

MODEL-BASED PREDICTIONS OF INTERRUPTED CHECKLISTS

Melanie Diez, Deborah A. Boehm-Davis and Robert W. Holt

Department of Psychology
George Mason University
Fairfax, VA

The improper completion of a cockpit checklist has contributed to a number of aviation accidents. In many of these cases, it can be shown that interruptions were a contributing factor in the failure to complete the checklist properly. Unfortunately, most studies of interruptions have provided only post-hoc explanations for these failures. Further, research has focused on whether or not tasks are resumed rather than on predictions of where people will resume a task after an interruption. This paper describes several generic models that were used to explore cognitive strategies for handling interruptions. One of these models was then modified to fit the specific real-world task of completing an aircraft checklist. This model produced detailed *a priori* predictions about where the interrupted checklist will be resumed. The implications of these predictions for task design and for the use of cognitive modeling as an approach to understanding interruptions are discussed.

INTRODUCTION

The world of aviation has long relied on checklists to help pilots configure their aircraft safely and efficiently. Checklists provide an external memory for storing standardized procedures, eliminating the need to recall lengthy lists of items from memory. However, the efficacy of a checklist can be diminished by a number of factors, including poor design and social issues which may lead to improper checklist usage (Degani & Wiener, 1990).

Several major aviation accidents testify to the importance of checklists as well as how quickly a normal flight can deteriorate when a checklist is performed improperly. In 1968, a B-707 crashed on takeoff because the flaps were fully retracted (NTSB, 1969). According to the cockpit voice recorder (CVR), the pilots were in the middle of executing the TAXI checklist when they received a call from the control tower. The Flaps item was "passed over" and never accomplished before takeoff. A similar situation occurred in 1987 when a DC-9 crashed shortly after takeoff (NTSB, 1988). The National Transportation Safety Board (NTSB) determined that the flaps and slats had not been extended because the TAXI checklist was neither initiated nor executed. The CVR revealed numerous disruptive events occurring during the taxi phase, prompting the NTSB to hypothesize that, "If there are interruptions and the checklist has not been initiated normally, when the airplane reaches a point in the taxi where the TAXI checklist typically has been completed, it is possible that the flightcrew will believe that the checklist was completed" (p. 59). One year later, history repeated itself as a B-727 crashed during takeoff due to fully retracted flaps and slats (NTSB, 1989). In this case, the pilots suspended several items on the TAXI and BEFORE TAKEOFF checklists due to airport delays. When the checklists were resumed, their execution was rushed and the Flaps item was not correctly accomplished.

Unfortunately, it is not surprising that checklists are often completed improperly. The flightdeck is often a flurry of

activity with Air Traffic Control (ATC), flight attendants, co-pilots, and instruments all competing for attention. Although the purpose of the checklist is to ensure that the most critical items are accomplished, it is not unusual for pilots to suspend the checklist in favor of more time critical tasks (e.g., responding to an urgent clearance from ATC). Depending on the phase of flight, an interrupted checklist may be the norm rather than the exception due to the dynamic nature and tempo of today's flight operations (Loukopoulos, Dismukes, & Barshi, 2001).

Even though interruptions are ubiquitous for most everyday tasks, they do not always lead to a failure to complete the primary task. Further, when they do, the price we pay is often limited to minor annoyance and inefficiency. However, a seemingly harmless interruption during a safety-critical task (e.g., an ATC clearance issued while a pilot is executing a checklist) can potentially cost lives if the interrupted task is not resumed properly.

The literature on task interruptions is sparse, yet growing. Numerous researchers from a variety of fields have recognized the unpredictable effects that interruptions can have on decision making (Speier, Valacich, & Vessey, 1999), performance time (Eyrolle & Cellier, 2000), productivity (O'Connell & Frohlich, 1995), and error rates (Flynn et al., 1999). Thanks to countless incident and accident reports, the aviation domain is now well aware of these effects as well as how vulnerable pilots are to interruptions and distractions on the flightdeck (Dismukes, Young, & Sumwalt, 1998). Clearly, many efforts have been made to identify and quantify interruptions as well as describe their effects (both positive and negative) on human performance. (For an excellent review of this literature, see Latorella (1999)).

