Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computers in Human Behavior 22 (2006) 685–708
Computers in Human Behavior
www.elsevier.com/locate/comphumbeh

# On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state

Brian P. Bailey [a,*], Joseph A. Konstan [b]

[a] Department of Computer Science, University of Illinois, 201 N. Goodwin Avenue, Urbana, IL 61801, United States
[b] Department of Computer Science, University of Minnesota, Minneapolis, MN 55455, United States

Available online 7 February 2006

**Abstract**

This paper reports results from a controlled experiment ($N = 50$) measuring effects of interruption on task completion time, error rate, annoyance, and anxiety. The experiment used a sample of primary and peripheral tasks representative of those often performed by users. Our experiment differs from prior interruption experiments because it measures effects of interrupting a user's tasks along both performance and affective dimensions and controls for task workload by manipulating only the time at which peripheral tasks were displayed – between vs. during the execution of primary tasks. Results show that when peripheral tasks *interrupt* the execution of primary tasks, users require from 3% to 27% more time to complete the tasks, commit twice the number of errors across tasks, experience from 31% to 106% more annoyance, and experience twice the increase in anxiety than when those same peripheral tasks are presented at the boundary *between* primary tasks. An important implication of our work is that attention-aware systems could mitigate effects of interruption by deferring presentation of peripheral information until coarse boundaries are reached during task execution. As our results show, deferring presentation for a short time, i.e. just a few seconds, can lead to a large mitigation of disruption.


*Keywords:* Affect; Attention; Interruption; Performance; Workload

* Corresponding author. Tel.: +1 217 333 6106; fax: +1 217 244 6869.
*E-mail addresses:* bpbailey@uiuc.edu (B.P. Bailey), konstan@cs.umn.edu (J.A. Konstan).

0747-5632/$ - see front matter 
doi:10.1016/j.chb.2005.12.009

## 1. Introduction

Interruption is becoming an increasingly common and frequent occurrence in human–computer interaction. E-mail notifications (Jackson, Dawson, & Wilson, 2001), instant messages (Cutrell, Czerwinski, & Horvitz, 2000), and agent-initiated interactions (Maes, 1994) are all contributing to a burgeoning epidemic of interruption at the user interface.

While application-initiated interruptions can be a nuisance, empirical investigation is needed to further quantify and better understand the effects of interruption (McFarlane & Latorella, 2002), to test models of cognitive processes that can reliably predict those effects (Gillie & Broadbent, 1989; McCrickard, Chewar, Somervell, & Ndiwalana, 2003) and to motivate the need for and posit computational strategies that can mitigate disruption caused by interruption (Horvitz, Jacobs, & Hovel, 1999). The experiment reported in this paper contributes to these areas by further quantifying effects of interruption on both users and their tasks, comparing moments in a primary task sequence previously speculated to cause more or less disruption, and using the results to motivate the use of *temporal* strategies in attention-aware systems for mitigating the effects of interruption.

Our experiment utilized a commonly used paradigm where users perform primary tasks and are occasionally interrupted to perform peripheral tasks. It differs from previous interruption experiments because it measures effects along performance and affective dimensions, controls for task workload by manipulating only the time at which peripheral tasks are presented, and uses a theoretical basis for selecting the timings. Many prior experiments have measured the effects of interruption using some combination of completion time, error rate, and decision-making; including (Czerwinski, Cutrell, & Horvitz, 2000a; Kreifeldt & McCarthy, 1981; Latorella, 1998; Monk, Boehm-Davis, & Trafton, 2002; Speier, Valacich, & Vessey, 1999; Trafton, Altmann, Brock, & Mintz, 2003). In addition to completion time and error rate, our experiment measures effects of interruption along affective dimensions of annoyance and anxiety. Understanding these effects is particularly important for information workers and end users, as a few extra seconds or errors made on a task may be of low consequence relative to unnecessary increases in stress (Motowidlo, Packard, & Manning, 1986). While the interruption experiment conducted by Zijlstra, Roe, Leonora, and Krediet (1999) also measured affective dimensions, that experiment did not control for task workload. Our experiment controls for task workload by manipulating only the *time* at which peripheral tasks are presented to a user relative to a sequence of primary tasks.

We selected moments for interrupting a primary task sequence from theoretical arguments by Miyata and Norman (1986). The authors speculate that task boundaries represent more opportune moments for interruption because users have reduced mental workload at these moments. When a user completes a task, the executive system releases allocated resources, momentarily reducing workload before the cycle of allocation/deallocation occurs again for the next task. The boundary period from when resources are released to when resources are allocated for the next task should represent an opportune, or less disruptive, moment for interruption. Combining this hypothesis with a coarse structure of a primary task sequence, our experiment compares two moments for presenting peripheral tasks – *between* vs. *during* the execution of primary tasks.

For the experiment, we developed six categories of primary tasks: adding, counting, reading comprehension, image comprehension, selection, and registration, and two categories of peripheral tasks: reading news headlines and reasoning about market actions.

Primary tasks were designed to be representative of tasks often performed within larger interactive activities and peripheral tasks were designed to be representative of information often maintained at the periphery of user attention. Users were divided into an Experimental and Control group. In the Experimental group, a user performed a sequence of three tasks from a primary category. The execution of one primary task was interrupted with a reading task, another was interrupted with a reasoning task, and the remaining task was not interrupted. This procedure was then repeated for the five remaining task categories. In the Control group, users performed the same primary tasks except that peripheral tasks were now presented at the boundaries between those tasks, which controlled for task workload. We measured completion time, error rate, annoyance, and anxiety, and validated difficulty of the primary tasks using subjective ratings.

Results show that when peripheral tasks interrupt the execution of primary tasks, users require more time to complete the primary tasks, commit more errors across tasks, and experience more annoyance and anxiety than when those same peripheral tasks are presented between the primary tasks. When extrapolated over the millions of computer users whose tasks are being increasingly interrupted by applications executing outside their focus of attention, our results show that the collective impact of these interruptions could be quite remarkable.

Since only the timing of peripheral tasks was manipulated, our results motivate the use of a temporal strategy in attention-aware systems for mitigating effects of interruption. By deferring delivery of peripheral information until coarse boundaries are reached during task execution, the resulting interruption would have considerably less disruptive impact. As shown by our results, a *small* delay in delivery could lead to a *large* mitigation of disruption.

## 2. Related work

In this section, we discuss existing empirical evidence about the effects of interruption, describe how our experiment differs from prior interruption experiments, and discuss methods for selecting moments in a task sequence for interruptions to occur.

### 2.1. Effects of interruption

Several prior experiments have measured the effects of interruption using combinations of task completion time, error rate, decision-making, and affective state. For task completion time, Kreifeldt and McCarthy (1981) found that interruptions caused users to perform slower on calculator-based tasks due to the time needed to re-orient to the suspended task after the interruption. Gillie and Broadbent (1989) found that users perform slower on interrupted tasks and suggest that the disruptive effect depends on at least the cognitive load required by the interrupting task and its similarity to the primary task. McFarlane (1999) found that interrupting a continuous-attention primary task disrupts performance on both the primary and interrupting tasks. Trafton et al. (2003) found that interrupting a battlefield simulation task disrupts completion time and attributed the degradation to the cognitive system seeking to re-activate previously suspended goals (Altmann & Trafton, 2002). Extending prior work, our experiment compares how presenting peripheral tasks at different moments in a primary task sequence affects completion time on both the peripheral and primary tasks.

Czerwinski, Cutrell, and Horvitz (2000b) performed a series of experiments investigating the effects of interrupting primary tasks during their planning, execution, and evaluation phases. The authors found that users required more time to switch to the interrupting task when it was presented during the execution phase of a primary task. They also found that interruptions disrupt the evaluation phase more than the planning and execution phases. In subsequent experiments (Cutrell, Czerwinski, & Horvitz, 2001; Czerwinski et al., 2000a), the authors found that notifications disrupted performance on search tasks and caused users to request more reminders of the search goal. While these experiments tested different phases of a task, the boundary points between those phases were not tested. Using a coarse structure of a primary task sequence, our experiment compares the effects of presenting peripheral tasks between vs. during the execution of primary tasks. By showing meaningful differences using coarse moments, our work provides the necessary scaffolding for future work that uses analysis of a specific task to select and compare finer-grained moments for interruption.

