The Effects on Behaviour and Efficiency by Varying Multitasking Conditions

(Bachelorproject)

Tom Janssen* (s1534335, tjanssen@ai.rug.nl)
Supervised by Niels A. Taatgen* (niels@ai.rug.nl), Jelmer P. Borst* (jpborst@ai.rug.nl)

July 11, 2011

Abstract

Recent studies have showed that when computer users are faced with a deferrable task interruption while multitasking, users tend to defer the processing of such an interruption until times of lower workload (Salvucci and Bogunovich, 2010). Also, users can be tempted to processing a deferrable task interruption at times of higher workload when faced with a forced pause during the primary task. (Timmerman, 2011) In our experiments, we investigate the influence on multitasking behaviour and efficiency by varying: (1) the difficulty of the interrupting task, (2) forced pauses during the primary task and (3) a penalty for not maintaining information required to do the primary task. We found that the difficulty of the interrupting task has little or no influence on behaviour and efficiency, whereas forced pauses and a time penalty can have different influences on multitasking behaviour and task efficiency.

1 Introduction

The ability for humans to multitask is a very important one. Many complex tasks like cooking or driving a car require many subtasks to be done. For example, someone driving a car must control the vehicle, but must also observe his surroundings and may even have a conversation with his passengers. Because of our limited mental resources, some tasks can be done simultaneously (e.g. walking and having a phone call) and other tasks need to be done sequentially (e.g. reading and watching the news). This dichotomy is also known as concurrent and sequential multitasking (Salvucci, Taatgen, and Borst, 2009). In sequential multitasking, a primary task may be interrupted by a secondary task (e.g. a phone call while reading a newspaper). One may choose to do the secondary task (and interrupt the primary task) or continue with the primary task (and leave the secondary task to do at a later time or not at all). Switching to an interrupting task and back may have time costs. Before and after the execution of the interrupting task, there may be an interruption lag and a resumption lag respectively (Trafton, Altmann, Brock, and Mintz, 2003). The interruption lag is used to store or rehearse information needed to do the primary task if necessary, whereas the resumption lag is used to recall this information (Salvucci et al., 2009). In terms of the ACT-R theory this information is called problem state information (Anderson, 2007; Borst, Taatgen, and Van Rijn, 2010). The problem state resource is limited to store one chunk of information about the task being done and may be used in simple tasks such as arithmetic or reading. We say the mental workload is high when a problem state is being maintained and low when no problem state is being maintained. Switching to an interrupting task during high workload may result in the loss of problem state information or additional costs for rehearsing or recalling problem state information. If an interrupting task is deferrable (e.g. answering an incoming chat message) it may be better to do the interrupting task at a moment where mental workload is lower for the primary task.

Salvucci and Bogunovich (2010) have shown in

*University of Groningen, Department of Artificial Intelligence

their research that computer users tend to defer incoming chat messages when their mental workload of the primary task is high, to a moment where their mental workload is lower. In their experiment, participants had to answer emails as an employee of a customer service. Emails were from customers asking for prices of certain products. Participants had to look up these prices in a web browser and send an email back to the customers. In this primary task, participants had times of high mental workload (remembering specific product or price) and times of low mental workload (processing emails). At evenly distributed times (high and low workload) chat messages would come in, asking if the participant had seen a certain movie. These could be answered with a simple yes or no. Participants switched to this secondary task 94% at times of low workload as opposed to 6% at times of high workload. Therefore, participants tended to switch to the secondary task at times of lower workload, thus deferring incoming task interruptions.

Timmerman (2011) adapted this experiment to investigate if computer users could be seduced into switching to the secondary task at high workload if there was some forced pause during the high workload execution of the primary task. In this experiment, the browser was modified to enforce a delay after clicking a link, before the next page would show up. Results showed that participants could indeed be tempted to switch to the secondary task if the primary task was interrupted by a pause. However, these results also showed that participants did not significantly lose efficiency by doing so. Not counting the time of the forced pauses, participants spent equal amounts of time to do the primary tasks, despite the change in behaviour. In this case, the benefits of switching during a forced pause were most likely countered by the costs for switching during times of high workload, resulting in little or no change in efficiency. Other experimental conditions may result in other cost/benefit ratios resulting in a change in efficiency.

More research may lead to a better understanding of the behaviour associated with task interruptions. Also, a better understanding of behaviour in ecologically plausible tasks as emailing and chatting may help improve the development of interfaces for computer users. In the two experiments described in this article we investigated what influence a more difficult secondary task has on the deferring behaviour of participants. We adapted the previous experiments to include chat messages for the secondary task that require more than just a yes or a no. In these messages, participants were asked for their favourite things in different categories and were thus required to reply in a more complex and creative way. Participants had to go through four rounds that could have either simple or complex chat messages, and with a browser delay (as previously described) or not.

Because of the increased complexity of the secondary task, more mental resources are needed for the participant to come up with a decent answer. If the participant is maintaining problem state information for the primary task, it might be very unwise to switch to the secondary task, even if a forced pause interruption comes in. The main question investigated in this paper is: What are the effects of a more difficult secondary task with respect to user behaviour and task efficiency? Will users be tempted to defer harder secondary tasks more easily than easy secondary tasks? And will users become less efficient by more difficult task interruptions?

