ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT

Int. J. Human-Computer Studies 58 (2003) 583–603

International Journal of Human-Computer Studies

www.elsevier.com/locate/ijhcs

Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal

J. Gregory Trafton[a,*], Erik M. Altmann[b], Derek P. Brock[a], Farilee E. Mintz[c]

[a] *Naval Research Laboratory, Code 5513, 4555 Overlook Av. S.W., Washington, DC 20375-5337, USA*

[b] *Department of Psychology, Michigan State University, East Lansing, MI 48824, USA*

[c] *ITT Industries, AES Division, 2560 Huntington Avenue, Alexandria, VA 22303, USA*

Received 1 May 2002; accepted 30 August 2002

Abstract

We examine people's strategic cognitive responses to being interrupted while performing a task. Based on memory theory, we propose that resumption of a task after interruption is facilitated by preparation during the *interruption lag*, or the interval between an alert to a pending interruption (e.g. the phone ringing) and the interruption proper (the ensuing conversation). To test this proposal, we conducted an experiment in which participants in a Warning condition received an 8-s interruption lag, and participants in an Immediate condition received no interruption lag. Participants in the Warning condition prepared more than participants in the Immediate condition, as measured by verbal reports, and resumed the interrupted task more quickly. However, Immediate participants resumed faster with practice, suggesting that people adapt to particularly disruptive forms of interruption. The results support our task analysis of interruption and our model of memory for goals, and suggest further means for studying operator performance in dynamic task environments.
Published by Elsevier Science Ltd.

Keywords: Interruption; Goals; Rehearsal; Prospective memory; Task switching

*Corresponding author.
E-mail address: trafton@itd.nrl.navy.mil (J.G. Trafton).

1071-5819/03/$ - see front matter Published by Elsevier Science Ltd.
doi:10.1016/S1071-5819(03)00023-5

1. Introduction

Consider a typical interruption. You are in conversation with a colleague when the phone rings. How do you respond? You might excuse yourself for a moment, pick up the phone and quickly reschedule the caller ("I'm in a meeting; can I call you back?"), then resume the interrupted conversation. If you were expecting a long and important call, you might instead reschedule your colleague ("Can I find you in your office in 20 min?"), and then simply take the call. In either case, you have taken explicit measures to prepare to resume a task (your conversation with your colleague) that was interrupted. This particular example, in which a ringing phone signals an upcoming interruption, represents a pattern that seems to be quite general. Even in an emergency, like an alarm that is not simply a drill or a prank, one might first hurry to save all modified files and only then evacuate the building.

In this paper, we offer a simple task analysis of interruptions, focusing on the preparatory opportunity afforded by the *interruption lag*, or the interval between an alert (e.g. the phone ringing) and the interruption proper (e.g. the phone call). We then offer a theoretical basis for supposing that people do in fact engage in preparatory cognitive activity during the interruption lag, activity analogous to the social negotiation with your colleague in the example above. We then present an experiment that supports the prediction and helps to show *how* people prepare—prospectively, by encoding specific goals to achieve at time of resumption, and retrospectively, by rehearsing state information from the point of interruption. Finally, we discuss future research and applications suggested by these findings.

2. A task analysis of interruption

Fig. 1 presents a timeline of task interruption followed by task resumption. The primary task is ongoing when an alert occurs, indicating a pending interruption by a secondary task.[1] In terms of the telephone example, the primary task would be the conversation with a colleague in your office, the alert would be the phone ringing, and the secondary task would be the conversation with the caller. The time between the alert and the start of the secondary task is the *interruption lag*. Eventually, after seconds, minutes or longer of working on the secondary task, the operator completes it or suspends it and returns to the primary task. The time between leaving the secondary task and beginning the primary task is the *resumption lag*.

A number of variables can complicate this simple representation of the interruption/resumption process. For example, the operator may have more or less freedom to control the length of the interruption lag (McFarlane, 2002). In an office setting, for example, people show a strong tendency to reach a logical stopping point in the primary task, before attending to the interruption (Zijlstra et al., 1999; Cutrell et al., 2001). In general, the operator's first decision during the interruption lag may

[1] We use the terms primary task and secondary task for ease of reference, with the primary task being the one being interrupted and the secondary task being the one that interrupts. We make no assumptions about one of these being more important or urgent than the other.

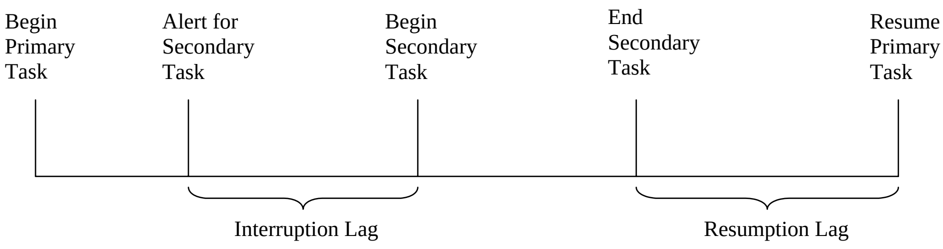

Fig. 1. The interruption and resumption process, involving a primary (interrupted) and a secondary (interrupting) task.

concern whether to switch immediately to the secondary task or to "wrap up" some aspect of the primary task first, in which case the interruption lag will be a bit longer. The alert itself can take any number of forms (visual, auditory, etc.), and may provide information about the urgency of the interruption and the type of action called for (Stanton and Edworthy, 1999). Appropriately interpreting an alert is critical because it might be the only indicator that multiple tasks require immediate operator action, for example in complex systems where system failures can cascade quickly (e.g. Three Mile Island; Rubinstein and Mason, 1979).

After the interruption, there is considerable variability in what it means to resume the primary task. It may be that the precise subgoal that was suspended during the interruption is the one resumed afterwards. For example, after a long phone call (in our scenario above), you may in fact chase down your colleague and resume the conversation. On the other hand, you may forget, and only later (if ever) remember the promise you made to find them in 20 min. Of course, forgetting a suspended goal like this could be catastrophic (see Latorella (1996) for a review of the effects of interruptions on commercial flight decks). Thus, the greater the importance of the interrupted subgoal, the more it matters whether the operator has somehow prepared to resume it. Finally, although interruptions usually disrupt performance of the primary task (Gillie and Broadbent, 1989; Zijlstra et al., 1999; Cutrell et al., 2001; McFarlane, 2002), they can also facilitate performance by increasing arousal or decreasing boredom (Speier et al., 1999).