For error rates, Latorella (1998) found that auditory and cross-modal interruptions cause pilots to commit more errors than other interruption strategies. Speier et al. (1999) found that interruptions during low task workload conditions facilitated decision performance, while interruptions during high workload conditions decreased performance. While decision-making was not measured, our experiment did measure how interrupting at different moments in a primary task sequence affects the number of errors committed.

Zijlstra et al. (1999) found that interruptions cause users to experience increased anxiety. However, in their study, task workload was not controlled between the 'not interrupted' and 'interrupted' conditions. That is, users in the not interrupted condition performed only the primary task while users in the interruption condition performed both primary and interrupting tasks. Thus, the increase in anxiety could be attributed to the difference in task workload, not to the act of being interrupted. Our experiment controls for task workload between timing conditions and measures annoyance as well as anxiety. For information workers and end users, the effects on affective state may have greater consequences than for task performance. A few extra seconds or errors made on a task may be of low consequence relative to unnecessary and prolonged increases in stress (Motowidlo et al., 1986; Narayanan, Menon, & Spector, 1999).

We previously reported a very small part of the results from this experiment in (Bailey, Konstan, & Carlis, 2001). The results reported in this paper include our first analysis of the error rates, a much more extensive analysis of primary and peripheral task completion time, ratings of annoyance, and changes in user anxiety. Also, this paper provides a thorough discussion of the results situated in related work and implications for the design of attention-aware systems.

### *2.2. Posited moments for interruption*

In prior interruption experiments, the structure of a primary task sequence has typically not been leveraged to select moments for interruption. Instead, primary tasks have been typically interrupted at random or periodic moments during execution. For example, Monk et al. (2002) interrupted users every few seconds during a VCR task, Trafton et al. (2003) interrupted users randomly during a tank battlefield simulation task, and Zijlstra et al. (1999) interrupted users at random points during document editing tasks. While appropriate in the contexts of the respective experiments, we believe that the structure of a primary task sequence must be considered when selecting moments for interruption and when interpreting results.

Miyata and Norman (1986) have speculated that task (and subtask) boundaries represent opportune moments for interruption since users should have reduced mental workload at those moments. They argue that when a user completes a task, the executive system releases the mental resources allocated for performing that task, momentarily reducing workload before the cycle of allocation/deallocation occurs for the next task. In studies of dual task behavior, for example, this may explain why users often work until a natural breakpoint in a task sequence before attending to peripheral information – they are intrinsically waiting for a boundary with reduced workload. While Czerwinski et al. (2000b) compared the effects of interruptions during different phases of a task, their experiments did not test boundaries between those phases.

Combining the reasoning of Miyata and Norman (1986) with a coarse structure of a primary task sequence, our study compares two moments for presenting peripheral tasks – *between* vs. *during* the execution of primary tasks. Interrupting the execution of a primary task should cause more disruption since mental resources have been allocated and are being actively engaged. In contrast, presenting peripheral tasks at the boundaries between primary tasks should cause less disruption since allocated resources should have been just released. While using task modeling techniques such as GOMS (Card, Moran, & Newell, 1983) or event perception theory (Zacks, Tversky, & Iyer, 2001), which both produce a hierarchical decomposition of a task, could be used to select finer moments for interruption, we wanted to first establish that there is a meaningful difference using coarser moments.

## 3. Method

We designed our experiment to answer the following questions:

- How much do the category of primary task, category of peripheral task, and timing of the peripheral task affect completion time on both the primary and peripheral tasks?
- How much do the category of primary task, category of peripheral task, and timing of the peripheral task affect errors committed on the primary tasks?
- How much do the category of primary task, category of peripheral task, and timing of the peripheral task affect user annoyance relative to both the primary and peripheral tasks?
- How much does the timing of peripheral tasks affect user anxiety? Note that while the category of primary and peripheral task may indeed cause transitory changes in anxiety, the paper-based instrument used to measure it does not easily allow for collecting finer-grained responses.

### *3.1. Subjects*

Fifty subjects (20 female) participated in the experiment. Subjects were between the ages of 18 and 40, had at least one year of computer experience, and were a mix of undergraduate and graduate students and working professionals.

### 3.2. Experimental design

The experiment used a 6 primary task (adding, counting, reading comprehension, image comprehension, selection, and registration) × 3 peripheral task (reading, reasoning, and none) × 2 timing (receiving peripheral tasks between vs. during primary tasks) mixed design. Primary task and peripheral task were within-subjects factors while timing was a between-subjects factor. A user completed a total of 18 primary tasks (3 tasks × 6 primary categories) and 12 peripheral tasks (2 tasks × 6 primary task categories). For timing, users receiving peripheral tasks between primary tasks will be referred to as the *Control group* while users receiving peripheral tasks during primary tasks will be referred to as the *Experimental group*. A user was randomly assigned to the Control or Experimental group subject to the constraint that each group had an equal number of males (15) and females (10).

### 3.3. Primary tasks

As shown in Figs. 1a–f, six categories of primary tasks were developed:

- *Addition*. Four numbers, each consisting of four digits, were presented to a user. The numbers were right aligned in a 4-row × 1-column table. The user mentally added the numbers and then entered the sum into a text field positioned underneath the last number.
- *Counting*. A set of 40 words arranged in a 10-row × 4-column table was presented to a user. The 40 words were randomly chosen from a base set of six words, i.e., each of the six words was repeated in the table. The user counted the words in the table that matched a given target word chosen from the base set, and then entered this count into a text field.
- *Image comprehension*. A completed tournament bracket starting with eight teams was presented to a user. The user answered five questions regarding various outcomes of the tournament, e.g., which team lost in the second round or which team won by four points.
- *Reading comprehension*. A short narrative passage (∼7–10 sentences) was presented to a user. The user read the passage and then answered three questions regarding its content.
- *Registration*. Eight registration-style questions were presented to a user, e.g., name, age range, favorite music, etc. The user entered the requested information using a combination of three interaction formats; toggle sets, drop-down lists, and free-form text fields.
- *Selection*. A set of 40 words along with checkboxes arranged in a 10-row × 4-column table was presented to a user. The 40 words were randomly chosen from a base set of six words, i.e., each of the six words was repeated in the table. The user selected each word in the table that matched a given target word chosen from the base set.

We designed the primary task categories to be of varying difficulty, but of similar length, requiring about 30 s to complete. From a pilot study with five users, we refined the tasks to improve clarity and to roughly meet the target completion time. Task difficulty was later validated from user rankings collected as part of the experiment. Because a user would perform more than one task from each category, multiple sets of similar tasks were designed. The task screens were implemented using HTML and were rendered using a standard Web browser.


Fig. 1. An example task from each of the primary task categories: (a) adding; (b) counting; (c) reading comprehension; (d) image comprehension; (e) registration; (f) selection.

Although these primary tasks are not exhaustive, they are representative of tasks or subtasks that are often performed in broader interactive activities. For example, the adding and counting tasks are representative of subtasks performed in numerical spreadsheet applications. The image comprehension tasks are representative of inspecting and evaluating diagrams. The reading tasks are representative of reading email messages, web pages, or instant messages. The registration tasks are representative of filling in product or personnel forms. The selection tasks are representative of visually scanning search results or documents for matching targets. Our sample of primary tasks should support reasonable external validity for the results.

### 3.4. Peripheral tasks

As shown in Figs. 2a and b, two categories of peripheral tasks were developed:

- *Reading comprehension*. A short (3–5 sentences) news summary was presented to a user. The user read the summary and then selected the most appropriate title from among three choices. We collected the news summaries and their actual titles from an existing news site to enhance realism.
- *Reasoning task*. A realistic stock scenario was presented to a user. Each scenario consisted of a fictitious company's name along with the quantity, date, and price of shares that the user hypothetically purchased from that company. Each scenario also contained the price of the stock and a one sentence "news-flash" regarding the company. After analyzing the scenario, the user selected one of five trading actions; do nothing, buy a few more shares, buy many more shares, sell a few shares, or sell all shares.

The peripheral tasks were designed to last approximately 20 s and were refined based on a pilot study with five users. Because a user would receive more than one peripheral task from each category, we designed multiple sets of similar tasks. These peripheral task categories were selected because they are representative of notifications that users often receive (Maglio & Campbell, 2000; McCrickard, Catrambone, Chewar, & Stasko, 2003) and because these tasks tap both language comprehension and analytic processing resources (Wickens, 1984).

