Instant Messaging and Interruption: Influence of Task Type on Performance

Mary Czerwinski, Edward Cutrell and Eric Horvitz
Microsoft Research
{marycz, cutrell, horvitz}@microsoft.com

Abstract

We describe research on the effects of instant messaging (IM) on ongoing computing tasks. We present a study that builds on earlier work exploring the influence of sending notifications at different times and the kinds of tasks that are particularly susceptible to interruption. This work investigates alternative hypotheses about the nature of disruption for a list evaluation task, an activity we had identified as being particularly costly to interrupt. Our findings replicate earlier work, showing the generally harmful effects of IM, and further show that notifications are more disruptive for fast, stimulus-driven search tasks than for slower, more effortful semantic-based search tasks.

Keywords: Notifications, interruption, user studies, attention, reminders

1. Introduction

Instant messaging systems, including America Online's Instant Messenger, Microsoft Network's Messenger, and Yahoo!'s Messenger service claim over 70 million users and research shows that the number of users of these services is growing rapidly (MediaMetrix, December, 1999). The benefits of instant messaging are numerous, including the ability to know when personal contacts are available, having access to nearly instantaneous communication, and the ability to carry on several informal conversations at once. However, the effects of incoming instant messages and other alerts, such as email, news summaries, and other updates, on ongoing computing tasks have been relatively unexplored.

The investigation of the costs of instant messages (IM) in the context of desktop computing tasks falls in the general arena of psychological research on alerting and disruption. To date, much of the research on interruption leverages abstract or theoretical task constructions. We have undertaken a series of studies to investigate the nature of interruptions commonly associated with computing such as instant messaging on computer users. In this paper, we shall first review related work, including several recent studies performed in our lab investigating the effects of alerting on the efficiency of desktop tasks. Then we will describe a study following up on these findings. Finally, we summarize our results and discuss directions for future research.

2. Related Work

Although little research has been performed on the effects of IM on real-world computing tasks, there have been numerous studies exploring interruptions more generally. McFarlane (1999) examined four methods for deciding when to interrupt someone during multitasked computing. He explored several interruption policies, including *immediate* (requiring an immediate user response), *negotiated* (user chooses when to attend), *mediated* (an intelligent agent might determine when best to interrupt) and *scheduled* (interruptions come at prearranged time intervals) notifications. He found that none of these methods was the single best way to interrupt users in tasks across all performance measures. McFarlane concluded that giving people the control to negotiate for the onset of interruptions resulted in good performance. However, he cautions that users may postpone attending to interrupting messages in these cases. Also, if forced to acknowledge an interruption immediately, users in his study got the interrupting task done promptly but were less efficient overall.

Gillie and Broadbent (1989) presented a series of experiments aimed at elucidating features of interruptions that make them more or less disruptive to an ongoing computer task. They manipulated interruption length, similarity to the ongoing task, and the complexity of the interruption. They showed that being able to rehearse one's position in the main task does not protect one from the disruptive effects of an interruption. In addition, they discovered that interruptions with similar content could be quite disruptive even if they are extremely short, replicating findings in earlier work by Kreifeldt and McCarthy (1981).

Other work has shown that if an interrupter imposes a high memory load or processing demands on the user, it is harmful to the primary task. Hess and Detweiler

(1994) showed that interruptions that were similar to an ongoing computer task are quite disruptive over the first two of three sessions, but are significantly less disruptive by the third session. In addition, they found that, if participants are allowed to train on the primary task without interruptions for two sessions, then presenting a third session with interruptions is significantly harmful to performance, despite the task being highly trained. It would appear from these last results that experience handling the interrupting tasks reduces their harmful effects over time.

Several studies have shown that the nature of the display of the notification influences performance on the primary computing task. Maglio and Campbell (2000) demonstrated that continuously scrolling displays were more distracting than discrete displays (those that start and stop) to ongoing word editing tasks. They found that all notification styles reduced word-editing performance in comparison to a no-notification condition. It should be noted that the methodology used in these studies was more akin to a dual task paradigm, as participants were instructed to monitor the scrolling notification displays. Memory tests did not reveal any differences across the different types of scrolling displays tested.