With so many investigations into the effects of interruptions, it is surprising to note that only a handful of these studies ground their findings within a cognitive theory and very few theories predict how and where an interrupted task will resume (Altmann & Trafton, 2002). This may be due to the fact that few theories specify the underlying

mechanisms involved in the interruption process. Consequently, they lack the ability to make precise *a priori* predictions and their conclusions are limited to generalizations. It is one thing to say that an interruption will cause a delay in performance. It is another thing to predict why the delay occurs, how performance will be affected, and where the interrupted task will be resumed.

Due to their susceptibility to interruptions, safety-critical nature and hierarchical structure, we chose to explore the effects of interruptions within the context of a cockpit checklist. This paper presents a model of an interrupted checklist that makes specific predictions about where the suspended task is resumed after attending to an interruption. It should be noted that this model was created in the absence of data and was not fit to any quantitative results. This approach was taken to avoid the common criticisms of 'data fitting' that typically plague computational modelers (Roberts & Pashler, 2000). Instead, we chose to take a well-established cognitive architecture and develop a generic model that could handle interruptions while performing a simplistic task. The architecture itself provided the constraints for modeling the task, rather than an actual data set.

THESIS

In this research project, computational cognitive models were used to explore and establish a set of cognitive strategies that could be used to handle interruptions in a real-world task. These models were grounded in the ACT-R (version 4.0) cognitive architecture (Anderson & Lebiere, 1998). ACT-R is a computational cognitive theory based on a production system that allows modelers to develop simulations of human behavior. Over 100 ACT-R models have been published to date on topics ranging from visual attention to arithmetic, and from playing backgammon to making scientific discoveries. This track record, covering such a wide variety of applications and cognitive phenomena, suggests that the ACT-R architecture offers a fair approximation of human cognitive ability.

For this research project, a two-pronged approach was adopted. First, several generic models were created to establish basic strategies of goal prioritization and retrieval. Next, one of the generic models was modified to simulate a real-world task that is known to be susceptible to interruptions (a cockpit checklist). Separating the first step from the second step ensured that the nature of the checklist did not influence development of the generic models. The specific checklist model makes precise predictions regarding where the task will be resumed after being interrupted. Implications for these findings are discussed.

SOURCES OF INFORMATION

Exploratory (Generic) Models

The purpose of the generic model was to establish basic strategies of goal prioritization and retrieval. Given that our primary interest was in modeling interruptions, we focused on modeling goal creation, the onset of an interruption, and finally, goal resumption.

Goal creation. In the ACT-R (version 4.0) cognitive architecture, "goals" direct behavior. These goals are stored on a goal "stack" (see Figure 1) upon which new goals are 'pushed' on or 'popped' off. ACT-R only permits one goal to be active at a time and the goal on top of the stack is always the active goal. That is, pushing a goal onto the stack makes that goal the active goal, allowing it to direct behavior. Popping a goal removes attention from that goal and focuses on the next goal underneath. Given these constraints, there are two methods for switching to a new goal (Table 1, column a). A "Push" simply pushes a new goal on top of the current one in the goal stack, even if the current goal is incomplete (see Figure 1). A "Focus-on" pops the current goal off of the goal stack, even if it is incomplete, before pushing a new one in its place (see Figure 2).

The first method (Push) is the one most commonly found in traditional ACT-R models. Because goals stay on the goal stack until they have been completed, the stack tends to get deeper the further down in the task hierarchy one goes and goals are never forgotten. When a goal is finally completed it is popped off, revealing the goal beneath it. In the second method (focus-on), the current goal is removed from the goal stack and returned to declarative memory (referred to as a goal 'chunk'). In its place, the new goal is pushed onto the stack. The consequence of this method is that the goal stack remains relatively flat because goals are popped off before new ones are pushed on. As a result, the current goal that is displaced must be recalled later on before it is forgotten.

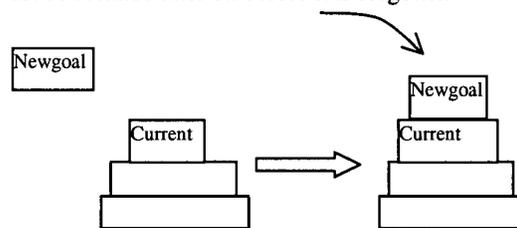


Figure 1. "Push" or Method 1 where a new goal is pushed directly onto the stack.