### 3.5. Hardware/software

The experiment was conducted on a standard desktop PC running MS Windows. Primary and peripheral tasks were implemented using HTML. JavaScript was used to

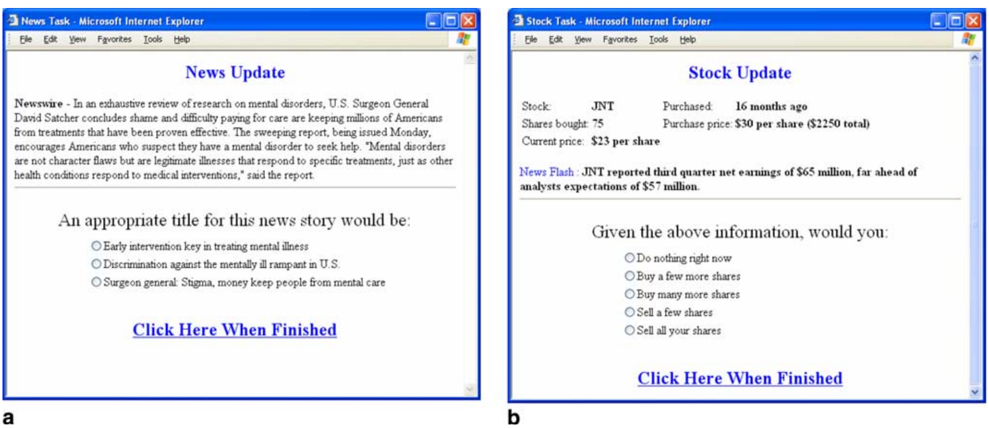


Fig. 2. An example task from each of the peripheral task categories: (a) reading news headlines; (b) reasoning about market actions.

implement the dynamics of the experiment such as randomizing, sequencing, and displaying tasks as well as logging performance data. Lotus ScreenCam was used to record a user's screen interaction.

### 3.6. Procedure

When a user arrived at the lab, we went through an informed consent process. The user was given a brief overview of the experiment and was then asked to complete the state anxiety form (Y-1) of the STAI (Spielberger, 1983), a commonly used and well-researched instrument for measuring anxiety in adults. While the trait form measures general anxiety, the state form measures anxiety at the present moment. The state form consists of 20 questions each with a four-point Likert response. The range of the anxiety scale is [20, 80].

Once complete, the user then moved to the computer to perform the tasks and was randomly assigned to either the Experimental or Control group:

- *Experimental group*. The experimenter described the peripheral tasks and a user performed several practice tasks. The first primary task category was then described and a practice task was performed by the user. The user was instructed to complete the primary and peripheral tasks as quickly as possible while maintaining accuracy on the tasks. The user was also instructed to immediately attend to a peripheral task whenever it appeared. After questions were answered, the experimenter left the testing area and the user performed three primary tasks. One primary task was interrupted with a reading task, another was interrupted with a reasoning task, and the remaining primary task was not interrupted and served as a control. If a peripheral task was presented, it was presented about halfway through execution of the primary task. Peripheral tasks were displayed in a modal window and covered the main work area of the primary task. To indicate completion of a primary or peripheral task, the user selected a ''finished'' link at the bottom of the task screen. For primary tasks, selecting the link would lead to the next task. This procedure was repeated for the five remaining task categories. The presentation order of the primary categories, tasks within each category, and peripheral tasks were randomized to minimize learning effects.
- *Control group*. The procedure for the Control group was similar to that of the Experimental group. The only difference was that peripheral tasks were now presented between primary tasks rather than during their execution. That is, a peripheral task was presented just after the primary task was complete, but before the next task began. A peripheral task was presented when the user selected the finished link of the primary task. As in the Experimental group, a peripheral task was presented for two of the three primary tasks from each category.

After performing the computer tasks, a user immediately completed another state anxiety form of the STAI. While we wanted a more fine-grained response for anxiety, it was not practical to administer this questionnaire after each condition in the experiment, nor did we have access to sensory-based hardware devices. After completing this form, the user filled out a paper questionnaire and rated the difficulty of the primary and peripheral task categories, the level of annoyance experienced when peripheral tasks were presented relative to each primary task category, and the level of annoyance experienced when performing the peripheral tasks. The entire experimental session lasted about 60 min.

### 3.7. Measurements

We measured task performance, error rate, task difficulty, annoyance, and anxiety. To measure task performance, we instrumented our system to log:

- *Time on primary task (TOT)*. This was the time spent performing a primary task, but did not include the time spent on a peripheral task, if presented.
- *Time on peripheral task (TOP)*. This was the time spent performing a peripheral task.

To measure errors, we reviewed the screen interaction videos and compared a user's answer to the correct answer for each task. Because registration tasks as well as the peripheral tasks did not have objectively right or wrong answers, these tasks were not included in the error analysis.

A pencil and paper questionnaire was used to measure the difficulty of primary tasks and the level of annoyance experienced when performing the tasks. The questionnaire asked users to:

- *Rank each primary task category according to difficulty*. A user ranked the categories from 1 (easiest) to 6 (hardest) by placing an 'X' in the appropriate location in a table and was allowed to rank two or more categories as being equally difficult.
- *Rate the level of annoyance experienced for each primary task category when a peripheral task was presented*. A user placed six labels representing the primary task categories along a continuous scale of annoyance ranging from 1 (not annoying) to 25 (intolerable).
- *Rate the level of annoyance experienced when performing each peripheral task category independent of a primary task category*. A user placed two labels representing the peripheral task categories along a continuous scale of annoyance ranging from 1 (not annoying) to 25 (intolerable).

Our annoyance scale was based on the 25 point scale used by Mital, McGlothlin, and Faard (1992) to measure levels of annoyance caused by noise in computer rooms. The scale was marked in nine equally spaced locations (1, 4, 7, 10, 13, 16, 19, 22, 25), identifying different levels of annoyance, including somewhat annoying (7), annoying (13), and very annoying (19). Since the scale had a wider range of values and was continuous, it allowed users to more easily position the task labels relative to each other on the scale.

To measure change in user anxiety, we asked a user to complete the state anxiety form of the STAI both *before* and *after* the experimental session. We subtracted the before measure from the after measure to compute change in state anxiety.

## 4. Results

In this section, we discuss results for task difficulty, completion time, errors committed, annoyance, and anxiety. Because Gender did not affect any of the dependent variables, the data was collapsed across this factor and will not be discussed further. For post hoc analysis, familywise error rates were controlled using the Bonferroni adjustment.

### 4.1. Task difficulty

Fig. 3 shows a graph of the difficulty rankings for each category of primary task between timing groups. Primary task had a main effect on rankings of difficulty, $F(5,240) = 43.60$, $p < 0.001$. Post hoc analysis showed that adding ($\mu = 5.06$) was ranked more difficult than counting ($\mu = 3.06$, $p < 0.001$), image comprehension ($\mu = 2.68$, $p < 0.001$), selection ($\mu$=2.28, $p < 0.001$), and registration ($\mu = 2.02$, $p < 0.001$). Reading ($\mu = 4.44$) was ranked more difficult than Counting ($p < 0.001$), image comprehension ($p < 0.001$), selection ($p < 0.001$), and registration ($p < 0.001$). Counting was ranked more difficult than registration ($p < 0.024$) and selection ($p < 0.027$).

Timing had a main effect on rankings of task difficulty, $F(1,48) = 5.43$, $p < 0.024$. While users in the Control group ($\mu$=3.42) ranked the primary tasks as being slightly more difficult to perform than users in the Experimental group ($\mu = 3.10$), rankings followed the same qualitative ordering. Users in the Experimental group may have ranked the primary tasks as being less difficult because they overly attributed part of the difficulty of performing the tasks to the interruptions. There were no interactions in the data.

The rankings did validate that users perceived the categories of primary tasks as having different levels of difficulty (workload). The categories in Fig. 3 were ordered hardest-to-easiest and this same order will be used in subsequent graphs containing the primary tasks.

### 4.2. Completion time

The following sections analyze completion time for the primary and peripheral tasks.