Despite the rich set of variables that could affect how people process interruptions, the basic timeline illustrated in Fig. 1 seems to capture one set of fundamental opportunities and constraints. The basic constraint is the resumption lag. A rapidly growing literature on executive control indicates that switching among even simple tasks incurs an overhead in terms of response time (Rogers and Monsell, 1995; Allport and Wylie, 2000). Behaviorally, then, there will always be a lag in shifting from the secondary task back to the primary task. The basic opportunity is the interruption lag, a brief window in which to lay the cognitive groundwork for returning to the primary task and therefore reducing the resumption lag. The question we address next, from a theoretical perspective, is whether the cognitive system is in principle able to make use of the interruption lag to reduce the resumption lag. Then, after introducing our theoretical perspective, we turn to an empirical examination of these issues.

3. A theoretical framework for studying interruptions

A cognitive construct that seems to be highly relevant to how people process interruptions is what we will refer to as the *goal*, or an intention to perform some action in the future. For example, people seem quite capable of writing themselves "mental notes" to take up where they left off on an interrupted task, where such mental notes can be thought of as goals to resume that task at a particular point. The specific questions we address here concern the conditions necessary for being able to retrieve a goal at resumption, and the conditions necessary for being able to encode a goal during the interruption lag. We will focus on the use of goals retrieved from memory to resume the primary task, rather than the strategy of reconstructing goals from the environment (Simon, 1975; VanLehn and Ball, 1991), because it seems that people do in fact retrieve goals even when reconstruction is a viable option (Anderson and Douglass, 2001; Altmann and Trafton, 2002).

The *goal-activation model* (Altmann and Trafton, 2002) is our approach to analysing goal encoding and retrieval. The goal activation model is based on the hypothetical construct of activation of memory items, in particular, activation as construed in the ACT-R cognitive theory (Anderson and Lebiere, 1998). A basic processing assumption in this theory is that when central cognition queries memory, memory returns the item that is most active at that instant. Activation thus represents relevance to the current situation. To capture the relevance of any particular item, the memory system computes that item's activation from both the item's history of use and from its associations to cues in the current mental or environmental context. In Bayesian terms, the logic is that history of use and current context together serve to predict the current relevance of that item (Anderson and Milson, 1989; Anderson and Schooler, 1991). In functional terms, the implication that we pursue here is that the cognitive system should be able to exploit the predictive computations of the memory system to overcome decay and keep certain information active for use in the future.

In our model, the history-of-use factor is captured by the equation below, which is adapted from the base-level learning equation in ACT-R.[2]

$$\text{Activation} = \ln\left(\frac{n}{\sqrt{T}}\right) \tag{1}$$

Eq. (1) computes activation as a function of frequency of use. The quantity n is the total number of times the memory item has been retrieved in its lifetime, and T is the

[2] Eq. (1), a simplified version of ACT-R's base level learning equation, appears to model behavior reasonably well at time scales of up to a few tens of seconds. For longer time scales, the regular base-level learning equation may provide a better account. The regular base-level learning equation in ACT-R is $m = \ln\left(\sum_1^n t_j^{-d}\right) + \beta$ where t_j is the time lag between retrieval j and the present, d is a decay parameter that is typically set to 0.5, and β is an initial-activation parameter that we set to 0. With these settings the equation simplifies to $m = \ln(2n/\sqrt{T})$ on the assumption that the t_j are evenly spaced; T represents the interval from the very first retrieval to the present (i.e., the lifetime of the trace). For convenience, we then omit a factor of ln(2) to scale activation to be zero when $n = T = 1$, at the start of the trace's lifetime.

length of this lifetime, from the item's initial representation in the system to the present. Thus, as time passes without use of an item, T for that item grows while n does not, producing decay (a decrease in activation). Conversely, if concentrated use of an item causes n to grow rapidly, activation will increase. These dynamics are similar to the usual algorithm for deciding which item to replace in cache memory in a computer: Recency is a good predictor of future need, so the least recently used item in cache is the one replaced when a new item is brought in from slower memory.

These dynamics suggest two ways in which people might use the interruption lag to prepare to resume the primary task: *prospective goal encoding* and *retrospective rehearsal*. Both mechanisms exploit the fact that deliberate use of an item in memory increases its activation, as per Eq. (1). Prospective goal encoding, which we envision as a key mechanism behind prospective memory (Goschke and Kuhl, 1993; Brandimonte et al., 1996; Patalano and Seifert, 1997), is an important functional component of our simulation of performance on the Tower of Hanoi task (Altmann and Trafton, 2002). In this task, the problem solver must often suspend a goal while he or she "looks ahead" mentally to determine how to proceed. After this look-ahead planning, the problem solver often retrieves the suspended goal from memory to see if progress on that goal is now possible (Anderson and Douglass, 2001; Altmann and Trafton, 2002). In our simulation, successful retrieval depends critically on the system having built up that goal's activation during what was effectively the interruption lag. In phenomenological terms, this activation buildup is simply the sense one has of focusing on or "paying attention to" a particular goal until it comes to represent one's current mental set (Posner and Boies, 1971). In terms of Eq. (1), activation buildup is implemented in terms of several consecutive cycles of incrementing n in a small amount of T. Based on theoretical constraints on system cycle time (roughly 100 ms/cycle), we estimated that time to encode a goal in the Tower of Hanoi data was roughly a second or two (Altmann and Trafton, 2002). This was time enough to raise activation to the point where the encoded goal would later overcome proactive interference from old goals that are no longer relevant.

The second form of preparation that one might expect to find during the interruption lag is retrospective rehearsal. Informally, the distinction between retrospective rehearsal and prospective goal encoding parallels the distinction between "Now what was I doing?" and "Now what was I about to do?" In essence, if one can answer the former question with reference to memory, this may be useful in answering the latter question about what to do next. The target information would simply be state information from the primary task at the moment of interruption, for example the last action taken before the alert occurred. At resumption, a memory for this action would then allow selection of the next logical action, based on procedural knowledge about action sequences appropriate for that task environment. In terms of our model, the mechanism for rehearsal is simply the same as that for encoding, namely deliberate use of the memory to increase n in Eq. (1).

Our expectation concerning rehearsal is based on the pervasiveness of rehearsal as a strategy in situations that require temporary maintenance of information in working memory (e.g. Rundus, 1971). Indeed, we are not the first to suggest that rehearsal might be a factor in interruption management (Gillie and Broadbent, 1989;

Czerwinski et al., 1991; Storch, 1992), though previous studies have claimed that it is not a factor. Gillie and Broadbent (1989), for example, suggest that "having the opportunity to rehearse the point of interruption does not automatically offer protection against the disruptive effect of an interruption" (p. 348). Similarly, Czerwinski et al. (1991) found no evidence that providing opportunities to rehearse facilitated performance of the primary task. However, these studies did not include a direct measure of rehearsal (such as verbal protocol data), so it is difficult to tell whether manipulations of opportunity for rehearsal had their intended effect. It could have been, for example, that rehearsal occurred regardless of whether opportunities were explicitly available, in which case an opportunity manipulation would have had no effect. It could also be that the measures used to tap the disruptiveness of interruptions were not sensitive enough to reveal differential effects of rehearsal. Thus, it seems premature to rule out the possibility that rehearsal affects retention of primary-task information, especially given that rehearsal is such a common retention strategy.