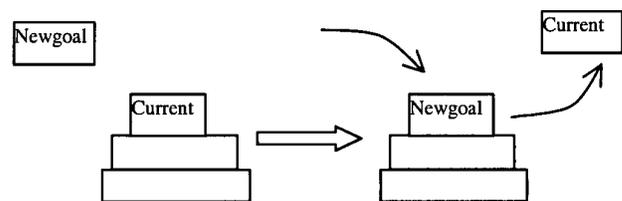


Figure 2. "Focus-on" or Method 2, where the current goal is removed before a new goal is pushed

Onset of interruptions. In our models, an interruption is always and instantly attended to. The same two methods described above could be used to switch to a new interruption goal (Table 1, column b). That is, the interrupt goal could simply be pushed onto the stack (Push), or the current goal could be removed before pushing the interruption task goal (Focus-on).

Task resumption. In a traditional ACT-R goal stack, a task resumes when the current goal is popped, revealing the suspended task beneath (see Table 1, column c). However, a model using the Focus-on method removes a suspended task from the goal stack and returns it to declarative memory before pushing a new goal. This means that an additional process is needed to select which goal should be resumed from the set of incomplete task goals that have been popped from the stack. Our approach was to direct the model, upon completing the interrupting task, to search its memory for the most active goal chunk. This process relies on the concept of

Table 1. Breakdown of eight generic models. Column a: strategy used to switch to new goals. Column b: strategy used to switch to interruption. Column c: Strategy used to resume original task. Column d: Model behavior when interrupted.

a	b	c		d
<i>Onset of new goals</i>	<i>Onset of interruption</i>	<i>Task resumption</i>	<i>Model</i>	<i>Result</i>
Push	Push	Pop	1	Perfect
		Focus-on	2	Redundant
	Focus-on	Pop	3	Repeat
		Focus-on	4	Repeat
Focus-on	Push	Pop	5	Perfect
		Focus-on	6	Redundant
	Focus-on	Pop	7	Failure
		Focus-on	8	Repeat

activation-based retrieval built into the ACT-R architecture. The activation of a goal chunk is a function of both the frequency and recency of use, as well as the presence of any cues that could prime one goal over another. The result is that the model looks for the next most highly active goal chunk in memory and makes that chunk the new goal. Besides facilitating task resumption, this process also allows the model to make mistakes and “forget” goals that have been suspended for some time. Once the most active goal has been identified, the process could either push that goal onto the stack or focus on that goal.

Final generic models. In the end, we created a total of eight models of the same generic task that represented all of the combinations of these mechanisms for goal creation, onset of interruption, and task resumption (see Table 1, columns a-c). We chose this path in order to explore the ways in which choices exercised by the modeler affect model outcomes.

Findings: Exploratory (Generic) Models

To compare the ability of each model to resume an interrupted task, a single interruption was introduced at the

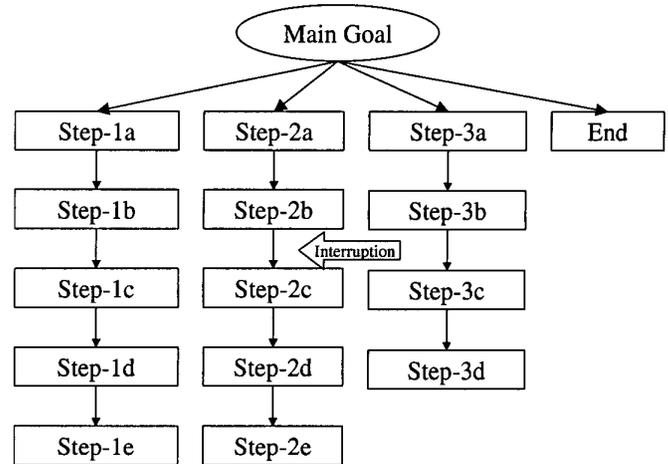


Figure 3. Hierarchy showing generic task with point of interruption.

same pre-determined point for each of the eight models (see Figure 3). We then observed the behavior of each model after being interrupted. Task resumption performance was qualitatively different for many of the models, ranging from perfect (picking up exactly where it left off) to total failure (see Table 1, column d).