*Primary tasks*. Time on primary task between timing groups is shown in Fig. 4a. Primary task affected completion time, $F(5,235) = 107.67$, $p < 0.001$, however, they were not expected to take exactly the same amount of time. While timing did not affect completion time, $F(1,47) = 0.16$, $p < 0.69$, peripheral task did affect completion time $F(2,94) = 16.68$, $p < 0.001$. This main effect was due to a significant two-way interaction between peripheral task and timing, $F(2,94) = 4.79$, $p < 0.01$, suggesting an interruption effect in the Experimental group. Although this effect was not detected between the timing groups, it is likely because the impact of interruption on completion time in the Experimental

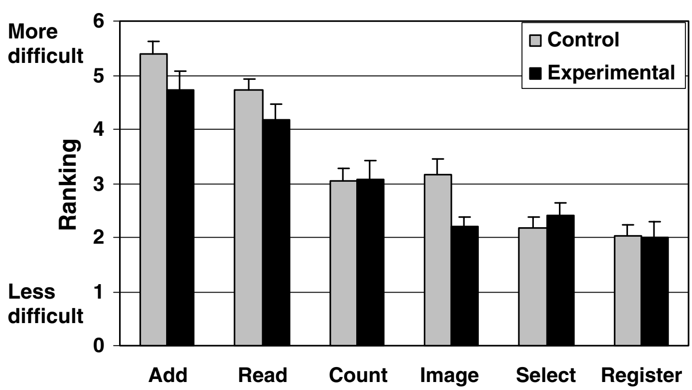


Fig. 3. Rankings of task difficulty ordered from hardest-to-easiest.

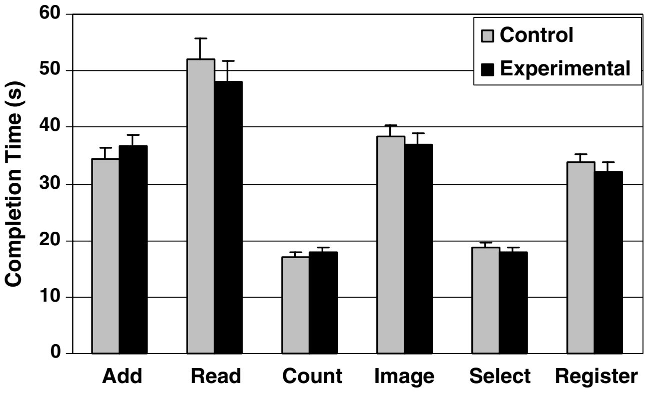


Fig. 4a. Time on primary tasks between groups.

group was not enough to overcome the increased variability in performance that is common between different users (Wickens, 1984).

In the Experimental group, peripheral task had a main effect on primary task completion time, $F(2,46) = 15.69$, $p < 0.001$. As shown in Fig. 4b, users required more time to complete primary tasks when interrupted with either a reading ($\mu = 33.60$ s) or reasoning ($\mu = 32.15$ s) task than when not interrupted ($\mu = 29.16$ s, $p < 0.001$, $p < 0.002$, respectively). While both categories of peripheral task caused a decrease in completion time, they did not affect it differently.

Interrupting the execution of primary tasks caused users to perform adding tasks 24% slower ($\mu_{int} = 39.06$ s, $\mu_{not} = 31.51$ s, $p < 0.001$), counting tasks 27% slower ($\mu_{int} = 19.31$ s, $\mu_{not} = 15.25$ s, $p < 0.001$), reading tasks 7% slower ($\mu_{int} = 49.32$ s, $\mu_{not} = 45.94$ s, $p < 0.04$), image comprehension tasks 10% slower ($\mu_{int} = 38.28$ s, $\mu_{not} = 34.71$ s, $p < 0.05$), selection tasks 16% slower ($\mu_{int} = 18.88$ s, $\mu_{not} = 16.21$ s, $p < 0.03$), and registration tasks 3% slower ($\mu_{int} = 32.39$ s, $\mu_{not} = 31.34$ s, $p < 0.001$). When compared with task difficulty, results show that the disruptive effect of interruptions on completion time tends to increase with the difficulty of the primary task. Because more difficult tasks likely induce higher mental loads on working memory, users require more time to re-orient to the suspended task. This finding is consistent with results in (Kreifeldt & McCarthy, 1981).

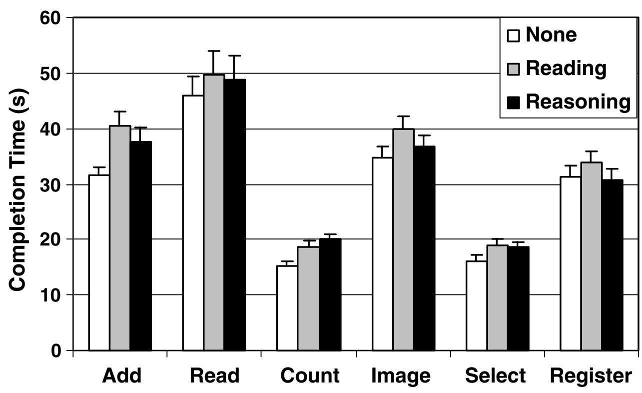


Fig. 4b. Time on the not interrupted vs. interrupted primary tasks within just the Experimental group.

In the Control group, peripheral task did not affect completion time, $F(2,48) = 2.37$, $p < 0.10$, and there were no interaction effects. Thus, when peripheral tasks interrupt the execution of primary tasks, performance on the primary tasks is degraded. However, when those same peripheral tasks are presented between primary tasks, there is no performance degradation.

*Peripheral tasks.* Figs. 5a–b show time on peripheral tasks clustered by the primary tasks between which (Control) or during which (Experimental) they were performed. Peripheral task had a main effect on completion time, $F(1,47) = 98.01$, $p < 0.001$. While reading tasks ($\mu = 25.96$ s) required more time to complete than reasoning tasks ($\mu = 17.10$ s), they were not expected to take exactly the same amount of time.

Timing marginally affected time on peripheral task, $F(1,47) = 3.05$, $p < 0.087$, with users in the Experimental group performing peripheral tasks ($\mu_{\text{Reading}} = 23.87$ s, $\mu_{\text{Reasoning}} = 15.50$ s) about 15% faster than users in the Control group ($\mu_{\text{Reading}} = 27.97$ s, $\mu_{\text{Reasoning}} = 18.64$ s; $p < 0.10$, $p < 0.08$, respectively). While users were instructed to perform all tasks as quickly as possible, performing the peripheral tasks in the context of having primary tasks interrupted ostensibly caused more intrinsic motivation and thus faster performance on the peripheral tasks (Card et al., 1983). This effect may not be present in

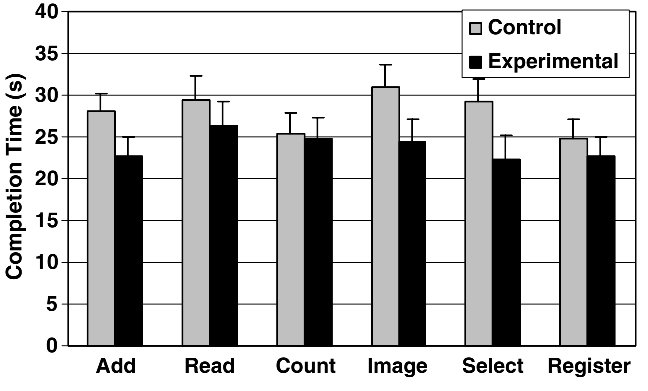


Fig. 5a. Time on the reading peripheral tasks between groups.

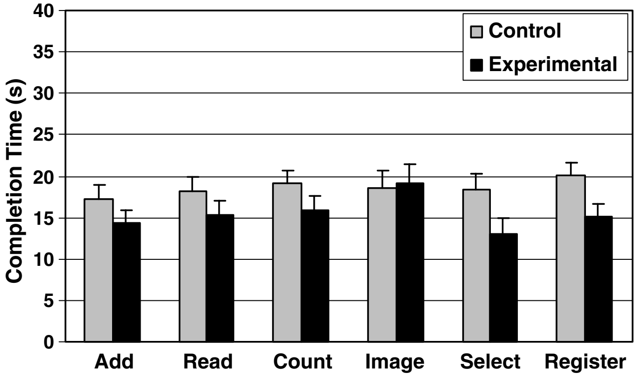


Fig. 5b. Time on the reasoning peripheral tasks between groups.

more realistic settings where user motivation does not necessarily need to be artificially instructed or otherwise induced. While cognitive resource conflicts have been shown to cause increased switching time among tasks (Altmann & Trafton, 2002; Rubinstein, Meyer, & Meyer, 2001), our results show that users were able to switch from the primary to peripheral tasks without significant cost, as there was no effect of primary task and there were no interactions in the data. Combined with the previous results for primary tasks, our results show that users are able to allocate resources for and initiate new (peripheral) tasks much more readily than they can resume previously suspended tasks.

