Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance

Edward Cutrell Mary Czerwinski Eric Horvitz

Microsoft Research, One Microsoft Way, Redmond, WA 98052 USA

{cutrell; marycz; horvitz}@microsoft.com

Abstract: We describe a study on the influence of instant messaging (IM) on ongoing computing tasks. The study both replicates and extends earlier work on the cost of sending notifications at different times and the sensitivity of different tasks to interruption. We investigate alternative hypotheses about the nature of disruption for a list evaluation task, an activity identified as being particularly costly to interrupt. Our findings once again show the generally disruptive effects of IM, especially during fast, stimulus-driven search tasks. In addition, we show that interruptions coming early during a search task are more likely to result in the user forgetting the primary task goal than interruptions that arrive later on. These findings have implications for the design of user interfaces and notification policies that minimize the disruptiveness of notifications.

Keywords: Notifications, user study, interruptions, information overload, divided attention

1 Introduction

With the advent of wireless communications, the ability to subscribe to services that deliver instant messages and notifications of all kinds has grown dramatically. AOL, Microsoft and Yahoo!, among a host of other companies, all provide the capability to have email, stock updates, buddy information and messages of various kinds automatically sent to your cellular phone or other, wireless device. The benefits of having instant access to your email, family, friends or other forms of information are many and previous research has shown that messaging of this nature can provide significant value to users (O'Connail & Frohlich, 1995). However, notifications also can be disruptive, both frustrating users and decreasing the efficiency with which they perform ongoing tasks. Methods for intelligent filtering and decision making about notifications show promise for minimizing the disruptiveness of notifications (Horvitz, Jacobs, & Hovel, 1999). Such systems can be enhanced by considering results gleaned from psychological studies of the influence of notifications on people performing real-world tasks.

We have undertaken a series of studies to investigate the nature of interruptions associated with instant messaging on computer users. We shall review related work, including a recent study performed in our lab. Then, we replicate and extend our earlier research with new findings. Finally, we summarize our results and discuss directions for future research.

2 Related Work

Although the research landscape is beginning to change as the influence of notifications on computer users grows in importance, to date, much of the psychological work on interruption has leveraged abstract or theoretical task constructions. A body of work on the effects of IM on real-world computing tasks is only just beginning to emerge (e.g., Altmann & Gray, 2000; Bailey, Konstan & Carlis, 2000; Entin, 2000; Gillie & Broadbent, 1989; Gopher, Greenshpan & Armony, 1996; Jambon, 1996; Kreifeldt & McCarthy, 1981; McFarlane, 1999; Renaud, 2000; Rhodes, Benoit & Payne, 2000). In this section, we focus a literature review on interruption studies related to memory, timing, task type and task switching, as these are the crucial areas of focus in the current work.

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. McFarlane found that none of these methods was the single best way to interrupt users in tasks across all performance measures. He found that if users are forced to acknowledge an interruption immediately, they completed the interrupting task promptly but were less efficient overall. He discovered that allowing users to negotiate when they attend to notifications enhances the performance of users on a background task. However, he points out that providing users with an ability to negotiate the receipt of notifications may lead to potentially indefinite postponement of notifications.

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 the position in the main task does not protect users from the disruptive effects of an interruption. They also 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). Since then, others have replicated and extended this result (Bailey, Konstan & Carlis, 2000; Czerwinski, Chrisman & Rudisill, 1991; Czerwinski, Chrisman & Schumacher, 1991; Hess & Detweiler, 1994; but see Linde and Goguen, 1987). In related work, Rhodes, Benoit & Payne (2000) point out that when formatting visual displays for dynamically updating environments, information should be presented in such a way as to reduce similarity interference (e.g., make the information highly distinctive across items in the display).

Other work has shown that while an interrupter that imposes a high memory load or processing demands on the user is harmful to the primary task, the effects of these interruptions are strongly influenced by training or expertise. Hess and Detweiler (1994) showed that interruptions that were similar to an ongoing computer task were quite disruptive over the first two of three sessions, but were significantly less disruptive by the third session. In addition, they found that, if participants were allowed to train on the primary task without interruptions for two sessions, then presenting a third session with interruptions was significantly harmful to performance, despite the task being highly trained. It would appear from the latter results that a user's experience with handling interrupting tasks reduces their harmful effects over time. Gopher, Greenshpan & Armony (2000) also point out that switching tasks—or even being notified simply to continue the same task—is costly. The investigators found that the costs were not uniform. The costs were discovered to be related to the nature of the current and pending activity, as well as the user's proficiency at both tasks; an anticipated task switch could be prepared for in advance. Again, such an ability to prepare for a switch depends on having some proficiency with the tasks.