Two models (1 and 5) exhibited perfect performance by resuming the suspended task exactly where they had left off. Two models (2 and 6) re-started the interrupted task, however, they later “remembered” that they had already done the task and merged the memories of both instances into one memory chunk. Three models (3, 4, and 8) re-started the suspended task after the interruption but never realized that they had already done some of the task. Finally, one model (7) simply stopped after attending to the interruption because all of the goals had been popped from the stack.

After examining the results of this exploratory modeling effort, we needed to select one version for further research. Although most of the models were able to resume and complete the task, some models were more elegant and cognitively plausible than others. For instance, the two models that exhibited perfect performance were eliminated for their perfection. Our goal was not to create a perfect model, but a model that could exhibit human-like performance. Although humans can do tasks perfectly, these models did not allow for the flexibility we desired. Likewise, the model that simply stopped was eliminated as well because it was not capable of getting back on track once it had been interrupted. We deemed this model cognitively implausible since few humans exhibit such catastrophic failures of memory. Finally, we favored the models that employed the Focus-on method as more cognitively plausible since the Push method assumes that all goals on the stack are perfectly preserved in memory. Consequently, we decided to adopt the model that consistently employed Focus-on for all of its goals (model 8). We could see no reason to treat new goals and interrupting goals differently, so each goal creation, onset of interruption, and task resumption was accomplished with a Focus-on strategy.

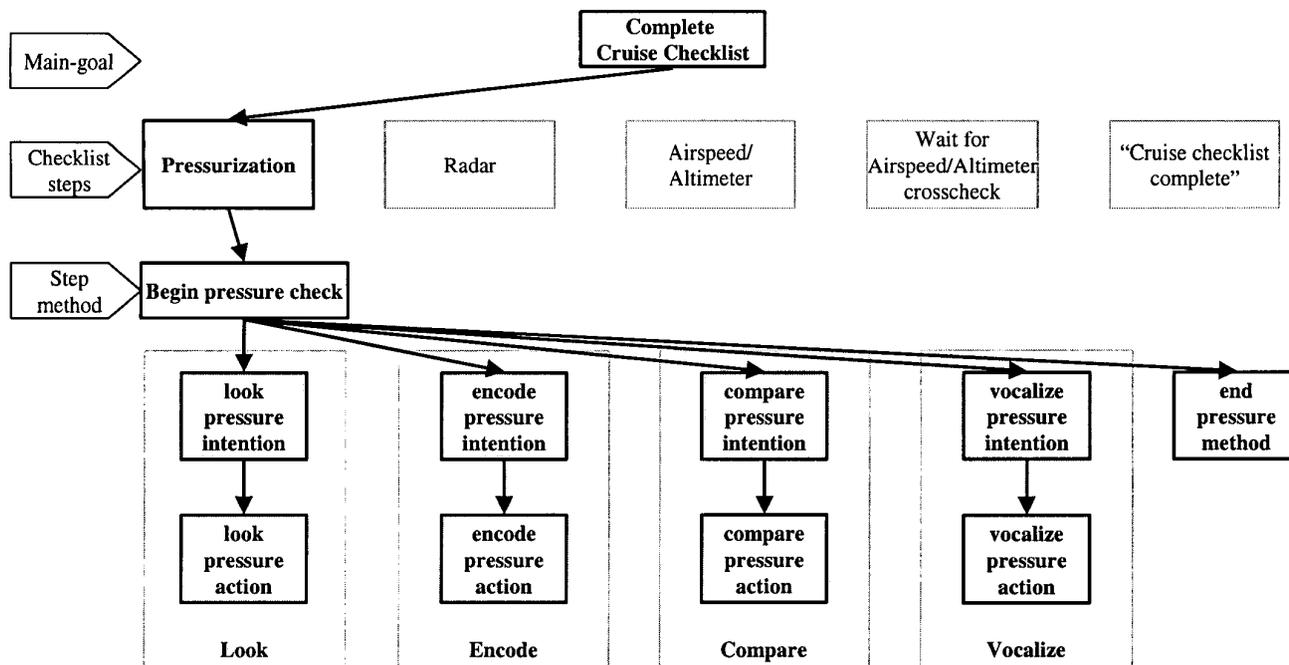


Figure 4. Cruise checklist task hierarchy.