### *4.3. Errors*

The number of errors committed for primary tasks (excluding registration) between groups is shown in Fig. 6a. Primary task had a main effect on errors committed, $F(4,192) = 4.35$, $p < 0.002$. Post hoc analysis showed that adding ($\mu = 0.26$) and image comprehension ($\mu = 0.21$) had more errors than reading ($\mu = 0.073$, $p < 0.002$, $p < 0.039$, respectively). While no other pairwise differences were detected, the number of errors committed generally increased with the difficulty of the primary tasks.

Peripheral task did not affect errors committed, $F(2,96) = 0.95$, $p < 0.39$, and analysis within the Experimental group (see Fig. 6b) further confirmed that there was no effect of peripheral task, $F(2,48) = 0.18$, $p < 0.83$. This was contrary to our expectation that interruptions would cause immediate, additional errors on the primary tasks. However, timing did have a strong main effect on errors committed, $F(1,48) = 67.30$, $p < 0.001$, with users in the Experimental group committing about *twice* the number of errors ($\mu = 0.22$, about two errors per 10 tasks) than users in the Control group ($\mu = 0.11$, about one error per 10 tasks). There were no interactions in the data.

While interruptions did not cause immediate error, the *expectancy* of interruption did cause more errors overall. A plausible explanation of this expectancy effect is that users were prospectively allocating cognitive resources to handle anticipated interruptions (Ellis, 1996; Trafton et al., 2003), diverting necessary resources away from primary tasks, resulting in additional error. These results, combined with the results for primary task completion time, show a multidimensional impact of interruption. Not only does the interruption

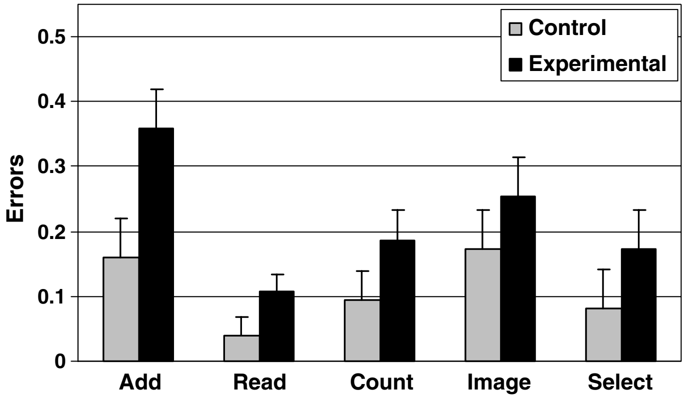


Fig. 6a. Errors committed on the primary tasks between groups.

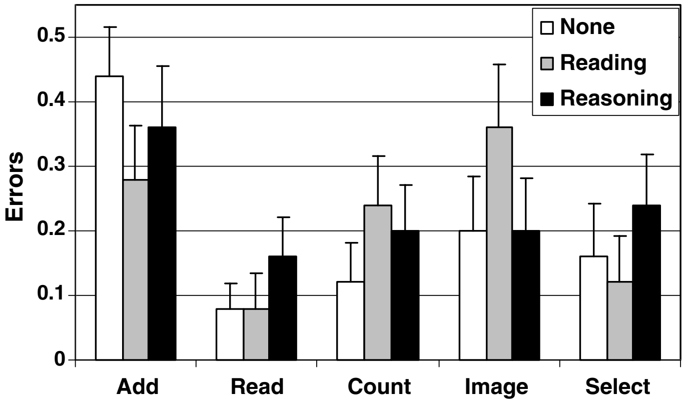


Fig. 6b. Errors committed on the primary tasks for the not interrupted vs. interrupted tasks in just the Experimental group.

proper cause a decrease in task completion time, but the expectancy of interruption causes increased error across tasks.

### 4.4. Annoyance

Fig. 7a shows user ratings for how annoying it was to receive peripheral tasks relative to the primary tasks. Primary Task had a main effect on ratings of annoyance, $F(5,240) = 35.93$, $p < 0.001$. Post hoc analysis showed that it was more annoying to receive a peripheral task relative to adding ($\mu = 13.75$) than to counting ($\mu = 10.61$, $p < 0.001$), image comprehension ($\mu = 7.40$, $p < 0.001$), selection ($\mu = 6.68$, $p < 0.001$), and registration ($\mu = 5.80$, $p < 0.001$). Also, it was more annoying to receive a peripheral task relative to both reading and counting than to image comprehension ($p < 0.001$, $p < 0.001$, respectively), Selection ($p < 0.001$, $p < 0.001$, respectively), and registration ($p < 0.001$, $p < 0.001$, respectively). No other differences were detected. These results show that the more difficult

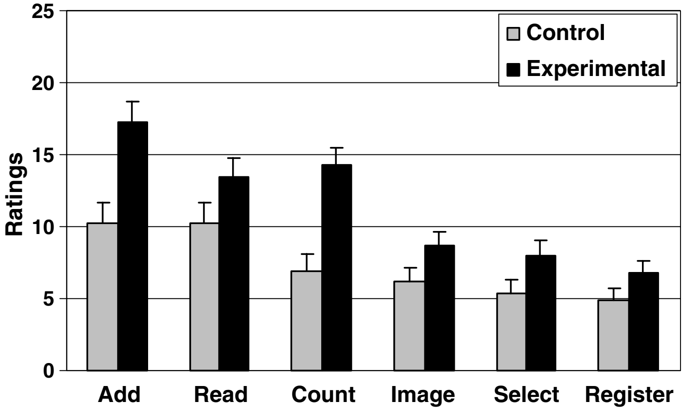


Fig. 7a. Ratings of annoyance for the primary tasks between groups.

the primary task, the more annoying it is to receive a peripheral task either during or between that primary task.

Although we expected ratings of annoyance to depend on the difficulty of the primary task in the Experimental group, we were somewhat surprised that a similar dependent pattern was detected in the Control group. This effect may be due to continued residual activation from the prior task (Allport, Styles, & Hsieh, 1994; Altmann & Trafton, 2002; Rubinstein et al., 2001), with more difficult tasks resulting in increased or longer residual interference. A user in the Control group succinctly summarized this effect, "the more difficult the [primary] task, the more I wanted a break before performing a peripheral task". If a short delay between completing a primary task and performing a peripheral task had been used in the Control group, the annoyance ratings may have been less dependent on the primary task.

Timing had a main effect on ratings of annoyance, $F(1,48) = 9.18$, $p < 0.01$. As illustrated in Fig. 7a, users experienced about 64% more annoyance when a peripheral task was presented during a primary task ($\mu_E = 11.40$) than when it was presented between primary tasks ($\mu_C = 7.29$). Post hoc tests showed that this was generally true across tasks. Annoyance increased by 68% for adding ($\mu_C = 10.26$, $\mu_E = 17.24$, $t(48) = 3.42$, $p < 0.001$), 31% for reading ($\mu_C = 10.22$, $\mu_E = 13.40$, $t(48) = 1.59$, $p < 0.055$), 106% for counting ($\mu_C = 6.92$, $\mu_E = 14.30$, $t(48) = 4.26$, $p < 0.001$), 41% for image comprehension ($\mu_C = 6.14$, $\mu_E = 8.66$, $t(48) = 1.81$, $p < 0.038$), 50% for selection ($\mu_C = 5.34$, $\mu_E = 8.02$, $t(48) = 1.89$, $p < 0.03$), and 40% for registration ($\mu_C = 4.84$, $\mu_E = 6.76$, $t(48) = 1.54$, $p < 0.065$). While there was an interaction between factors, $F(5,240) = 5.18$, $p < 0.001$, it was due to the counting and adding tasks in the Experimental group having larger increases in annoyance than the other tasks. These results show that the annoyance experienced when a peripheral task is presented depends not only on the difficulty of the primary task, but also on *when* that task is presented.