Ware, Bonner, Knight and Cater (1992) reported an experiment designed to test the use of simple linear motion as an attention-getting device for computer displays. The experiment also utilized a dual task paradigm. A primary task required the transcription of a document typed into a computer screen and a secondary task involved detecting and responding to a moving icon signal. The icon was a rectangular bar that grew and shrank vertically in an oscillatory fashion. Both the amplitude and velocity of the icon's motion were varied systematically and response time was recorded. The results from the secondary task showed that there was an inverse relationship between the velocity of the moving icon and time to respond to the icon movement, but no effect was found for amplitude. Observed reponse speeds appeared to indicate that simple motion was an effective attention-getting device for events in the periphery of the visual field. These results could be useful to designers that wish to have users quickly switch attention to high priority instant messages.

Mollenhauer, Lee, Cho, et al. (1994) reported a driving simulator study in which participants were presented road sign information from a visual dash-mounted LCD display or from digitized auditory voice. Participants either received all road sign information or only "filtered" high-priority sign information. The effects of display type and filtering on information recall, driver performance, and driver preferences were measured. The results indicated that auditory information presentation was associated with increased road sign recall, but decreased the subjects' driving performance. Participants also rated auditory information as more distracting than visual information. Participants were able to recall more road sign information and drive at a higher level of performance during the filtered conditions. These results suggest that holding off notifications unless they are of high priority would improve overall performance, but it is unclear how well the results might generalize to the computing domain.

In previous work (Czerwinski, Cutrell & Horvitz, 2000), we explored several conditions using as measures of disruption the times required for the user to move from tasks to the notification, to read the notification, and to return to the primary task following review of the notification. We found that the degree of disruption to ongoing productivity tasks depended on the specific point in a task that a notification was presented. We found that it was less costly if notifications came early in a task, before the user had become deeply engaged in the task goal. More specifically, we found that the costs of the disruption depended on the nature of the ongoing task or subtask. Finally, we found that IMs that were relevant to ongoing tasks were less disruptive than those that were irrelevant. This influence of relevance was found to hold for both notification viewing and task resumption times, suggesting that notifications that were unrelated to ongoing tasks took longer to process.

3. Deeper Study of Disruption

In an effort to better frame our new results, we shall review our previous efforts to elucidate the relationship between the nature of the ongoing task and the disruptive effects of IMs. In our earlier work, we formulated a web-based search task and divided the overall task into several phases. We explored the differential influence of IMs on each phase. We were influenced in our definition of the phases by conjectures about interruptions made over a decade ago by Miyata and Norman (1986). Following their conjectures, we described formulating a web search query as "planning," typing the search query and using buttons or menus as "execution", and reviewing search results for the desired target web page as "evaluation." Miyata and Norman had speculated that interfering with any of these stages would be problematic, and that better interruption points would be at breaks between these hypothesized stages.

We found that receiving an IM was particularly disruptive during the execution and evaluation stages. That is, the costs were highest when a message was received when users were typing or interacting with toolbars and while users were scanning a list of web search results for a target. The first of these findings is consistent with the notion of *chunking behaviors* (Sellen, Kurtenbach, & Buxton, 1990), suggesting that some behaviors are grouped tightly into chunks and thus are difficult to control or guide once their execution begins. Chunking would provide an explanation for a user delaying a transition to a new task until a currently executing chunked subtask is completed. Chunking may characterize automated behaviors such as typing a word or phrase.

In this paper, we report on our effort to better understand why the evaluation phase was particularly harmed by the notifications. We formulated two hypotheses. The first hypothesis is that the result may reflect the time required for users to visually re-orient themselves to where they recently left off in the search results list, and the concomitant re-scanning of the web search results after the interruption. An alternative explanation is that the additional delays arose from latencies in the access of the memory of the goal—in this case why/if a particular result was a candidate target. We set out to explore the finer structure of disruptions during evaluation of a list of items.

4. Experiment: Messages and List Evaluation

Why are instant messages rendered during the evaluation stage of web searching more costly than other stages? We sought to identify whether or not harmful effects were observed during visual scan, target identification or remembering the goal, a task we refer to as conceptual reacquisition. As part of this work, we explored the value of leaving a displayed "marker" as a reminder to users where they left off in their primary task when returning from a notification.