Specific Checklist Model

To better understand cockpit interruptions, it was necessary to apply our generic model to a specific cockpit checklist task. We chose to model a cruise checklist used by a North American commercial carrier. The checklist consisted of five steps: Check pressurization, check radar, check airspeed and altimeter, wait for cross check from co-pilot, and announce “cruise checklist complete.” Some steps could be broken down into sub-steps (e.g., acquiring and comparing two pressurization values), while others did not require additional actions (e.g., say, “Cruise checklist complete.”)

The model uses the main-goal (“Complete cruise checklist”) to keep track of which steps (the five specific checklist items) need to be completed (see Figure 4). If a step has not been completed, the main-goal sets a sub-goal to begin that step. Each step has a method associated with completing that step. After completing each sub-goal, the model retrieves the main-goal again, records the step that has just been completed, and then recalls the next step. As before, interruptions were attended to immediately. At the end of an interruption, the model would request the most highly active goal chunk and resume the task from that point.

Findings: Specific Checklist Model

The focal question for the specific model was, If the checklist is interrupted, where, in the task hierarchy does the model resume? To answer this question, the checklist model was interrupted at each checklist item and the point of resumption was recorded. When the interruption occurred during the first checklist item, the model exhibited perfect performance by picking up exactly where it had left off. However, when the interruption occurred during the later four steps, the model was unable to recall the exact step to resume.

Instead, the model resumed the checklist task at the next highest goal from where the interruption occurred.

This pattern of behavior is a result of the dynamic properties of memory activation and the design of the model. In the case of the first run of the model, the interruption occurred early enough in the task that the main-goal had not been retrieved any more frequently than the sub-goals. When the model requested the most highly active goal chunk after the interruption, it retrieved the most *recently retrieved* chunk—that of the last step performed before the interruption.

When the interruptions occurred later in the task, results were different. Since the model uses the main-goal after every sub-task to decide what to do next, the main-goal chunk became more highly active. When the model requested the most highly active goal chunk after the interruption, it retrieved the most *frequently retrieved* chunk, which was the main-goal.

DISCUSSION

The behavior exhibited by our checklist model makes two interesting qualitative predictions. First, it suggests that a person who is interrupted early in the midst of a task should be able to resume that task faster (and more accurately) than if the interruption came later in the task. Although this seems intuitive, the model offers a deeper understanding of *why* this occurs. Early in a task, recency tends to play a greater role in determining which goal is recalled, whereas frequency tends to play a greater role in recalling goals later in a task.

This brings up the second, more surprising prediction: that interruptions later in a task will force the person to resume the task at a higher level. This is surprising because, although one might expect it to take longer to recall where they were in a task, one might not expect that when they do resume the task, it is not where they left off. The model predicts that tasks interrupted at a deeper goal level will take longer to

resume because additional “cognitive steps” are required to get back to the resumption point.

The predictions made by this model are precise but plausible. For instance, a person who is interrupted while relaying a story to a friend does not resume the story with the very next word in the sentence, but rather pauses to reflect on what she was about to say. This suggests resumption of a higher goal (tell the story), rather than to the sub-goal (say this sentence). Similarly, musicians who are interrupted while practicing a musical piece frequently “take it from the top” or back up to the beginning of a phrase. Such anecdotal evidence is consistent with the predictions offered by this model.

On a more empirical note, Monk, Boehm-Davis, and Trafton (this volume) report findings consistent with this prediction in their study of interruptions of in-vehicle information-style tasks (IVIS). Specifically, they found that interruptions introduced at the start of an IVIS task were more quickly resumed than those introduced mid-task. Taken together, these predictions and findings offer some general guidance for the design of task aids and environmental cues. Given the ubiquitous nature of interruptions, tasks that offer cues to aid the user back to the task after being interrupted should help task resumption. Likewise, training interventions could improve task prioritization by emphasizing the negative effects of attending to an interruption in mid-task.