Fig. 7b shows a graph of user ratings for how annoying it was to perform peripheral tasks in each timing group. Peripheral task had a main effect on annoyance, $F(1,48) = 104.04$, $p < 0.01$. Reading tasks ($\mu = 11.23$) were rated as more annoying to perform than reasoning tasks ($\mu = 9.19$), which may have been due to users needing more time to complete reading tasks. Timing also had a main effect on these ratings of annoyance, $F(1,48) = 13.10$, $p < 0.001$. Users in the Experimental group ($\mu = 12.85$) rated the

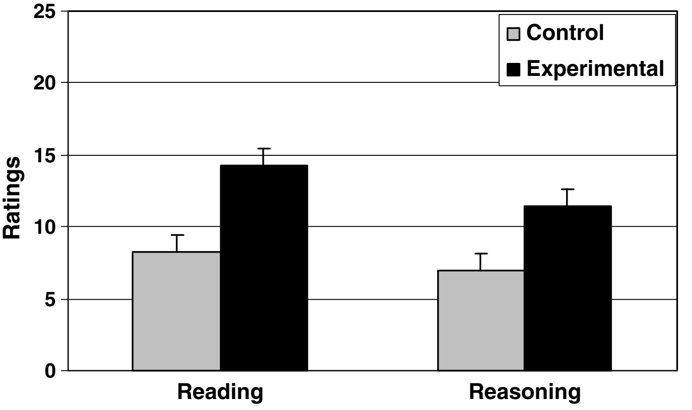


Fig. 7b. Ratings of annoyance for the peripheral tasks between groups.

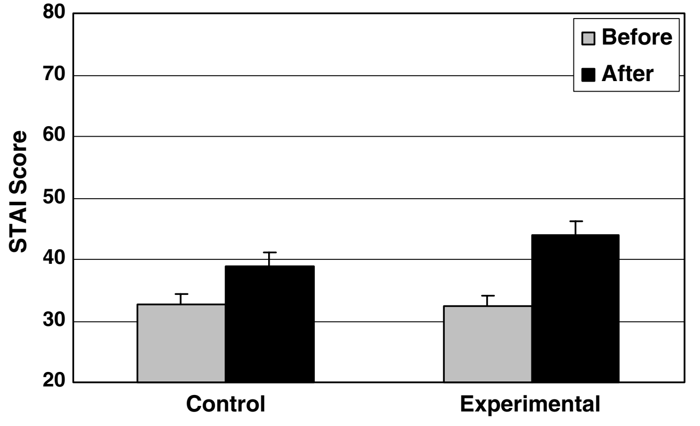


Fig. 8. User anxiety just before and after the experimental session between groups.

peripheral tasks as being about 60% more annoying to perform than users in the Control group ($\mu = 7.57$). No interactions were present in the data.

Consistent with completion time and error rate, these results show that the timing of peripheral tasks affects how much annoyance a user experiences when the primary task sequence is interrupted as well as when performing the peripheral tasks.

### 4.5. Anxiety

Fig. 8 shows a graph of the before vs. after differences in user anxiety between timing groups. Session (before vs. after) had a main effect on anxiety, $F(1,48) = 41.97$, $p < 0.001$, showing that users had more anxiety after participating in the experiment ($\mu = 41.50$) than before it began ($\mu = 32.50$). While timing did not have a main effect, there was a significant two-way interaction between session and timing, $F(1,48) = 3.90$, $p < 0.05$.

As can be seen in Fig. 8, the interaction was due to users in the Experimental group reporting about *twice* the increase in anxiety ($\mu_{\text{before}} = 32.32$, $\mu_{\text{after}} = 44.04$, $\mu_{\Delta} = 11.72$) than users in the Control group ($\mu_{\text{before}} = 32.68$, $\mu_{\text{after}} = 38.96$, $\mu_{\Delta} = 6.28$, $p < 0.028$). This represents about a 10% increase on the STAI scale. When peripheral tasks interrupt the execution of primary tasks, users experience considerably more anxiety than when they are presented between primary tasks.

## 5. Discussion and implications for attention-aware systems

### 5.1. Summary of results

The purpose of this experiment was to investigate how manipulating the timing of peripheral tasks relative to primary tasks would influence the resulting impact on completion time, error rate, and affective state. For completion time, interrupting the execution of primary tasks caused users to perform primary tasks from 3% to 27% slower than when not interrupted. While the degradation tended to increase with the difficulty of the primary task, it did not depend on the category of peripheral task. Degradation was likely due to users needing more time to re-orient to tasks that induced higher mental demands at the point of interruption. However, when peripheral tasks were presented *between* primary tasks, there was no degradation of task completion time.

Peripheral tasks were performed about 15% faster when they interrupted the execution of primary tasks. Increased performance may be attributed to a user's increased motivation to get back to and complete interrupted primary tasks as quickly as possible. The category of primary task did not affect peripheral task completion time in either timing group. These results show that users can initiate new tasks much more readily than they can resume previously suspended tasks.

For error rate, presenting peripheral tasks during the execution of primary tasks did not cause immediate, additional errors on the primary tasks. However, periodic interruption to the primary tasks did cause users to commit about twice the number of errors overall, indicating that interruptions cause increased or longer residual interference lasting beyond the immediate task.

In terms of affective state, results show that peripheral tasks interrupting the execution of primary tasks cause from 31% to 106% more annoyance than when presented between them. The more difficult the primary task, the more annoying it is to be interrupted. Interestingly, this pattern was evident regardless of when peripheral tasks were presented, suggesting that resources allocated to the primary task were not being immediately released upon completion of the task. The timing of peripheral tasks also influenced the amount of annoyance experienced. When peripheral tasks interrupted primary tasks, users experienced about 64% more annoyance relative to the primary tasks and about 60% more annoyance relative to the peripheral tasks. Also, users experienced about twice the increase in anxiety from the experimental session when peripheral tasks interrupted the execution of primary tasks than when they came between the primary tasks.

When extrapolated over the millions of computer users whose tasks are being increasingly interrupted by applications executing outside their focus of attention, these results show a remarkable impact of interruption on users and their tasks. Since only the time at which peripheral tasks were displayed was manipulated, our results demonstrate that effective timing of interruptions relative to primary tasks can substantially mitigate their disruptive effects. This is easily shown by interpreting our results from the opposite perspective. When peripheral tasks are presented at the boundary between primary tasks, users complete primary tasks faster, commit far fewer errors, and experience much less annoyance and anxiety than when those same peripheral tasks interrupt the execution of primary tasks.

### *5.2. Relations to existing work*

Our results for task completion time and error rate are consistent with many previous studies, including (Czerwinski et al., 2000a; Kreifeldt & McCarthy, 1981; Latorella, 1998; McFarlane, 1999; Trafton et al., 2003). However, by using a different sample of primary and peripheral tasks, our results provide further external validity to previous findings.

While Zijlstra et al. (1999) found that interruptions caused *increased* task performance, users in their experiment were instructed to work at their own pace. Increased arousal from the interruptions likely facilitated increased performance (Sanders & Baron, 1975), which more than compensated for any resumption lag after an interruption. Although users in our experiment were instructed to work as quickly as possible, we suspect that increased arousal was responsible for the increased performance on peripheral tasks in the Experimental group.

Our results on error rate support previous findings that interruptions can have a residual effect lasting well beyond the immediate primary task (Latorella, 1998). Within the Experimental group, users did not commit more errors on the interrupted tasks than the non-interrupted tasks, but they did commit more errors overall than users in the Control group. This indicates that users were experiencing residual interference from the interruptions.

For affective state, our results support and extend results in (Zijlstra et al., 1999). We found that interruptions impact affective state along the dimensions of anxiety as well as annoyance. However, our experimental design ensured that this effect cannot be attributed to differences in task workload, as only the timing of a peripheral task was manipulated. For information workers and end users, the impact of interruptions on affective state may be more important than on performance and error rate. Future work should seek to measure a broader spectrum of affective dimensions and consider the use of physiological measures such as galvanic skin response (Edelberg, 1972), which can provide a continuous and more objective measure of anxiety.