4.1. Subjects

Twelve participants (6 female) aged between 25 and 54 years of age (average age was 41 years old) came to the lab for a single session. All subjects were screened to be at least at an intermediate level of proficiency at using Microsoft Windows and Office products. Six of the subjects had used MSN's Messenger v. 2.0 prior to the study, and six had never used Messenger before. All subjects were run singly for one session.

4.2. Design and materials

Sixty-four Excel target and distracter book title sets were derived from 6400 book titles obtained from the Microsoft Library. Book items were chosen to be targets if they were found to be distinctive within a group of 80 distracter titles (i.e., book titles that did not have similarly titled, competing alternatives during a search trial for that book). The Excel spreadsheets were designed so that every trial comprised a worksheet with that trial's number on the worksheet's tab (*i.e.*, each workbook for a given subject had 64 worksheets, for 64 trials). Each worksheet then contained the search target at the top of the list, and a list of 80 book titles below it (~ 3 pages worth of search results at a screen resolution of 640x480). The spreadsheets were fixed so that the description of the search target did not scroll off the top of the screen when the subject moved more than a screenful through the list. Figure 1 shows an example of an Excel spreadsheet with stimuli from the experiment.

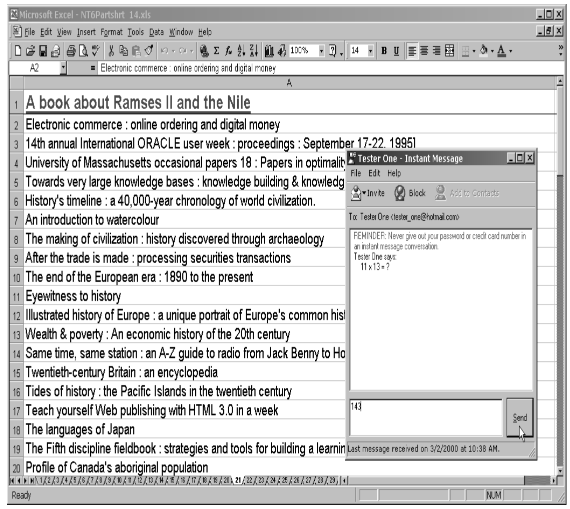

Figure 1. Example of "gist" search with no marking cursor and an instant message.

Users navigated the lists using either the Cursor Up/Down (arrow) keys or using the Page Up/Down keys. When they used the arrow keys to navigate, a marker (the cursor) outlined the currently selected box, and items scrolled off the top of the screen one by one as they moved down the list. In contrast, when they used the Page Up/Down keys, the entire page was replaced with the next page and no marker outline was visible.

The difficulty of remembering the goal when returning from a notification was manipulated by altering the type of search target. For half of the trials, subjects were given the verbatim title of the book. This made the task a relatively straightforward visual scan for the first few letters of the title with little cognitive demand. For the other half of the trials, subjects were given a *gist* (e.g., "A book about Ramses II and the Nile.") We assumed that these tasks were cognitively more demanding, requiring more resources for recall, and for the real-time guiding of a search for semantic content. The average length of titles and gists were roughly equivalent.

The experimental design was a 2 (title v. gist search trial) x 2 (marker—arrow keys with cursor outline v. no marker—Page Up/Down keys without cursor outline) x 2 (notification trial or no notification trial) x 8 (replications per condition) for a total of 64 trials per session. Dependent variables included total task time, time to switch to a notification and time spent on a notification when one occurred.

4.3. Procedure

Participants were greeted and given a tour of the laboratory before starting. As part of this tour, subjects were introduced to a second experimenter, and told that this experimenter would be sending them notifications throughout the experiment. This was done so that participants would feel familiar with the person sending notifications once the study began. Next, participants were asked to read directions describing the search procedures, including how to navigate both using the arrow keys and Page Up/Down keys. Once they had completed reading the instructions, the experimenter walked them through two practice trials in order to familiarize the participants with the experimental procedure. One practice trial had participants navigate the Excel spreadsheet lists with the arrow keys, and the other trial had them use Page Up/Down. In addition, one trial was a "title" search trial; one was a "gist" search trial. Participants were sent notifications during both practice trials, so that they could become familiar with how to work with the notification user interface.