Altmann & Gray (2000) used functional decay theory in a series of studies to demonstrate that in dynamic task environments, if users don't have enough time to pay attention to an update and enough time to let a previous task item fade from memory, situation awareness could degrade "catastrophically." Their work also showed the importance of developing some expertise with the current task instructions.

As workplace interruptions are frequently beneficial to users (O'Conaill & Frohlich, 1995), developing methods for recovering from the loss of primary task focus and context promises to be valuable. Renaud (Renaud, 2000) describes an attempt to aid recovery from interruptions by providing a visualization of application activity. She discussed a user study that showed that subjects using the system could recover more efficiently from errors, but she could not demonstrate that users were able to reinstate context after an error was provided.

In other related work, Linde and Goguen (1987) reported a linguistic study of interruptions to crews who performed difficult flight landing simulations. The authors found that the best crews used "explicit holds," or linguistic markers, that let the crew know that a checklist was being momentarily suspended at step *n*, in order to attend to an interrupter, after which the checklist procedures would be reinstated at the step on hold.

3 A Previous Study

An earlier user study, reported in Czerwinski, Cutrell & Horvitz (2000b) demonstrated the harmful effects that notifications have during the task of searching through a list. We suspected that this result would generalize to other types of evaluation tasks. Briefly, participants searched through lists of book titles, performing either a cognitively effortful search task requiring memory of the semantic content of the target (a description of the book, or 'gist'), or a relatively easier stimulus-driven search task (the

verbatim title of the book). We found that notifications reliably harmed the faster, stimulus-driven search tasks more than slow, effortful search tasks. We conjecture that this finding may be based on users needing to re-engage their visual search scanning mechanism from higher- to lower-level features, such as letters, which may be difficult after attending to a notification. A priori, we thought we might observe a beneficial effect for a "highlight", or marker, in this regard. Our approach of persisting a highlighted marker in a list of search results to aid users with task resumption following interruptions can be viewed as analogous to the "linguistic hold," described by Linde and Goguen (1987) for managing interruption during the review of checklists, as described in Section 2.

Our previous study showed that marking the position in a search list with a cursor improved our subjects' performance only during the title search tasks. A more salient marker may have helped more overall, but this was unclear from that study. We offered several potential explanations for the results. One explanation is that users may not have actively employed the cursor for position management and memory, especially in the gist condition. Also, it might 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. Finally, the title or gist of the book participants searched for was continuously displayed during the search trial. Perhaps this design element obviated the need for participants to rely on the marker to reinstate context after an interruption.

We noted a basic navigational confound with the marking procedure used in the first study: to mark the 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 have been that the differences between the efficiencies of these two navigational techniques were masking any performance advantage that a cue to a spatial location might provide.

The experiment in this paper was designed specifically to control for this confound, by equating both search conditions so that the arrow keys are used throughout all trials. In addition, the current experiment allowed for finer timing control of both the subject behaviors, as well as at which times the notifications were actually sent. In the previous work, the experimenter collected reaction time data using observer software (in a *Wizard of Oz* manner). The experimenter also controlled the sending of instant messages. In the experiment we describe here, the software performed all of the data collection, as well as the randomization and timing of when instant messages were sent. Finally, in the current experiment, the title or gist was not available at the top of the search list throughout the trial, as it had been in the previous work. In the current experiment, participants had to explicitly request a reminder of their search target via a button press in order to get a reminder of what they were looking for. We thought this might make the search more difficult across both the title and gist search conditions, possibly improving the chance that the marker might help get back on task after a disruption.

4 User Study: Messages and List Evaluation

Why are instant messages delivered during the evaluation of a list of results in a web search task more costly than messages delivered in other stages of the task? We sought to identify whether or not harmful effects were observed during visual scan, target identification or remembering the task following an interruption, a goal 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. In addition, a "reminder" button was available to assist the user in remembering the search target. The inclusion of this button allowed us to assess the effects of notifications on memory for the primary task following an interruption.

4.1 Subjects

Sixteen participants (nine female) were run through the second study. All were between the ages of 20 and 57, and all but one of the participants had at least tried a chat or instant messaging system before. All were intermediate to advanced computer users. All subjects were run singly for one session.