The behavioral prediction associated with both forms of preparation (prospective goal encoding and retrospective rehearsal) is that they are time-consuming, and thus require a temporal window in which to execute. The interruption lag seems the most likely candidate for such a window, though we cannot rule out that preparation will intrude on the secondary task as well. Thus, in the following experiment, our hope is to manipulate interruption lag so as to find evidence for preparation when a lag is present, and little or no evidence of preparation when a lag is absent. If this manipulation of interruption lag were successful, we would then expect to find that preparation predicts greater efficiency when resuming the primary task after the interruption.

The second factor that contributes to the activation of a goal in our model is priming from contextual cues to which a goal is linked. When such cues are attended to in the environment, they spread activation to any goals with which they are associated. This spreading activation is simply added to the activation produced by the history-based mechanism represented by Eq. (1). In our model, priming is necessary to retrieve a suspended goal, because that goal will have decayed during the interval of the interruption. Because the most active goal governs behavior, the most recent goal (associated with the secondary task, upon returning from an interruption) would continue to govern the system's behavior in the absence of priming of some other goal, because it has decayed less. Thus, during the resumption lag, the system must have access to appropriate cues to allow it to remember anything about the primary task. We do not manipulate the availability of cues in this particular experiment, but do expect the verbal protocol data we collect to reflect an emphasis on perceptually available information.

4. Experiment

Our goal with this experiment was to manipulate the opportunity to prepare to resume after an interruption, and then to measure the effects of any differences in preparation across conditions. To measure preparation, we recorded verbal protocols, on the view that goal encoding and rehearsal are sufficiently deliberate or controlled that they rise to the level of inclusion in verbal reports (Ericsson and Simon, 1993). Participants in the Warning condition were given a visual alert followed by an interruption lag of 8 s. Participants in the Immediate condition were taken directly to the secondary task, with no warning and hence an interruption lag of zero. We expected that participants in the Warning condition would engage in more preparation and would resume the primary task more quickly than those in the Immediate condition. In particular, we expected that participants in the Immediate condition would resume more slowly because it would take them longer to retrieve the relevant goal, or because retrieval would fail more often and lead to more goal reconstruction. Because we were interested, for the purposes of this study, in measuring people's natural propensity to prepare, we provided no instruction that preparation might be a useful strategy.

4.1. Method

4.1.1. Participants

Participants were 17 NRL employees (14 men and three women) who volunteered for the study. They were assigned randomly to one of two conditions, with eight assigned to the Warning condition and nine to the Immediate condition. Ages ranged from 18 to 63 with an average of 37. Completed education ranged from high-school (six participants) to doctorate (three participants). There were no differences between conditions in terms of average age, education or gender.

4.1.2. Task and materials

The primary task was a complex resource-allocation task (Brock and Trafton, 1999) that we will refer to as the *tank task*. Resources in this task are a set of tanks (heavy and light) and their associated munitions, fuel and fuel tanks. Participants are assigned a mission to use these tanks to attack and destroy three destinations, or locales. Locales can defend themselves, and to attack a destination a tank must overcome obstacles (like a river or sand dunes) that can use up extra fuel or cause the tank to crash. The effects of these defenses and obstacles are stochastic, so planning can only be approximate. Resources are subject to constraints, in that tanks can hold only limited amounts of fuel or munitions, and resources (including tanks) are costly. Thus, the challenge for the operator is to plan successful missions while minimizing the use of resources.

The operator's interface uses a standard point-and-click paradigm and is composed of several dialog-box style windows in which the operator can review and select destinations, equip and allocate tanks, and subsequently evaluate the success or failure of a mission. The operator also has access to a map showing the location of each locale. Fig. 2 shows a screen snapshot of the tank task.

As we noted earlier, contextual cues may play an important role in the process of resuming the primary task after an interruption. However, in this experiment we wanted to minimize the possibility that participants would manipulate

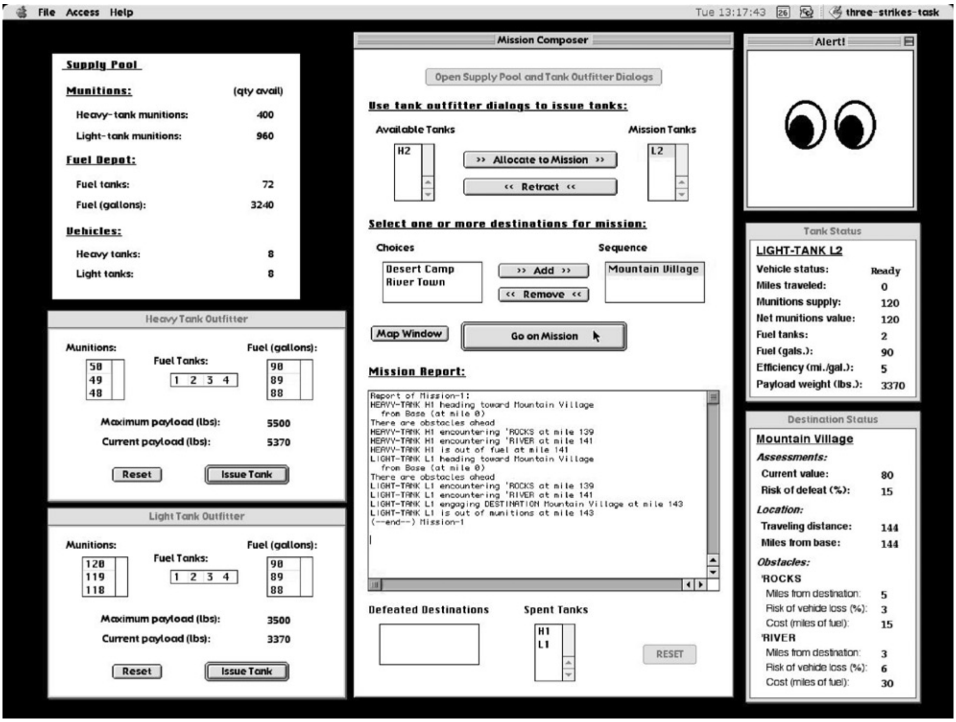

Fig. 2. A screen snapshot of the tank task. The upper left window shows available resources. Below this window are two others that allow the operator to outfit heavy (middle) or light (lower) tanks. The large, center window provides information about locales, missions, and the outcome of missions. The middle right window shows information about a selected tank, and the lower right window shows information about a selected locale. The upper right window is the alert window as seen by participants in the Warning condition. It becomes visible and flashes briefly to signal an alert to Warning participants, and remains visible during the interruption lag.