During the experiment proper, participants were asked to do two kinds of searches: In half of the trials, we gave them a book title as their target and they simply had to scan the list for the title. In the other half, we gave them a short gist of what the book was about and they had to scan the list for the title of the book associated with our description. In addition, on half of all trials the experimenter interrupted the participant within 30 seconds of beginning the search and before the title was found. The interruption consisted of the experimenter sending a message from MSN's Messenger with a simple multiplication or division problem as the message content. The participant was asked to respond to the message by solving the math problem and then to return to the search task and continue until the correct book title was found. When participants found the correct title match, they alerted the experimenter and moved on to the next trial by clicking on the worksheet tab just to the right of the current worksheet. All tabs were numerically labeled to show the trial number. After 32 search trials using either the arrow keys (marker condition) or the Page Up and Down keys, the participant took a short break and used the alternate navigation technique for the second half of the session. Order of navigation technique was determined at random for the first participant and then alternated between subjects thereafter. All other variables were run within subjects and were counterbalanced and randomized in terms of presentation for a given session. Subjects completed satisfaction questionnaires at the end of the one-hour experimental session, were debriefed, and escorted to the lobby. All participants received a software or book gratuity for their participation.

5. Results

Although some book titles were more difficult to find in the list than others, users were able to find all of the book titles. If the search time became excessive (defined to be when the subject went past the target in the list for a third time) for a given trial, the experimenter would give a hint as to which third of the list the book title was in. On average, this occurred less than once per session for a given user. Therefore, accuracy data need not be dealt with, and only time data will be reported.

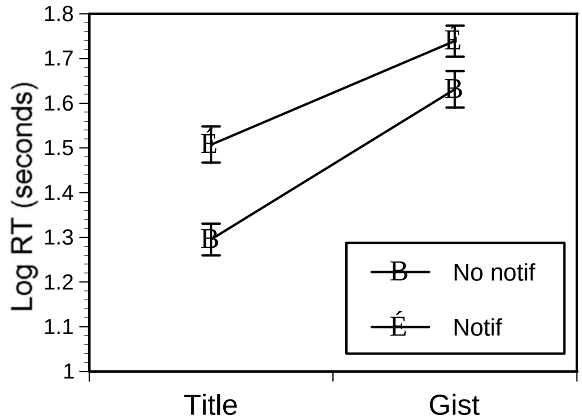

Figure 2. The effects of notification and search type (title or gist) on overall average task times.

Log response times were used in the analyses to normalize the common skewing and variability associated with response time data. The findings for total task time are shown in Figure 2. Receiving notifications reliably slowed down performance on the primary task of searching for a book title. In addition, searching for the title of the book was reliably faster than using the gist of what the title was. There was no reliable difference between task times when navigating with the arrow keys versus the Page Up/Down keys during search trials, nor was there an interaction between this variable and any other. Our analysis of the task time data detailed below shows that the slowdown in task performance after notifications cannot be blamed entirely on the necessity of switching from the keyboard to the mouse and back again, and that significant costs remain when the device switching times are subtracted out from the total task times. This provides at least some evidence for the argument that a large proportion of the cost of notifications on primary task performance comes from the influence of notifications on memory. An analysis investigating any differential item, repetition or position effects allowed us to collapse across these variables, as they did not contribute to a significant amount of the variance. Therefore, a 2 (notification trial or not) x 2 (marked navigation v. page up/down) x 2 (title v. gist search) within subjects Analysis of Variance (ANOVA) was carried out on the overall trial completion time data. A significant main effect was observed for whether or not a notification was received during a search trial, $F(1,11)= 236.2$, $p<.001$. As seen in Figure 2, trials with notifications were reliably slower than those without.

A significant main effect was also observed for whether or not participants were completing a title or a gist search trial, $F(1,11)= 99.3$, $p<.001$; title search trials were significantly faster than gist search trials. In addition, there was a reliable interaction between notification and title vs. gist, $F(1,11)=6.4$, $p<.05$ (see Figure 2). No other main effects or interactions were significant.