### 5.3. Timing and task structure

As this was one of the first experiments to focus on timing of peripheral tasks, we used a coarse structure of a primary task sequence to select moments for presenting peripheral tasks – at the boundaries between vs. during the execution of primary tasks. We expected that presenting peripheral tasks at a coarse boundary between primary tasks would be less disruptive than presenting them during the execution of primary tasks, since a user's mental workload should be lower at those moments (Miyata & Norman, 1986). Our results were consistent with this expectation.

An important question is whether finer-grained temporal manipulation of a peripheral task can mitigate the effects of interruption similar to the coarser timings used in this work. By leveraging task modeling techniques such as GOMS (Card et al., 1983) or event perception theory (Zacks et al., 2001), a primary task could be hierarchically decomposed and different boundaries in the task model could serve as candidate moments for presenting peripheral tasks (i.e. interruption). Comparing moments selected by alternative techniques will hopefully lead to more systematic methods for determining when interruptions would be least disruptive to a primary task.

Because boundaries in a task sequence are only *speculated* to represent moments of lower mental workload, linking physiological measures of workload such as pupil dilation (Beatty, 1982), heart rate variance (Rowe, Sibert, & Irwin, 1998), or blink rate (Kramer, 1991) to models of task execution may offer an improved technique for predicting how opportune different moments in a task are for interruption and offer a psycho-physiological explanation. Iqbal, Adamczyk, Zheng, and Bailey (2005), Iqbal, Zheng, and Bailey (2004), and Shell, Selker, and Vertegaal (2003), among others, are pursuing this direction of work.

### 5.4. Issues of practical relevance, user strategies, and limitations

An issue of practical relevance was our decision to display peripheral tasks using a modal dialog window. While the use of modal dialogs is representative of many practical situations such as receiving email notifications, system alerts, and instant messages, it is not representative of others, e.g., where salient visual or auditory cues are used to attract, but not force user attention. Our decision to use a modal dialog balanced the need to make the experiment representative of at least some practical situations with still being able to control interruption timing.

Another practical issue is whether interrupting a similar task sequence at coarse boundaries would produce a similar effect if those tasks themselves were part of a broader interactive activity. While the task sequences used in our experiment had well-defined boundaries with little or no cognitive carryover between tasks, if they were embedded within a broader interactive activity, such carryover would probably exist. However, the "cognitive rhythms" that Card et al. (1983) reported observing when users performed manuscript editing tasks, along with the concept of chunking behaviors discussed by Buxton (1986), strongly suggest that similar boundaries exist during the execution of broader interactive activities. If those boundaries could be identified, perhaps by linking physiological measures of workload to models of task execution (Iqbal et al., 2005), effects of interruption similar to the effects produced in this experiment may be achieved. This is an exciting opportunity for future research.

Peripheral tasks were delivered to a user at similar moments throughout the experiment. Thus, users could have employed strategies to create their own mental breaks during the tasks. For example, in the Control group, users could have paused just before selecting the 'finished' link, knowing that a peripheral task may ensue. In the Experimental group, users could have waited at the beginning of a task to determine if a peripheral task would appear. Several steps were taken to eliminate the use of these or similar strategies. Users were allowed to rest for several minutes between categories of primary tasks, the two tasks in each sequence that would be followed by interruptions were randomized, and most importantly, users were reminded to perform tasks as quickly as possible just prior to the start of each category. A review of the screen interaction videos confirmed that all users immediately began the tasks and, after completing them, quickly moved to select the 'finished' links. Thus, it does not appear that users employed any detectable strategy to circumvent the controls put forth in the experiment.

While we measured the effects of interruption using task performance, error rate, and affective state, our study did not include other measures such as the latency to attend to peripheral tasks (McCrickard, Catrambone, & Stasko, 2001) or the time to resume previously suspended tasks (Trafton et al., 2003). Interruption latency could not be meaningfully measured because users always attended to the peripheral tasks immediately, otherwise we could not control for timing. Resumption lag was not measured because the design of the primary tasks did not support cues to effectively measure when a user was back on task. We also did not manipulate the frequency, complexity, rehearsal time, or visual presentation style of the peripheral tasks (McCrickard et al., 2003). These were necessary tradeoffs to keep the complexity of the experiment within manageable limits. However, our work may motivate future research that seeks to better understand the relationship between these variables and interruption timing.

### *5.5. Implications for attention-aware systems*

Our results have important implications for the design of an attention-aware system called an attention manager that seeks to computationally balance a user's need for minimal disruption with an application's need to effectively deliver information. Motivated by our results, an attention manager could employ the use of a temporal strategy, useful either for manipulating when to present the information itself or for manipulating when to render attentional cues in peripheral displays, such as those in (MacIntyre et al., 2001; Maglio & Campbell, 2000; Van Dantzich, Robbins, Horvitz, & Czerwinski, 2002).

Leveraging a temporal strategy, an attention manager would reason about *when* peripheral information, or its attentional cue, should be presented, deferring presentation until a user reaches appropriate points during task execution. While our work indicates that waiting for coarse boundaries would mitigate disruption, future research may point to finer moments that produce similar mitigation. Most importantly, our results show that a *small* delay in the delivery of information could result in a *large* mitigation of disruption. In the Control group, peripheral tasks were delayed by no more than a few seconds compared to the Experimental group, yet the relative mitigation of disruption was substantial. In office-based or other environments where awareness of peripheral information is desired, but generally not safety critical, a small decrease in awareness may be well worth the large reduction in the disruption that would otherwise be caused by ill-timed interruptions.

Properties of peripheral information such as urgency and relevance to the primary task must also be considered when reasoning about interruption using a temporal strategy (Horvitz et al., 1999). For more urgent or relevant information, an attention manager could limit deferral to a shorter timeframe, e.g., the next boundary in a task sequence. Less urgent or relevant information could be deferred more optimistically, waiting until a coarser boundary is reached. To avoid indefinite waits, an attention manager must not delay presentation beyond a provided time limit. From a hierarchical decomposition of a task, the level of a boundary could also serve as part of a non-intrusive approximation for the cost of interruption in a broader reasoning framework, such as the one developed by Horvitz and Apacible (2003). However, tuning the relationship between boundaries in a task model and the cost of interruption will require more empirical research.

Developing an attention manager would require at least a mechanism to specify user tasks and observe when those tasks are performed, a system to learn a model of task execution based on the specifications and observed execution, and an iterative decision algorithm that uses the model to forecast which boundaries will be reached during task execution (Bailey, Adamczyk, Chang, & Chilson, 2005). While building such a system would require significant research effort, our results show that this effort is warranted. An effective system could enable users to perform tasks faster, commit fewer errors, and experience less annoyance and anxiety, dramatically improving the human–computer interaction experience.

## 6. Conclusion and future work

Interruption is becoming increasingly common in the human–computer interaction experience. It is imperative to further quantify and better understand the effects of interruption, test models of cognitive processes that can reliably predict their effects, and seek novel computational strategies for mitigating those effects. Our work has made contributions to each of these areas.

First, our results show that interruptions have a disruptive impact on completion time and error rate for primary tasks. Results provide further validity to previous findings and further show that interruptions have a residual effect that transcends the immediate task.

Our results are the first to show that interruptions have a negative impact on affective state such that this impact cannot be attributed to a difference in task workload. For information workers and end users, the effects of interruptions on affective state may be equally or more important than on performance.

Second, our results show that the time at which information is presented relative to a primary task influences its disruptive impact. As theorized in prior work and empirically supported in this work, interrupting users at boundary points during task execution is less disruptive, ostensibly due to increased availability of mental resources. Our future work will investigate whether finer temporal manipulation of peripheral tasks can produce similar mitigation.

Finally, our results imply that attention-aware systems could mitigate effects of interruption by deferring peripheral information until opportune moments such as coarse boundaries during task execution. The system could use the relevancy and urgency of the information to determine how long it could be deferred, and then wait until an appropriate boundary or expiration of the time limit to present it. Our future work seeks to develop such a system, which as shown by our results, could substantially improve the human–computer interaction experience.

## Acknowledgments

We thank the users who volunteered their time to participate in the experiment and the anonymous reviewers for providing helpful comments on an earlier draft of this article.

## References

Allport, A., Styles, E., & Hsieh, S. (1994). Shifting intentional set: exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 421–452). Cambridge, MA: MIT Press.

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

Bailey, B. P., Adamczyk, P. D., Chang, T. Y., & Chilson, N. A. (2005). A framework for specifying and monitoring user tasks. *Journal of Computers in Human Behavior* (special issue on attention aware systems).