environmental cues in ways that we would not be able to detect. In particular, we wanted to prevent use of the mouse cursor and the active window to signal where to resume after the interruption. Thus, when the tank task interface was restored after an interruption, the software always placed the mouse cursor in the same position (the upper left-hand corner of the screen), regardless of where it had been before the interruption. We expected that this would prevent the operator from using the cursor to mark what he or she was working on or wanted to work on next (cf., Czerwinski et al., 2000; Cutrell et al., 2001). The software also deactivated (or defocused, or "back-grounded") all windows of the tank task, so that during resumption the operator would not be able to infer from perceptual cues which window he or she had been working in. Again, we expected that this would prevent the user from finding the window that was "in front" and using that information as an environmental cue to facilitate the resumption process. Note that our goal was not to eliminate the use of all environmental cues, but only those that could be

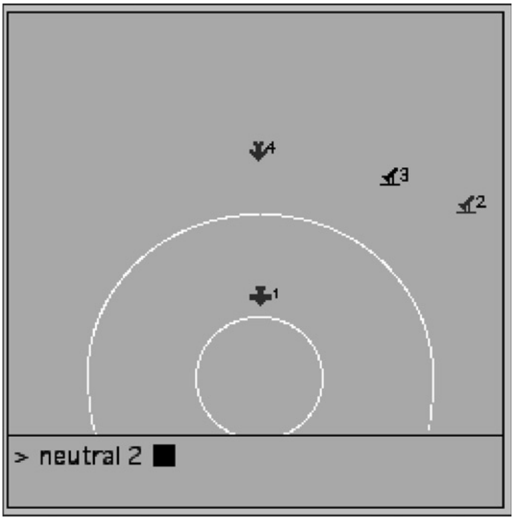

Fig. 3. The tactical assessment task. Participants had to identify whether each icon was neutral or hostile. The original is in color.

exploited by the operator to direct his or her attention to a particular region of the interface during resumption.

The secondary task is a simulated tactical assessment task that has been used extensively in other studies (Ballas et al., 1999; Brock et al., 2002a, b). In the tactical assessment task, approaching objects (or "tracks") must be classified as hostile or neutral based on a set of rules for their behavior. The objects appear as numbered icons representing fighters, cargo planes and surface-to-air missile sites that move down a radar-screen-like display as the operator's own aircraft supposedly travels forward. An automation component assists the operator by designating each object as hostile (red), neutral (blue) or unknown (yellow) when an assessment cannot be made. The operator's task is to indicate whether a yellow object is hostile or neutral on the basis of its behavior and, otherwise, to confirm the automated assessments of red and blue. Tactical decisions are entered with two left hand keystrokes on the numeric keypad of a standard keyboard, with the first keystroke indicating hostile or neutral and the second identifying the track number. Fig. 3 shows a screen snapshot of the tactical assessment task.

4.1.3. Design and procedure

Participants were trained for approximately 1 h on how to perform the two tasks singly and in combination, and on how to give a talk-aloud protocol. Participants returned the next day for a testing period again lasting approximately 1 h. Participants gave talk-aloud protocols throughout the testing period (similar to Trafton et al., 2000).

There was one between-participants factor (Warning or Immediate) and one within-participants factor (session). In the Warning condition, the interruption lag began with a warning indicating that the participant was about to be taken to a secondary task. The warning took the form of a window popping up and flashing for 600 ms, with a pair of eyes inside it. As shown in Fig. 2, the alert window appeared in the upper right-hand corner of the display. We used this kind of alert because evidence from pilot studies suggested that other kinds (in particular, a count-down timer) could engage the participant's attention for the duration of the interruption lag, possibly to the detriment of preparation. Once the alert window appeared, the keyboard and mouse were frozen and no further actions could be taken to advance the primary task. The interruption lag in the Warning condition lasted 8 s. In the Immediate condition there was no alert window; participants were taken immediately to the secondary task.

Within the experimental testing period, there were three sessions of approximately 20 min each. Sessions ended after participants had spent 20 min on the tank task. The time spent on interruptions was not part of this 20 min.

During each session, a participant was given 10 interruptions. Ten temporal "markers" were first randomly scheduled for the session. Then, the first mouse click or keystroke that occurred after one of these markers triggered an interruption. In other studies, interruptions were triggered based on elapsed time (i.e., an interruption every 30 s) or at random intervals (Gillie and Broadbent, 1989; Zijlstra et al., 1999; Czerwinski et al., 2000; Cutrell et al., 2001). We hoped to make interruptions more uniformly disruptive by linking them to the performance of an action, on the assumption that actions are often taken in service of active goals.

During an interruption, participants performed the tactical assessment task for about 30 s. After the interruption, they were immediately returned to the tank task.

4.1.4. Measures

Keystroke and mouse-click data were recorded for every participant. The primary measure of interest was how fast people were able to resume the primary task after being interrupted by the secondary task. To measure this quantity, we computed resumption lag as the interval from the moment the tank task interface was restored following the interruption to the first mouse click or key press a participant made to resume the primary task. However, we also computed a derived measure of time to resume the primary task. Pilot studies showed that individual differences in time to perform the task were quite large. To reduce these individual differences as a source of variance, we first determined the average inter-click lag, or time between actions in the tank task. The inter-click lag was calculated by taking the average time elapsed between mouse clicks/keyboard actions over all mouse clicks/keyboard actions in a session. Inter-click lags were quite variable between participants, but were reasonably stable within participants. We then computed a *disruption score* by subtracting the average inter-click lag of each participant from the average resumption lag for that participant. Thus, each session for each participant had a resumption lag, an inter-click lag and a disruption score.

Finally, we used the talk-aloud protocols to determine the amount, timing and type of preparation, both during the interruption lag in the Warning condition and during the secondary task in both conditions.

4.2. Results and discussion

Three resumption lags were over 10 standard deviations from the session mean, so were excluded from all analyses. Furthermore, software problems prevented seven resumption lags from being properly recorded. These exclusions accounted for less than 2% of the collected data. One Immediate participant's verbal protocol was not properly recorded, but none of that participants' other measures were affected so all were included in the analysis.

4.2.1. Kinds of preparation

Table 1 shows excerpts from the verbal protocols, to illustrate the different kinds of preparation.[3] Excerpts 1 and 5 are coded as retrospective because they involve rehearsal of state information from the primary task prior to the interruption. In excerpt 1, the state information is the distance that a particular tank needed to travel to reach a destination (the refinery). This is the only excerpt in Table 1 that is taken from the Immediate condition; it appears that in this case, preparation was deemed important enough to be attempted concurrently with the secondary task. Excerpt 5 is retrospective because it reflects a constraint (105 gallons of fuel needed) implied by a travel distance (210 miles) visible on the display at the time of the alert. Excerpts 2–4 are coded as prospective because they explicitly specify goals to be achieved at resumption time. Excerpt 2, marked by the comment "when I get back" (meaning back to the primary task), is a particularly clear instance of prospective goal encoding.