As alluded to earlier, we analyzed the task time data after subtracting out the notification times. We did this because it might be argued that the costs of the interruption come solely from switching devices (from the keyboard to the mouse to interact with Messenger). We observed a similar pattern of results as we saw with the total task times: reliable main effects of notification [$F(1,11)=23.26$, $p<.001$] and title vs. gist [$F(1,11)=97.94$, $p<.001$] were obtained. A reliable interaction between marking method (arrow v. page up/down) and title v. gist [$F(1,11)=4.9$, $p=.047$], was also obtained. Participants were reliably faster in the title condition when the marker was present (see Figure 3). No other main effects or interactions were significant. Therefore, the primary costs of the interruptions remain even after subtracting out device switching time.

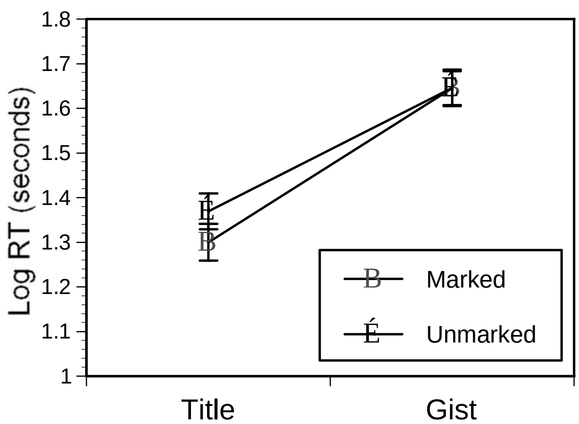

Figure 3. The effects of notification and search type (title or gist) on overall average task times minus notification time.

6. Discussion

This study bolsters the findings from our previous work, again demonstrating the harmful effects that notifications have during the task of searching through a list. We suspect this will generalize to other similar types of evaluation tasks. Additionally, we demonstrated that notifications reliably harm faster, stimulus-driven search tasks more than effortful, cognitively taxing search tasks, such as the gist search condition. This is true even though both kinds of search are used during the evaluation stage of computing tasks, and when device-switching time is accounted for. A possible explanation for this is that for tasks centering on high-speed visual scanning, users need to disengage and then re-engage their scanning mechanism(s) after an interruption, both potentially effortful endeavors (Shiffrin, 1988). Other explanations exist, and further research is needed to isolate the determining factors behind this result. A priori, we had thought we might observe a beneficial effect of the marker in this regard. However, we found that having a marked position in a search list improved our subjects' performance for the title searches only when notification times were subtracted from overall task times. A more salient marker may have helped more overall, but this is uncertain. There are several potential explanations for these results. In one, users may not have actively employed the cursor for position management and memory, especially in the gist condition. Also, it may have taken users longer to engage the rapid visual scan mechanism when one returned to a feature-based title search than it did in the slower gist condition. In any case, these data do not clearly support the hypothesis that the effect of notifications on this task was due to a difficulty in visual reorienting to the task.

There was a basic navigational confound with the marking procedure used in the study: To mark current search position, the user had to navigate via the arrow keys. In the *no marking* condition, participants used the Page Up and Page Down keys. It may be that the differences between the efficiencies of these two navigational techniques are masking any performance advantage that might be provided by a cue on spatial location.

Although our marking procedure was not as effective a reminder as we had hoped in this experiment, we remain optimistic that reminders about an interrupted task might prove to be valuable in reducing the disruptiveness of notifications. In addition to navigational cues, such reminders might include the use of graphical and linguistic summaries of the interrupted task. For example, a system might remind the user with words of what they were doing prior to a notification, also providing links back to the primary task or subtask. We plan to continue to explore designs and user interface mechanisms for helping users to recover their context regarding the primary task after an interruption, and to pursue a better understanding of how policies and designs for notifications might mitigate the harmful effects of interruptions.

This study was a systematic step in a larger, ongoing research effort to examine the psychological effects of notifications during different computing tasks. Our research in this area has been carried out with an eye toward principles of human-computer interaction tools, metaphors, and designs that could reduce the disruptiveness of notifications. We also hope that the results of this work and related psychological studies will provide parameters that support ongoing work on systems that employ automated reasoning to control notifications (Horvitz, Breese, Heckerman, et al., 1998; Horvitz, Jacobs & Hovel, 1999).