4.2 Design and Materials

Sixty-four 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 book lists were designed as trials within a Visual Basic 6.0 executable program run on the participants' machines. Each list contained approximately 3 pages worth of search results at a screen resolution of

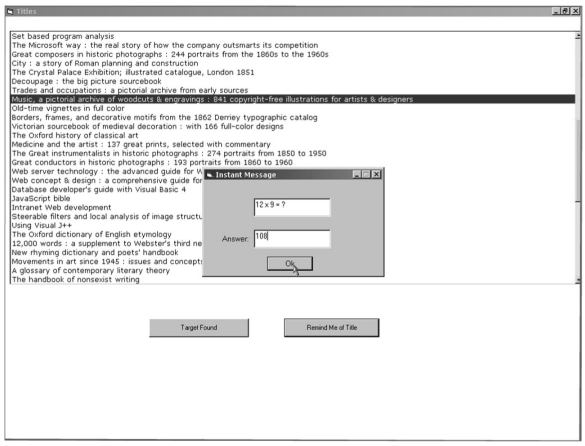

Figure 1. Example of search trial with a marking cursor and an instant message.

1024x768. Figure 1 shows an example of a search trial with stimuli from the experiment. Users navigated the lists in all conditions using the Cursor Up/Down (arrow) keys. In the "marker" condition, a highlight (the default blue highlighted cursor in MS Windows) outlined the currently selected book title, and the list "paged down" automatically when the user got to the last item on the screen. In the no marker condition, navigation was the same, but the cursor was invisible. The paging behavior was identical between the two marker conditions and was similar to the behavior of "Page Down" in web browsers, where the lowest 3 lines remain at the top of the new page.

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. These results were confirmed in an experiment reported earlier (Czerwinski, Cutrell & Horvitz, 2000b). The average length of titles and gists were roughly equivalent.

The experimental design was 2 (title v. gist search trial) x 2 (marker—cursor reverse high-lighting v. no marker—without cursor highlighting) x 2 (notification trial or no notification trial) x 8 (replications per condition) for a total of 64 trials per session. The notifications were sent with pseudorandom timing depending on the quartile of the position of the target within the list of titles (e.g., earlier targets required that the notification be sent to the user earlier). Participants were instructed to request a reminder of what title they were looking for if they forgot it, and selecting a "Reminder" button on the bottom of the search list accomplished this. Dependent variables included *total task time*, *time to switch to a notification*, *number of reminders requested* and *time spent on a notification* (when one occurred).

4.3 Procedure

Participants were greeted and given a tour of the laboratory before starting. Next, participants were asked to read directions describing the search procedures, including how to navigate using the arrow 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.

During the experiment, participants were asked to do 2 kinds of searches: For half of the trials, we gave them a book title as their target and they simply had to scan the list for the title. For 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 the trials participants were notified with an instant message in a custom window that mimicked the sound and onset of MSN Messenger 2.0, containing a math problem to solve (simple multiplication and division problems). 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 indicated this by pressing the "Book Found" button on the screen, or by hitting Enter. After 32 search trials, the participant took a short break and then returned for the second half of the session. Order of marking technique was determined at random for each participant and blocked. 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 gratuity for their participation.

5 Results

Analyses were carried out for both the average trial times, as well as the average log trial times. Since the pattern of results was the same, we report the results from the average log trial times in order to comply with the assumptions required for performing Analysis of Variance. Logging the data normalizes the reaction time distributions and countermands much of the skew common in reaction time distributions.

5.1 Reaction Time

A repeated measures Analysis of Variance (ANOVA) of the log reaction time data was carried out for the factors of search trial type (title v. gist), notification (present or absent) and marker (present or absent), for a 2x2x2 ANOVA. Results showed no overall main effect of marker, but a significant main effect of title v. gist search, $F(1,11)=123.2$, $p<.001$, and of notification, $F(1,11)=48.8$, $p<.001$. Title searches were reliably faster than gist searches, and notification trials were reliably slower than those without notifications. Also, participants took longer to switch to the instant message notification in the gist condition, $F(1,11)=18.9$, $p<.01$. These findings are shown in Figures 2 and 3. This confirms that we replicated our results from the first study (Czerwinski, Cutrell & Horvitz, 2000b), using improved data collection techniques, a more prominent marker, and despite hiding the search target behind a reminder button.