The excerpts are also coded along a second dimension. Excerpts 1–3 are coded as external because they refer to perceptually available information. Excerpt 2 is external because the munitions and payload of this particular tank were visible on the display (the "Heavy Tank Outfitter" in Fig. 2). Excerpts 4 and 5, in contrast, are coded as inferred because they involve intermediate products. In excerpt 4, the inference is that *if* the tank arrives at its destination, then it will need 40 munitions to attack that destination; the value 40 is not visible on the display. Excerpt 5 shows an inference (that each fuel tank needs to be filled with 105 gallons of fuel) being initiated just before the alert and then being completed during the interruption lag.

Of all preparation utterances, 62% involved prospective goal encoding, whereas only 38% involved retrospective rehearsal of state information from the time of interruption. This difference was reliable, $\chi^2(1) = 8.2$, $p < 0.05$. On the other dimension, 79% of preparation utterances referred to perceptually available information, whereas 21% referred to intermediate products, $\chi^2(1) = 33.6$, $p < 0.001$. Note that because there were relatively few rehearsals in the Immediate condition (see Table 2 and the discussion in the next section), we collapsed rehearsals across conditions for these analyses.

Thus, the most common form of preparation involved prospective encoding of a goal to be achieved at resumption, with the goal formulated in terms of specific

[3] One of the authors coded all instances of preparation. A second author coded a 10% subset and agreed 90% of the time, $\kappa = 0.85$, $p < 0.001$.

Table 1
Utterances from five preparation instances and how the utterances were coded

Utterance before alert	Utterance continued after alert	Code
1. Defeatable. Uh. 150, 150 miles	150 miles from the refinery. *That'll be hostile.* 150 miles to the refinery. *Neutral. That'll be hostile 3. Neutral 1. 4. 1 is neutral. 4 is neutral…*	Retrospective external
2. I'm already over payload. Hmm	Okay, adjust the munitions to get below the payload level when I come back.	Prospective external
3. Train depot is not much risk. Okay. So we have, so 275. And then uh light tanks	Issue tank, a little one I guess.	Prospective external
4. Its munitions value was, what was its munitions value? Um, it was 3 per thing. So	It's gonna need at 120, defeated, it's gonna need about 40 munitions. Um, that assumes of course that it gets there.	Prospective inferred
5. 210 miles figuring about. That will allow 210 miles. 2 miles per gallon. So that's	105 gallons. 105 gallons per tank.	Retrospective inferred

Note: Retrospective means that the utterance refers to state information from the primary task prior to the interruption. Prospective means that the utterance refers to a goal to be achieved at resumption. External means that the referent of the utterance is visible on the display. Inferred means that the referent is an intermediate product (see text). Instance 1 is from the Immediate condition; utterances in italics refer to the secondary task.

Table 2
Average number of preparation instances, by session and warning condition

Session	Condition	Preparation instances (out of 10)
1	Warning	4.0
	Immediate	0.4
2	Warning	4.6
	Immediate	1.3
3	Warning	4.6
	Immediate	1.3

values or items visible on the display for the primary task. It seems likely that the marked preference for prospective encoding, in contrast to retrospective rehearsal, reflects an information-processing efficiency of some kind that warrants further study.

4.2.2. Effects of warning on preparation

We then examined whether participants prepared differentially across the Warning and Immediate conditions. Each interruption was counted as an opportunity to prepare, and was scored as an *instance* of preparation if the protocol contained any evidence of preparation (prospective goal setting or retrospective rehearsal) for that particular resumption. Preparation could occur either during the interruption lag or during the secondary task. The data are presented in Table 2. Warning participants prepared an average of 13.5 times out of 30 opportunities, whereas Immediate participants prepared only 3.1 times out of 30 opportunities, $\chi^2(1) = 53.5$, $p < 0.001$.[4] Ninety-five percent of preparation utterances in the Warning condition occurred during the interruption lag, whereas all preparation utterances in the Immediate condition necessarily occurred during the secondary task.

It is important to remember that evidence from a protocol may be sufficient but is not necessary to signal the occurrence of preparation activities. It could be, for example, that participants performed more preparatory activities that they did not verbalize. Our analyses assume only that protocol evidence measures *relative* proportions of preparatory activity across conditions.

Thus, our manipulation was successful in that Warning participants took advantage of the interruption lag to prepare. As we mentioned earlier, we gave no explicit instructions as to what participants should do during the interruption lag, so it appears that there was a natural inclination to use the interruption lag to prepare. Whether preparation was beneficial to task resumption is a separate issue we address later in this section.

4.2.3. Disruption of the primary task

Before examining the effects of preparation on task resumption, we first asked whether interruptions actually disrupted the primary task. To measure disruptiveness, we compared resumption lag to inter-click lag. Both lags measure in part the time to plan the next action, and are comparable in that they measure time from one action to the next. Thus, if resumption lag is longer than inter-click lag, then we can infer that planning the first action after an interruption takes longer than the baseline planning time for an action.

Resumption lag and inter-click lag are shown in Fig. 4, separated by session. Resumption lag was in fact longer than inter-click lag, $F(1,16) = 22.2$, $\text{MSE} = 6.5$, $p < 0.001$. There was also a general speedup across sessions, $F(1,16) = 12.3$, $\text{MSE} = 3.6$, $p < 0.005$, suggesting that people improved as they became more familiar with the task. There was a marginally significant interaction between session and type of lag, $F(2,32) = 2.4$, $\text{MSE} = 1.87$, $p = 0.07$, suggesting that people improved more rapidly on the processes performed during the resumption lag than on processes performed during the inter-click lag.

Thus, the data suggest that interruptions were disruptive, in that resumption lag was longer than a baseline measure of time to formulate the next action within the tank task. We also found a general practice effect, especially for the processes involved in resuming the primary task.

[4] The variances were unequal across conditions for this measure, $\text{Levene}(1,14) = 7.0$, $p < 0.05$, so we used a non-parametric statistic.