Bailey, B. P., Konstan, J. A., Carlis, J. V. (2001). The effects of interruptions on task performance, annoyance, and anxiety in the user interface. In: *Proceedings of the IFIP TC.13 international conference on human–computer interaction* (pp. 593–601), Tokyo, Japan.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*(2), 276–292.

Buxton, W. (1986). Chunking and phrasing and the design of human–computer dialogues. In: *Proceedings of the IFIP world computer congress* (pp. 475–480), Dublin, Ireland.

Card, S., Moran, T., & Newell, A. (1983). *The psychology of human–computer interaction*. New Jersey: Lawrence Erlbaum Associates.

Cutrell, E., Czerwinski, M., & Horvitz, E. (2000). Effects of instant messaging interruptions on computing tasks. *Extended Abstracts of Human Factors in Computing Systems*, 99–100.

Cutrell, E., Czerwinski, M., & Horvitz, E. (2001). Notification, disruption and memory: effects of messaging interruptions on memory and performance. In: *Proceedings of the IFIP TC.13 international conference on human–computer interaction* (pp. 263–269), Tokyo, Japan.

Czerwinski, M., Cutrell, E., Horvitz, E. (2000a). Instant messaging and interruption: influence of task type on performance. In: *OZCHI 2000 conference proceedings* (pp. 356–361), Sydney, Australia.

Czerwinski, M., Cutrell, E., & Horvitz, E. (2000b). Instant messaging: effects of relevance and timing. *People and Computers XIV: Proceedings of HCI, 1*, 71–76.

Edelberg, R. (1972). Electrical activity of the skin: Its measurement and uses in psychophysiology. In N. S. Greenfield & R. A. Sternbach (Eds.), *Handbook of psychophysiology* (pp. 367–418). New York: Holt.

Ellis, J. (1996). Prospective memory or the realisation of delayed intentions: a conceptual framework for research. In M. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.), *Prospective memory: Theory and applications* (pp. 1–22). Mahwah, NJ: Lawrence Erlbaum.

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

Horvitz, E., & Apacible, J. (2003). Learning and reasoning about interruption. In: *Proceedings of the fifth ACM international conference on multimodal interfaces* (pp. 20–27).

Horvitz, E., Jacobs, A., & Hovel, D. (1999). Attention-sensitive alerting. In: *Conference proceedings on uncertainty in artificial intelligence* (pp. 305–313).

Iqbal, S. T., Adamczyk, P. D., Zheng, S., & Bailey, B. P. (2005). Towards an index of opportunity: understanding changes in mental workload during task execution. In: *Proceedings of the ACM conference on human factors in computing systems* (pp. 311–320).

Iqbal, S. T., Zheng, X. S., & Bailey, B. P. (2004). Task evoked pupillary response to mental workload in human–computer interaction. In: *Proceedings of the ACM conference on human factors in computing systems* (pp. 1477–1480).

Jackson, T. W., Dawson, R. J., & Wilson, D. (2001). The cost of email interruption. *Journal of Systems and Information Technology, 5*(1), 81–92.

Kramer, A. F. (1991). Physiological metrics of mental workload: a review of recent progress. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 279–328). London: Taylor and Francis.

Kreifeldt, J. G., & McCarthy, M. E. (1981). Interruption as a test of the user–computer interface. In *Proceedings of the 17th annual conference on manual control* (pp. 655–667). Jet Propulsion Laboratory, California Institute of Technology, JPL Publication 81-95.

Latorella, K. A. (1998). Effects of modality on interrupted flight deck performance: implications for data link. In: *Proceedings of the 42nd annual meeting of the human factors and Ergonomics Society* (pp. 87–91).

MacIntyre, B., Mynatt, E. D., Voida, S., Hansen, K. M., Tullio, J., & Corso, G. M. (2001). Support for multitasking and background awareness using interactive peripheral displays. In: *Proceedings of the ACM conference on user interface and software technology* (pp. 41–50).

Maes, P. (1994). Agents that reduce work and information overload. *Communications of the ACM, 37*(7), 30–40.

Maglio, P., & Campbell, C. S. (2000). Tradeoffs in displaying peripheral information. In *Proceedings of the ACM conference on human factors in computing systems* (pp. 241–248).

McCrickard, D. S., Catrambone, R., Chewar, C. M., & Stasko, J. T. (2003). Establishing tradeoffs that leverage attention for utility: empirically evaluating information display in notification systems. *International Journal of Human–computer Studies, 58*(5), 547–582.

McCrickard, D. S., Catrambone, R., & Stasko, J. T. (2001). Evaluating animation in the periphery as a mechanism for maintaining awareness. In: *Proceedings of the IFIP TC.13 international conference on human–computer interaction* (pp. 148–156), Tokyo, Japan.

McCrickard, S., Chewar, C. M., Somervell, J. P., & Ndiwalana, A. (2003). A model for notification systems evaluation–assessing user goals for multitasking activity. *ACM Transactions on Computer–Human Interaction, 10*(4), 312–338.

McFarlane, D. C. (1999). Coordinating the interruption of people in human–computer interaction. In: *Proceedings of the IFIP TC.13 international conference on human–computer interaction* (pp. 295–303).

McFarlane, D. C., & Latorella, K. A. (2002). The scope and importance of human interruption in HCI design. *Human–computer Interaction, 17*(1), 1–61.

Mital, A., McGlothlin, J. D., & Faard, H. F. (1992). Noise in multiple-workstation open-plan computer rooms: measurements and annoyance. *Journal of Human Ergonomics, 21*, 69–82.

Miyata, Y., & Norman, D. A. (1986). The control of multiple activities. In D. A. Norman & S. W. Draper (Eds.), *User centered system design: New perspectives on human–computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Monk, C. A., Boehm-Davis, D. A., & Trafton, J. G. (2002). The attentional costs of interrupting task performance at various stages. In: *Proceedings of the human factors and Ergonomics Society 46th annual meeting*.

Motowidlo, S. J., Packard, J. S., & Manning, M. R. (1986). Occupational stress: its causes and consequences for job performance. *Journal of Applied Psychology, 71*, 618–629.

Narayanan, L., Menon, S., & Spector, P. E. (1999). Stress in the workplace: a comparison of gender and occupations. *Journal of Organizational Behavior, 20*(1), 63–73.

Rowe, D. W., Sibert, J., & Irwin, D. (1998). Heart rate variability: indicator of user state as an aid to human–computer interaction. In: *Proceedings of the ACM conference on human factors in computing systems* (pp. 480–487).

Rubinstein, J. S., Meyer, D. E., & Meyer, D. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance, 27*(4), 763–797.

Sanders, G. S., & Baron, R. S. (1975). The motivating effects of distraction on task performance. *Journal of Personality and Social Psychology, 32*, 956–963.

Shell, J. S., Selker, T., & Vertegaal, R. (2003). Interacting with groups of computers. *Communications of the ACM, 46*(3), 40–46.

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.

Spielberger, C. D. (1983). *Manual for the state-trait anxiety inventory*. Redwood City, CA: Mind Garden.

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.

Van Dantzich, M., Robbins, D., Horvitz, E., & Czerwinski, M. (2002). Scope: providing awareness of multiple notifications at a glance. *Proceedings of Advanced Visual Interfaces*.

Wickens, C. D. (1984). *Engineering psychology and human performance*. Boston: Scott Foresman & Co..

Zacks, J., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General, 130*(1), 29–58.

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.

**Brian P. Bailey** is an Assistant Professor in the Department of Computer Science at the University of Illinois. His research investigates developing interactive design tools that better support human creativity, user interfaces for pervasive computing, computational systems that manage human attention, and other areas of human–computer interaction. Dr. Bailey received his Ph.D. from the University of Minnesota, Minneapolis, in 2002. His is a member of the ACM and the current editor of the ACM SIGCHI Bulletin.

**Joseph A. Konstan** is an Associate Professor in the Department of Computer Science and Engineering at the University of Minnesota. His research spans the areas of recommender systems, interactive multimedia, online medical applications, visualization, on-line community, and general human–computer interactions. Dr. Konstan received his Ph.D. from the University of California, Berkeley, in 1993. He is President of ACM SIGCHI, an ACM Distinguished Lecturer, and an IEEE Distinguished Visitor.