc. CHI 2005*, 331-340.
16. Gonzalez, V. M., & Mark, G. (2004). "Constant, constant, multi-tasking craziness": Managing multiple working spheres. *Proc. CHI 2004*, 113-120.
17. Horvitz, E., Apacible, J. and Subramani, M. (2005). Balancing awareness and interruption: Investigation of notification deferral policies. *Proc. User Modeling 2005*, 433-437.
18. Hudson, S.E., Fogarty, J., Atkeson, C.G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J.C. and Yang, J. (2003). Predicting human interruptibility with sensors: A Wizard of Oz feasibility study. *Proc. CHI 2003*, 257-264.
19. Iqbal, S. T., Adamczyk, P. D., Zheng, S. X., & Bailey, B. P. (2005). Towards an index of opportunity: Understanding changes in mental workload during task execution. *Proc. CHI 2005*, 311-320.
20. Iqbal, S.T. and Bailey, B.P. (2005). Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. *Proc. CHI 2005*, 1489-1492.
21. Iqbal, S.T. and Bailey, B.P. (2006). Leveraging Characteristics of task structure to predict costs of interruption. *Proc. CHI 2006*, 741-750.
22. Iqbal, S.T. and Horvitz, E. (2007). Disruption and recovery of computing tasks: Field study, analysis and directions. *Proc. CHI 2007*, 677-686.
23. Jones, R. M., Laird, J. E., Nielsen P. E., Coulter, K., Kenny, P., & Koss, F. (1999). Automated intelligent pilots for combat flight simulation. *AI Magazine*, *20*, 27-42.
24. Kieras, D. E. (2007). Control of cognition. In W.D. Gray (Ed.) *Integrated Models of Cognitive Systems* (pp. 327-355). Oxford University Press.
25. Lallement, Y., & John, B. E. (1998). Cognitive architecture and modeling idiom: An examination of three models of the Wickens task. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 597-602). Hillsdale, NJ: Erlbaum.
26. Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*, 375-419.
27. Martin-Emerson, R., & Wickens, C. D. (1992). The vertical visual field and implications for the head-up display. In *Proceedings of the Thirty-Sixth Annual Symposium of the Human Factors Society* (pp. 1408-1412). Santa Monica, CA: Human Factors Society.
28. McCrickard, D. S., Chewar, C. M., Jacob, P. S., & Ndiwalana, A. (2003). A model for notification systems evaluation—assessing user goals for multitasking activity. *ACM Transactions on Computer-Human Interaction*, *10*, 312-338.
29. Monk, C.A., Boehm-Davis, D.A. and Trafton, J.G. (2004). Recovering from interruptions: Implications for driver distraction research. *Human Factors*, *46*, 650-663.
30. Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*, 134-140.
31. Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
32. Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*, 220-244.
33. Salvucci, D. D. (2001). Predicting the effects of in-car interface use on driver performance: An integrated model approach. *International Journal of Human-Computer Studies*, *55*, 85-107.
34. Salvucci, D. D. (2005). A multitasking general executive for compound continuous tasks. *Cognitive Science*, *29*, 457-492.
35. Salvucci, D. D., & Beltowska, J. (2008). Effects of memory rehearsal on driver performance: Experiment and theoretical account. *Human Factors*, *50*, 834–844.
36. Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, *115*, 101-130.
37. Schumacher, E. H., Seymour, T. L., Glass, J. M., Fencsik, D. E., Lauber, E. J., Kieras, D. E., et al. (2001). Virtually perfect time sharing in dual-task performance: uncorking the central cognitive bottleneck. *Psychological Science, 12*, 101-108.
38. Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science*, *12*, 462-466.
39. Taatgen, N. A., & Lee, F. J. (2003). Production compilation: A simple mechanism to model complex skill acquisition. *Human Factors*, *45*, 61-76.
40. Taatgen, N. A., Juvina, I., Schipper, M., Borst, J., & Martens, S. (in press). Too much control can hurt: A threaded cognition model of the attentional blink. *Cognitive Psychology.*
41. Taatgen, N. A., Rijn, H. v., & Anderson, J. R. (2007). An integrated theory of prospective time interval estimation: The role of cognition, attention and learning. *Psychological Review*, *114*, 577-598.
42. Telford, C. W. (1931). The refractory phase of voluntary and associative response. *Journal of Experimental Psychology*, *14*, 1-35.
43. 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.
44. Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*, 159-177.