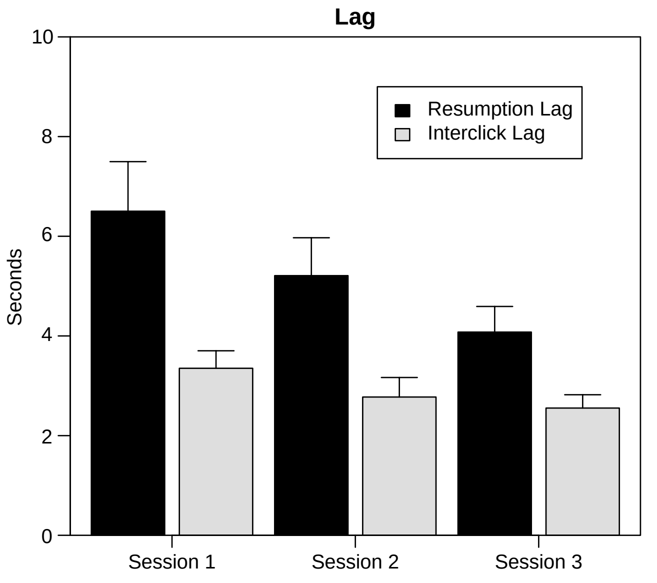

Fig. 4. Resumption lag and inter-click lag, separated by session. Error bars are standard error of the mean.

Table 3
Average RT and accuracy in the secondary task (standard deviations in parentheses)

Session	Condition	RT (s)	Accuracy
1	Warning	3.3 (0.9)	82% (0.14)
	Immediate	3.2 (1.1)	88% (0.09)
2	Warning	2.6 (1.0)	88% (0.20)
	Immediate	3.0 (1.1)	87% (0.11)
3	Warning	2.5 (0.9)	90% (0.09)
	Immediate	2.7 (1.0)	88% (0.16)

4.2.4. Disruption of the secondary task?

To determine whether providing participants an opportunity to prepare their goals affected performance generally, and not just resumption of the primary task, we examined whether having a warning affected performance on the secondary (tactical assessment) task also. Table 3 shows average reaction time (RT) and accuracy for classifying individual targets. Across sessions, participants became faster at classifying targets, $F(1,15) = 11.0$, $\text{MSE} = 0.287$, $p < 0.006$, and also became marginally more accurate, $F(1,15) = 3.0$, $\text{MSE} = 0.003$, $p = 0.10$. However, the main effect of having a warning was not reliable, nor was the interaction between session and warning condition, $p > 0.15$ in both cases. Thus, the interruption lag did not have a generalized effect on performance overall. This is important, because it helps to narrow the interpretation of any effects of the type of warning on resumption lag. That is, if resumption is facilitated in the Warning condition, then this is not simply because these participants had more rest (say), because in that case we would have expected improved performance on the secondary task as well.

4.2.5. *Effects of warning on task resumption*

Finally, we examined the effects of the warning manipulation on task resumption. Fig. 5 shows resumption lag separated by warning condition and by session. Fig. 6 shows disruption scores, again separated by warning condition and by session. Recall that the disruption score is the average resumption lag for each participant minus his or her average inter-click lag. (The disruption score is also simply the disruptiveness measure we analysed previously.) Thus, Figs. 5 and 6 show essentially the same pattern, but because the disruption score has less variance, it will be used in the following statistical analyses.

As Fig. 6 suggests, there was a linear decrease in disruption score across sessions, $F(1,15) = 7.4$, $\mathrm{MSE} = 2.6$, $p < 0.05$. However, warning condition and session also interacted in their effect on disruption score, $F(1,15) = 6.3$, $\mathrm{MSE} = 2.6$, $p < 0.05$. To explore this interaction, we computed linear contrasts separately for each warning condition. The disruption score in the Immediate condition decreased across the three sessions, $F(1,8) = 10.6$, $\mathrm{MSE} = 3.6$, $p = 0.01$. In contrast, the disruption score was unaffected in the Warning condition, $F(1,7) < 1$, $\mathrm{MSE} = 1.5$. These results

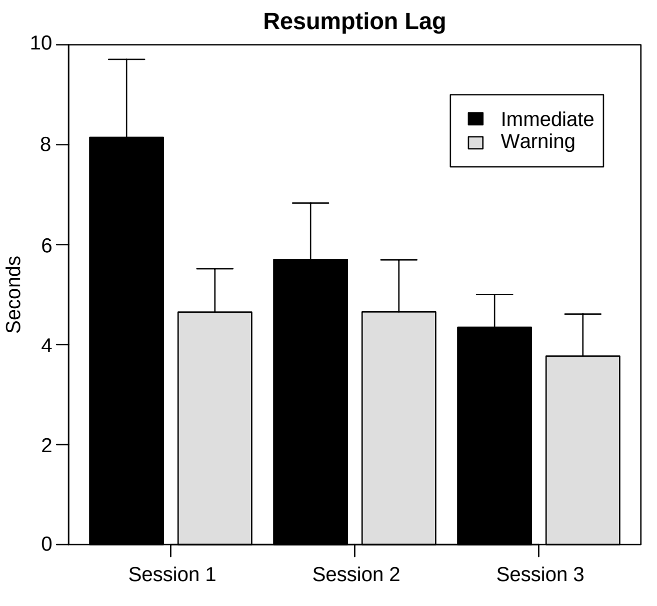

Fig. 5. Resumption lags for Immediate and Warning conditions, separated by session. Error bars are standard error of the mean.

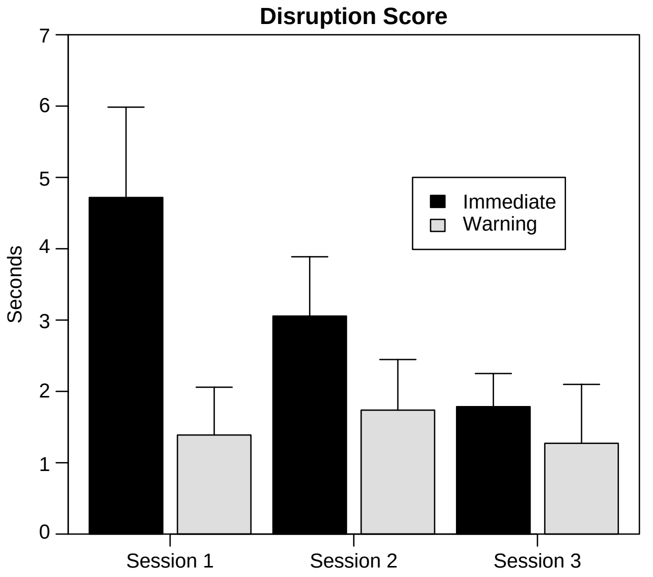

Fig. 6. Disruption score for Immediate and Warning conditions, separated by session. Error bars are standard error of the mean.

suggest that participants were able to improve their ability to resume the task, but only in the Immediate condition. It may be that in this condition, interruptions were sufficiently disruptive (as reflected in higher overall disruption scores) that participants were driven to adapt. An intriguing possibility is that, if the exact nature of these adaptations could be pinpointed, operator training could improve interruption management in situations where there is little time to think in response to an alert. When there is time to think, as in the Warning condition, this may mitigate the disruptive effects of the interruption to the point where the system is not driven to adapt.