There are, of course, several caveats to our theoretical approach. First, our model assumed that the pilot performed the checklist steps from memory and did not take into account external memory aids such as checklist cards, instrument panels, or environmental cues. While some airlines allow pilots to perform checklists from memory, the vast majority requires them to use paper, mechanical, or electronic formats. This removes some of the burden from working memory by effectively maintaining an external “goal stack.” Our model assumed no such aids. An interesting line of future research would be to expand the model to incorporate such memory aids. Although we would expect task resumption to improve, we are mindful of the fact that all of the accidents discussed in the introduction took place *in spite of* checklist aids.

Another limitation of this modeling effort is our assumption that all interruptions were attended to immediately. In the real world, people have a choice whether or not to respond to an interruption, as well as some flexibility in how quickly they respond. Recent modeling efforts by Altmann and Trafton (2002) suggest that the time between task suspension and attending to an interruption (called an *interruption lag*) can be used advantageously by rehearsing the interrupted location to facilitate task resumption. Currently, our model does not provide a mechanism for ignoring interruptions or rehearsing suspended goals.

Finally, our model only focused on one aspect of interruptions-- that is, how the location of an interruption within a task hierarchy affects task resumption. We made no attempt to examine the role of similarity, complexity or time duration on task resumption. Our goal in this research project was to tease apart the mechanisms of task resumption under one specific set of circumstances. We relied on the built-in constraints of a computational cognitive architecture to guide our modeling efforts and viewed our theoretical approach as a

means of hypothesis generation. Future research will focus on collecting and comparing human data against our model's predictions. Once the predictions of this model have been tested, further modeling efforts will explore factors such as external aids, rehearsal, and interruption characteristics.

REFERENCES

- Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science*, 26, 39-83.
- Anderson, J. R., & Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Degani, A., & Wiener, E. L. (1990). *Human factors of flight-deck checklists: The normal checklist* (NASA CR-177549). Ames Research Center, Moffett Field, CA: NASA.
- Dismukes, K., Young, G., & Sumwalt, C. R. (1998, December 1998). Cockpit interruptions and distractions. *ASRS Directorate Magazine*.
- Eyrolle, H., & Cellier, J.-M. (2000). The effects of interruptions in work activity: field and laboratory results. *Applied Ergonomics*, 31, 537-543.
- Flynn, E. A., Barker, K. N., Gibson, J. T., Pearson, R. E., Berger, B. A., & Smith, L. A. (1999). Impact of interruptions and distractions on dispensing errors in an ambulatory care pharmacy. *American Journal of Health-System Pharmacy*, 56(13), 1319-1325.
- Latorella, K. A. (1999). *Investigating interruptions: Implications for flightdeck performance* (Technical Memorandum NASA/TM-1999-209707). Hampton, VA: NASA Langley Research Center.
- Loukopoulos, L. D., Dismukes, R. K., & Barshi, I. (2001). Cockpit interruptions and distractions: A line observation study. *Proceedings of the Eleventh International Symposium on Aviation Psychology, Columbus, OH*.
- Monk, C. A., Boehm-Davis, D. A., & Trafton, J. G. (this volume). The attentional costs of interrupting task performance at various stages. *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society*.
- NTSB. (1969). *Aircraft accident report: Pan American World Airways, Inc., Boeing 707-321C, N799PA, Elmendorf Air Force Base, Anchorage, Alaska, December 26, 1968* (NTSB/AAR-69/08). Washington, DC: National Transportation Safety Board.
- NTSB. (1988). *Aircraft accident report: Northwest Airlines, Inc., McDonnell Douglas DC-9-82, N312RC, Detroit Metropolitan Wayne County Airport, Romulus, Michigan, August 16, 1987* (NTSB/AAR-88/05). Washington, DC: National Transportation Safety Board.
- NTSB. (1989). *Aircraft accident report: Boeing 727-232, N473DA, Dallas-Fort Worth International Airport, Texas* (NTSB/AAR-89/04). Washington, DC: National Transportation Safety Board.
- O'Connell, B., & Frohlich, D. (1995). Timespace in the workplace: Dealing with interruptions. *Proceedings of the Conference on Human Factors in Computing Systems (CHI'95), Denver, CO*.
- Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? *Psychological Review*, 107(2), 358-367.
- Speier, C., Valacich, J. S., & Vessey, I. (1999). The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences*, 30(2), 337-360.