Over a series of experiments, we have shown the disruptive effects of notifications on a variety of ongoing computing tasks. We have confirmed the predictions of Miyata and Norman (1986) that some task phases are less amenable to interruption than others. In particular, we found that sending an instant message while a participant is typing, using buttons or menus, or evaluating search results is harmful to overall task performance (Czerwinski, Cutrell and Horvitz, 2000). We believe that the results that we have observed over a series of experiments on IMs can provide guidance for designers of instant messaging systems. As both the popularity of instant messaging and the number of providers of messages grows, it will become increasingly valuable to consider designs for messaging systems that minimize the cost while providing the most value. In particular, guidelines can be developed based on a set of initial results gleaned from psychological studies on (1) the influence of the relevance of message content to the current task and (2) the sensitivity of performance to task phase that is interrupted. Prior studies have suggested that withholding messages until key task and subtasks are detected could mitigate the disruptive effects of instant messages (Czerwinski, Cutrell and Horvitz, 2000; Mollenhauer, Lee, Cho, Hulse, & Dingus, 1994). In particular, waiting until the user is task switching (e.g., just beginning or finishing a task), or until a query or other short keyboarding event has completed (e.g., typing in a field) will help the user be more efficient. From the experiment reported in this paper, it appears that notifications sent during the fragile evaluation stage of a task are harmful if the user is quickly scrolling through search results. In this case, better user interface tools might minimize the disruptive effects of instant messages on primary task performance.

7. References

[1] Anderson, R.C. and Pitchert, J.W. (1978). Recall of previously unrecallable information following a shift in perspective, *Journal of Verbal Learning and Verbal Behavior*, 17, 1-12.

[2] Bryan, W. and Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits, *Psych Rev*, 6, 345-75.

[3] Czerwinski, M., Cutrell, E. & Horvitz, E. (2000). Instant messaging: Effects of relevance and time. In S. Turner & P. Turner (Eds.), *People and Computers XIV: Proceedings of HCI 2000, Vol. 2*, British Computer Society, pp. 71-76.

[4] Gillie, T. and Broadbent, D. (1989). What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychol Res*, 50, 243-50.

[5] Horvitz, E., Jacobs, A. & Hovel, D. (1999). Attention-sensitive alerting. *15th Conf. on Uncertainty and AI (UAI '99)*, Stockholm, Sweden. Morgan Kaufmann Publishers: San Francisco, pp. 305-13.

[6] Horvitz, E., Breese, J., Heckerman, D., Hovel, D., and Rommelse, K. (1998). The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users. *14th Conf. on Uncertainty in Artificial Intelligence (UAI '98)*, Madison, WI. Morgan Kaufmann Publishers: San Francisco, pp. 256-65.

[7] Maglio, P.P. & Campbell, C.S. (2000). Tradeoffs in displaying peripheral information. In *Proceedings of Association for Computing Machinery's Human Factors in Computing Systems, CHI 2000*, pp. 241-248.

[8] McFarlane, D. (1997). Interruption of people in human-computer interaction: A general unifying definition of human interruption and taxonomy. Naval Research Laboratory Report # NRL/FR/5510--97-9870.

[9] McFarlane, D. (1999). Coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction - INTERACT'99*, IOS Press, Inc., The Netherlands, pp. 295-303.

[10] Miyata, Y. and Norman, D.A. (1986). Psychological issues in support of multiple activities. In D.A. Norman and S.W. Draper (Eds.), *Participant Centered Design: New Perspectives on Human Computer Interaction*. Lawrence Erlbaum, pp. 265-84.

[11] Mollenhauer, M.A., Lee, J., Cho, K., Hulse, M.C., & Dingus, T.A. (1994). Effects of sensory modality and information priority on in-vehicle signing and information systems. *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*, v.2, pp. 1072-1076.

[12] Shiffrin, R.M. (1988). Attention. In Atkinson, R.C., Herrnstein, R.J., Lindzey, G., & Luce, R.D. (Eds.), *Stevens' Handbook of Experimental Psychology, 2nd Ed.* New York: Wiley, pp. 739-811

[13] Ware, C., Bonner, J., Knight, W. & Cater, R. (1992). Moving Icons as a Human Interrupt, *International Journal of Human-Computer Interaction*, v.4, pp. 341-348