Lastly, we examined effects of warning on disruption scores separately for each session. For session one, the Warning condition had a lower disruption score than the Immediate condition, $F(1, 15) = 5.0$, MSE = 9.4, $p < 0.05$. However, there was no effect of warning for session two, $F(1, 15) = 1.4$, MSE = 5.2, N.S. or session three, $F(1, 15) < 1$, MSE = 3.6. Thus, Immediate participants improved their ability to resume the primary task, to the point where they resumed as quickly as Warning participants. Disruption scores remained positive through the experiment (averaging slightly over a second in session three), but this may simply reflect the baseline cognitive switch cost associated with switching from one task to another (Rogers and Monsell, 1995; Allport and Wylie, 2000). In the final analysis, the cost of an interruption, measured in terms of time to resume the primary task, may have two components, one that can be overcome by strategic adaptations and one that represents deeper architectural constraints.

4.3. Summary

This experiment showed several things. The critical results are that interruptions were disruptive, and that participants did use the interruption lag to prepare to resume, producing smaller disruption scores. Preparation focused either on encoding a goal to be achieved after the interruption, or on information from the state of the primary task that could be used during resumption to infer the next step. Our theoretical interpretation is that preparation boosted the activation of whatever information was the target of the preparation, facilitating retrieval of this information from memory during resumption. Participants did not always explicitly prepare, even in the Warning condition, so it is possible that this effect would be strengthened by explicitly training participants to rehearse their current state or prospectively set a goal before beginning the secondary task. When participants did prepare, the target information was most often represented directly in the environment, which is preliminary support for the notion that environmental cues play an important role in task resumption.

The experiment also showed that people improve not only on the primary task (measured by the effect of session on inter-click lag), but also at task resumption (measured by the effect of session on disruption score for Immediate participants). The practice effect has been shown countless times (Crossman, 1959; Newell and Rosenbloom, 1981; Anderson, 1983; Trafton and Trickett, 2001), but we know of only one other example of interruptions becoming less disruptive over time (Hess and Detweiler, 1994).

5. General discussion

In this paper, we have developed a simple but general task analysis of task interruption and resumption, along with several measures (resumption lag and disruption score) that characterize the disruptive effects of interruption. We also discussed a theoretical framework for analyzing memory for goals, and presented some concrete theory-based predictions. We tested one prediction in particular, namely that preparing before an interruption allows people to resume their primary task more quickly. In the future it will be important to test directly the second basic prediction of our model, which is that environmental cues play a potentially large role in the resumption of an interrupted task. The experiment also provided preliminary evidence on the nature of the information that operators consider important in preparing to resume the interrupted task. In most cases, the focus is on prospective actions to be taken at task resumption (as opposed to state information from the point of interruption), and involves directly available perceptual information (as opposed to intermediate products).

In the following discussion, we first focus on practical implications of our theoretical and empirical results for designers of systems in which interruptions are frequent. We then address implications about the nature of the alert itself, and finally comment on the role of practice in coping efficiently with interruptions.

5.1. Implications for building alerts into systems

As we suggested earlier, interruptions have a complex structure in which many variables are likely to affect the ease with which the primary task is resumed. Nonetheless, we can argue based on our results that a brief interruption lag of a few seconds can facilitate resumption, by allowing an opportunity to lay the cognitive groundwork for returning to the primary task later. It is not obvious that a duration of a few seconds for the interruption lag should be enough to make a difference, but this is what our model predicts and what our experimental results support. Indeed, although the interruption lag in our Warning condition was eight seconds, the model suggests that a shorter lag of 1–2 s should serve equally well, especially if operators become skilled in using the interruption lag to full advantage. A shorter interruption lag would also improve the cost–benefit ratio of the interruption lag. (The 8-s lag in our Warning condition saved only four seconds at resumption time, and less in later sessions once Immediate participants had adapted; Fig. 5.) Of course, the optimal duration of the interruption lag may also depend on other parameters of the task environment. One approach to assessing these parameters is simply to have the alerting system try to predict from experience how long the interruption lag should be (Horvitz, 1999). Alternatively, it may be best to hand the operator control of the interruption lag, when this is possible, as the operator could then complete key goals prior to the interruption proper (Zijlstra et al., 1999; Cutrell et al., 2001) and thus reduce the burden on efficient resumption.

A second recommendation concerns retrieval cues that facilitate retrieval of primary task information at resumption. Our model makes the strong prediction that retrieval cues are necessary to be able to retrieve information about the primary task at the point of resumption. Such information will have decayed during the secondary task and thus will be less active than information from the secondary task, and hence unavailable based on the past-use component of activation (Eq. (1)). Our results indicate that people do make use of environmental cues in preparing to resume, despite the fact that we prevented use of candidate cues like cursor position in the primary task. In future studies it will be important to investigate the factors that make a good retrieval cue, the extent to which operators search out such cues on their own, and whether cues must be available both during the interruption lag (to be encoded with a prospective goal) and at resumption.

In summary, we suggest that an effective mechanism for managing the disruptiveness of interruptions should provide the operator with a brief opportunity to prepare to resume the interrupted task later, and with salient retrieval cues associated with primary-task information.

5.2. The nature of the alert

Our results also raise questions about the nature and time course of the alert itself. In our experiment, the alert was window flashing for a few hundred milliseconds, but longer alerts are certainly possible, and may in fact be used to communicate additional information to the operator. For example, a tone may change pitch gradually as the interruption approaches, to signal the remaining duration of the interruption lag. One might think that such diagnostic information could only help the operator to prepare to deal with the interruption. However, our results suggest that any alert that is more complex or longer than necessary has an associated opportunity cost in terms of preparing to resume. That is, if operators do prepare, as our theory and data suggest, then the best thing the interface can do to facilitate preparation is to avoid burdening the operator with extra information during the interruption lag. Thus, for example, in pilot studies of ours the alert was a countdown timer showing the number of seconds remaining until the start of the secondary task. This form of alert appeared to engage our participants, apparently distracting them from preparations to resume, and hence we simplified the alert in the current study. Note that we do not claim that the simplest alert is always the best, but only that a tradeoff exists, in which time to parse a complex alert will be deducted from time invested in preparing to resume. In a sense, the more information-rich the alert becomes, the more like an interruption it becomes, with all the attending disruptive consequences. These potential trading relations in alert complexity and length clearly need to be explored in future studies.

5.3. Practice effects

Another potentially important finding in the current experiment was the gradual decrease in resumption lag as Immediate participants gained practice with the task and with the process of being interrupted and resuming. Contrasting the Immediate and Warning conditions, it appears that training can compensate for lack of an interruption lag. The implication is that with practice, people are less disrupted by interruptions that occur without warning. If, based on our theoretical perspective, we can link this practice effect to improved processing of environmental cues, this could point the way to improved design of information displays and display-specific operator training protocols.