5.2 Use of Reminders

A 2 (search type) x 2 (marker or not) x 4 (quartile of 80 titles in which notification was sent) ANOVA of the number of times participants requested reminders revealed several interesting results. Reminders were requested significantly more often in the gist condition than during title search, $F(1,11)=14.9$, $p<.01$, and significantly more often when a notification was sent during the search trial, $F(1,11)=9.1$, $p<.01$. No reliable effect of marking was observed in the reminder data. This pattern of results suggests that gist search was indeed more difficult, as evidenced by participants needing to request reminders of their target more often during gist trials.

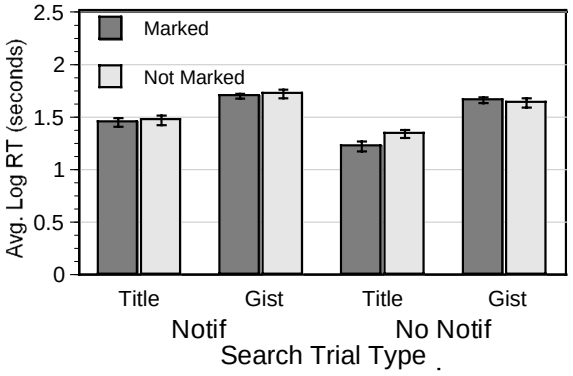

Figure 2. Effects of notification, search type and marker on overall mean log trial times.

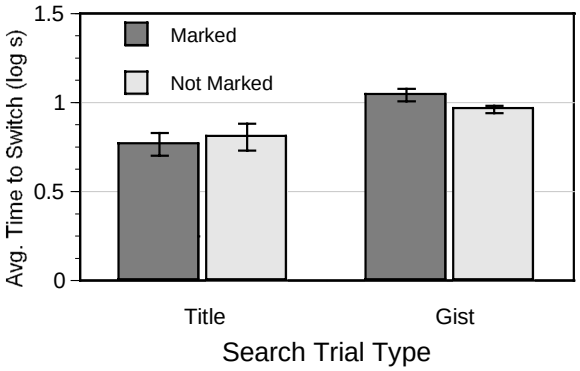

Figure 3. Mean time to switch to a notification depending on the type of search.

We also found that interruptions can influence the use of reminders. For the gist condition, reminders were requested significantly more often in cases where users were interrupted than in cases where no instant message was sent.

Finally, we found a significant relationship between the use of reminders and the timing of the interruption. We discovered that the use of reminders depends on which quartile of the search list the user is reviewing when a notification is sent. If a notification was sent earlier in the trial, participants were significantly more likely to request a reminder than if the notification came later in the trial, $F(3,33)=3.0$, $p<.05$. This was an interesting finding and may have to do with the amount of time participants had available to focus on, or rehearse the

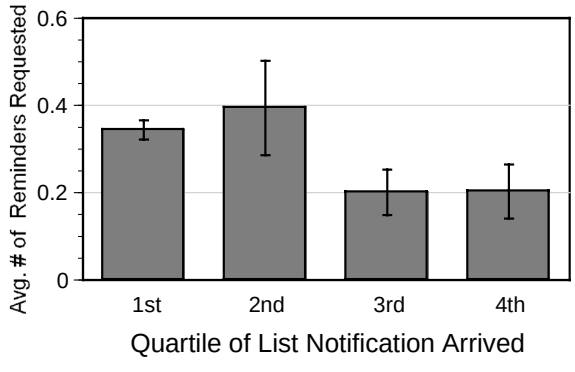

Figure 4. Mean number of reminders based on which quartile of the search list a notification was received.

target prior to getting interrupted. These results are shown in Figure 4.

6 Discussion

This experiment was carried out as an attempt to verify the findings of previously reported work, correcting for a confound between the marked and unmarked condition in the first experiment, using a more distinctive marker, and automating the experimental procedures. In addition, the target's title or gist was not displayed throughout the search trial as it had been in the previous work. Instead, a reminder button was available for participants to invoke if necessary. Despite all of these changes to the experimental procedure, the findings from this experiment closely replicate and extend those from the previous work. Although there were significant main effects for the presence of notifications and for gist search trial types, we confirmed our earlier assessment that there is little to no benefit of having a marker present after a notification was received. Participants were reliably slower overall after receiving an instant message, and the cost of the interruption was higher in the faster title search condition. Participants took significantly longer to switch to a message during the slower gist search trials.