Acknowledgements

This research was supported in part by grant number N0001402WX20870 from the Office of Naval Research to Greg Trafton. We thank Paula D. Raymond, Sheryl Miller, Chris Monk, Bob Holt, Debbie Boehm-Davis and five anonymous reviewers for their comments on the manuscript.

References

Allport, A., Wylie, G., 2000. Task-switching, stimulus–response binding, and negative priming. In: Monsell, S., Driver, J.S. (Eds.), Attention and Performance XVIII: Control of Cognitive Processes. MIT Press, Cambridge, MA, pp. 35–70.

Altmann, E.M., Trafton, J.G., 2002. Memory for goals: an activation-based model. Cognitive Science 26, 39–83.

Anderson, J.R., 1983. The Architecture of Cognition. Harvard University Press, Cambridge, MA.

Anderson, J.R., Douglass, S., 2001. Tower of Hanoi: evidence for the cost of goal retrieval. Journal of Experimental Psychology: Learning, Memory, and Cognition 27 (6), 1331–1346.

Anderson, J.R., Lebiere, C., 1998. Atomic Components of Thought. Erlbaum, Mahwah, NJ.

Anderson, J.R., Milson, R., 1989. Human memory: an adaptive perspective. Psychological Review 96 (4), 703–719.

Anderson, J.R., Schooler, L.J., 1991. Reflections of the environment in memory. Psychological Science 2, 396–408.

Ballas, J.A., Kieras, D.E., Meyer, D.E., Brock, D., Stroup, J., 1999. Cueing of display objects by 3-D audio to reduce automation deficit. Proceedings of the Fourth Annual Symposium on Situational Awareness in the Tactical Environment, Patuxent River, MD.

Brandimonte, M., Einstein, G.O., McDaniel, M.A. (Eds.), 1996. Prospective Memory: Theory and Applications. Erlbaum, Mahwah, NJ.

Brock, D., Ballas, J.A., Stroup, J.L., 2002a. Using an auditory display to manage attention in a dual task, multiscreen environment. Proceedings of the Eighth International Conference on Auditory Display, Kyoto, Japan.

Brock, D., Stroup, J.L., Ballas, J.A., 2002b. Effects of 3D auditory display on dual task performance in a simulated multiscreen watchstation environment. Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society, Baltimore, MD.

Brock, D., Trafton, J.G., 1999. Cognitive representation of common ground in user interfaces. In: Kay, J. (Ed.), User Modeling: Proceedings of the Seventh International Conference. Springer-Wien, New York, NY.

Crossman, E.R.F.W., 1959. A theory of the acquisition of speed skill. Ergonomics 2, 153–166.

Cutrell, E., Czerwinski, M., Horvitz, E., 2001. Notification, disruption, and memory: effects of *messaging* interruptions on memory and performance. Interact 2001: IFIP Conference on Human–Computer Interaction, Tokyo, Japan.

Czerwinski, M.P., Chrisman, S.E., Rudisill, M., 1991. Interruptions in multitasking situations: the effects of similarity and warning (JSC-24757). Lyndon B Johnson Space Center, Houston, TX.

Czerwinski, M.P., Cutrell, E., Horvitz, E., 2000. Instant messaging and interruption: influence of task type on performance. OZCHI 2000 Conference Proceedings, Sydney, Australia.

Ericsson, K.A., Simon, H.A., 1993. Protocol Analysis: Verbal Reports as Data, 2nd Edition. MIT Press, Cambridge, MA.

Gillie, T., Broadbent, D.E., 1989. What makes interruptions disruptive? A study of length, similarity, and complexity. Psychological Research 50, 243–250.

Goschke, T., Kuhl, J., 1993. Representation of intentions: persisting activation in memory. Journal of Experimental Psychology: Learning, Memory, & Cognition 19 (5), 1211–1226.

Hess, S.M., Detweiler, M.C., 1994. Training to reduce the disruptive effects of interruptions. Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting, Vol. 2, pp. 1173–1177.

Horvitz, E., 1999. Principles of mixed-initiative user interactions. Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '99) Pittsburgh, PA.

Latorella, K.A., 1996. Investigating Interruptions: Implications for Flightdeck Performance. PhD Thesis. State University of New York, Buffalo, NY.

McFarlane, D.C., 2002. Comparison of four primary methods for coordinating the interruption of people in human–computer interaction. Human–Computer Interaction 17 (1), 63–139.

Newell, A., Rosenbloom, P.S., 1981. Mechanisms of skill acquisition and the law of practice. In: Anderson, J.R. (Ed.), Cognitive Skills and Their Acquisition. Erlbaum, Hillsdale, NJ, pp. 1–55.

Patalano, A.L., Seifert, C.M., 1997. Opportunistic planning: being reminded of pending goals. Cognitive Psychology 34, 1–36.

Posner, M.I., Boies, S.J., 1971. Components of attention. Psychological Review 78, 391–408.

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

Rubinstein, E., Mason, J.F., 1979. The accident that shouldn't have happened: an analysis of Three Mile Island. IEEE Spectrum 16, 33–42.

Rundus, D., 1971. Analysis of rehearsal processes in free recall. Journal of Experimental Psychology 89, 63–77.

Simon, H.A., 1975. The functional equivalence of problem solving skills. Cognitive Psychology 7 (2), 268–288.

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.

Stanton, N.A., Edworthy, J. (Eds.), 1999. Human Factors in Auditory Warnings. Ashgate, Aldershot, UK.

Storch, N.A., 1992. Does the user interface make interruptions disruptive? A Study of Interface Style and Form of Interruption (UCRL-JC-108993), Lawrence Livermore National Laboratory.

Trafton, J.G., Trickett, S.B., 2001. Note-taking as a strategy for self-explanation and problem solving. Human–Computer Interaction 16 (1), 1–38.

Trafton, J.G., Kirschenbaum, S.S., Tsui, T.L., Miyamoto, R.T., Ballas, J.A., Raymond, P.D., 2000. Turning pictures into numbers: extracting and generating information from complex visualizations. International Journal of Human–Computer Studies 53 (5), 827–850.

VanLehn, K., Ball, W., 1991. Goal reconstruction: how Teton blends situated action and planned action. In: VanLehn, K. (Ed.), Architectures for Intelligence. Erlbaum, Hillsdale, NJ, pp. 147–189.

Zijlstra, F.R.H., Roe, R.A., Leonora, A.B., Krediet, I., 1999. Temporal factors in mental work: effects of interrupted activities. Journal of Occupational and Organizational Psychology 72, 163–185.