We identified several results with regard to the use of reminders during scenarios where users performing a primary task were interrupted by messages. Participants used reminders more often during gist trials, especially after an instant message was received. We found that reminders were used more often if the instant message was received earlier in the search trial, regardless of search type. This finding could be a direct result of participants getting more time to learn or rehearse the target prior to receiving an instant message later in the search list. This result has implications for automated systems for delivering notifications. For example, it may be less disruptive in some situations to delay the transmission of notifications in situations where a user has just initiated a new task. Overall, these results on interruption and memory suggest that methods for securing and recovering task focus can provide value.

7 Summary

Over a series of experiments we have shown the disruptive effects of notifications on a variety of ongoing computing tasks. We confirmed the predictions of Miyata and Norman (1986) that some task phases are less amenable to interruption than others. In previous work, 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). The results from the present study extend those findings by demonstrating the harmful effects of notification delivery on memory for the prior task early in a task's lifecycle, despite earlier findings that sending notifications early is better for overall task time. We believe that these combined results can provide guidance for designers of instant messaging systems.

Although our marking procedure was not as effective a reminder as we had hoped in this experiment, we remain optimistic that notification systems might one day employ a variety of reminders, including 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. Such tools could be helpful in getting users back on task more quickly after notifications. Pilot studies in our lab have shown that as little as a text sentence describing a previous task can be an effective tool to get users back on track after a notification. We are pursuing research on HCI designs and procedures for reinstating context after a notification.

References

Altmann, E.M. and Gray, W.D. (2000). Managing attention by preparing to forget. Proceedings of the IEA 2000/HFES 2000 Congress. San Diego, CA: Human Factors and Ergonomics Society, pp. 152-155.

Bailey, B.P., Konstan, J.A. and Carlis, J.V. (2000). Measuring the Effects of Interruptions on Task Pedrformance in the User interface. IEEE Conference on Systems, Man and Cybernetics. Nashville, TN.

Czerwinski, M., Cutrell, E. & Horvitz, E. (2000a). 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, p. 71-76.

Czerwinski, M., Cutrell, E. & Horvitz, E. (2000b). Instant Messaging and Interruption: Influence of Task Type on Performance, To appear in *Proceedings of OZCHI 2000*, Sydney, Australia, Dec. 4-8.

Czerwinski, M. P., Chrisman, S. E., & Rudisill, M. (1991). Interruptions in Multitasking Situations: The Effects of Similarity and Warning (Technical Report JSC-24757): National Aeronautics and Space Administration, Lyndon B. Johnson Space Center, Houston, Texas.

Czerwinski, M., Chrisman, S. & Schumacher, B. (1991). The effects of warnings and display similarities on interruptions in multitasking environments, SIGCHI Bulletin, 23(4), pp.38-39.

Entin, E. (2000). Representing and visualizing a dynamically changing tactical situation. . In Proceedings of the IEA 2000/HFES 2000 Congress. San Diego, CA: Human Factors and Ergonomics Society, pp. 427-430.

Hess, S. M., & Detweiler, M. C. (1994). *Training to Reduce the Disruptive Effects of Interrutpions.* Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting.

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

Gopher, D., Greenshpan, Y., Armony, L. (1996). Switching attention between tasks: Exploration of the components of executive control and their development with training. In Proceedings of the 40th Annual Meeting of the Human Factors and Ergonomics Society, *: Human Factors and Ergonomics Society, pp. *.

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.

Jambon, F. (1996). Formal modeling of task interruptions. In Proceedings of CHI ’96, Human Factors in Computing Systems, Volume 2, pp*, ACM Press.

Linde, C. & Goguen, J. (1987). Checklist interruption and resumption: A linguistic study. NASA Technical Report NASA-CR-177460, Ames Research Center, Moffett Field, CA.

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.

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.

O’Connail, , B. & Frohlich, D. (1995). Timespace in the workplace: Dealing with interruptions. In Proceddings of the ACM CHI;95 Conference on Human Factors in Computing Systems, Vol. 2, pp. 262-263.

Renaud, K. (2000). Expediting rapid recovery from interruptions by providing a visualization of application activity. To appear in *Proceedings of OZCHI 2000*, Sydney, Australia, Dec. 4-8.

Rhodes, J.S., Benoit, G.E. and Payne, D.G. (2000). Factors affecting memory for dynamically changing system parameters: Implications for interface design. In Proceedings of the IEA 2000/HFES 2000 Congress. San Diego, CA: Human Factors and Ergonomics Society, pp. 284-285.