2 Experiment 1

Two experiments were done in total. The first experiment consisted of four conditions created from two dichotomies, easy/hard secondary task and a (approximately) three-second browser delay or not. The four conditions were used within participants and the sequence of conditions (out of 24 possible sequences) was counterbalanced and unique for all participants. The second experiment is a slightly modified version of the first experiment that was conceived after analysing the data of the first experiment.

2.1 Tasks

Participants were instructed to perform two tasks in a fictional desktop environment on an Apple computer. An example of the windows is shown in Figure 1. The primary task consisted of: (1) reading and answering emails from customers as if the participant was working for a customer-service department and (2) browsing through pages to find the corresponding answer to the emails. The sec-

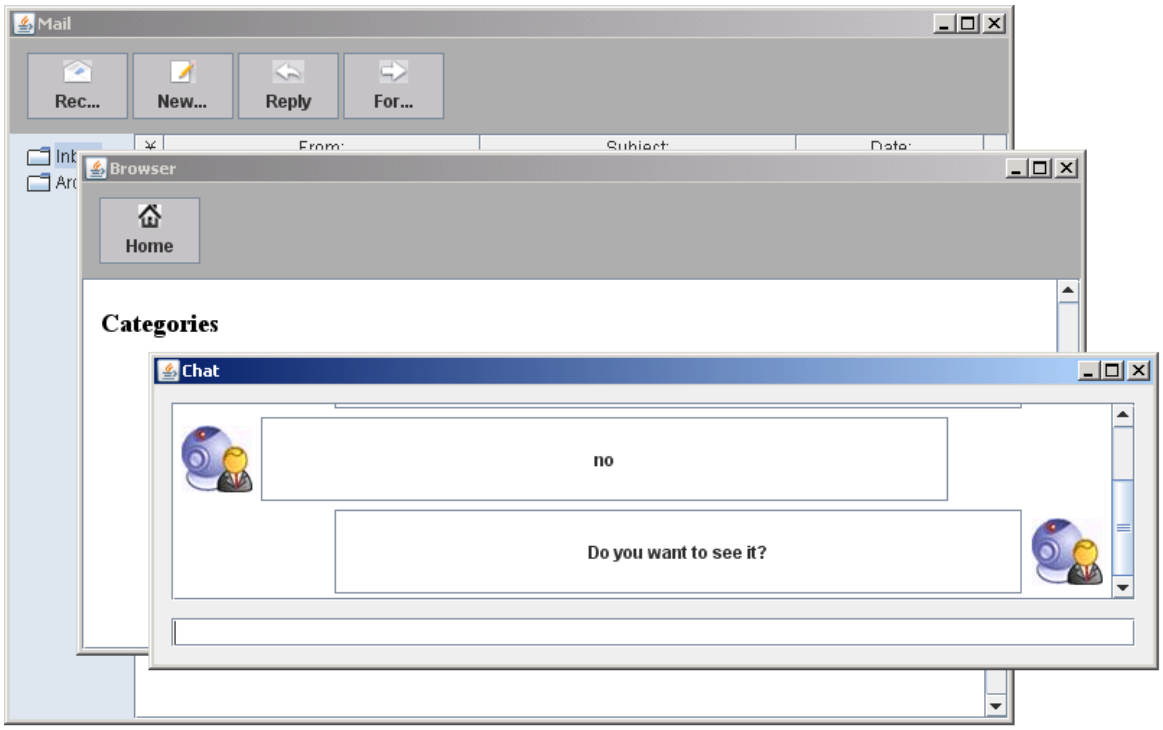

Figure 1: During the experiment, windows overlapped so the participant could only see the contents of one window at a time.

ondary task was reading and answering incoming chat messages.

In the primary task, emails had to be selected and read in a specific mail window. All emails were inquiries about the price of a specific product. Emails were randomly generated from lists of different customer names and different products. The products were defined by type of product, manufacturer and model code. (e.g. photocamera, Woksan, W-40) The names of the manufacturers were fictional and the model code always consisted of one letter and two digits. The products and corresponding prices could be looked up in a browser window by navigating through pages by clicking on the products type, manufacturer and model. The navigation could be reset by clicking the home-button in the browser window. After finding the price, participants had to click on reply in the mail window, which opened up a compose window in which the price could be typed and sent by pressing a button. The price of the product alone was enough to type as an answer. There was no effect if the participant answered the wrong price, but all answers were logged by the program. Before selecting the next email to repeat the task, the answered email had to be archived by dragging the email to the archive folder. Depending on the condition of the experiment, navigating through the pages of the browser by clicking the product's type or manufacturer could result in a delay before showing the requested page. The delay lasted for 3000 milliseconds with a 35% noise rate. The noise was added to make the duration of the delay unpredictable for the participant and to make the simulated loading of the page look more natural. During the delay, the browser window would show an animated icon suggesting the loading of a page.

The second task was reading and responding to chat messages. During different stages of the primary task a chat message would appear in a chat window asking a question. An incoming message was notified by an alerting sound and the chat window getting a yellow colour instead of grey. (Part of the chat window was visible at all times without showing the contents.) Depending on the stage of the experiment, the questions were either easy

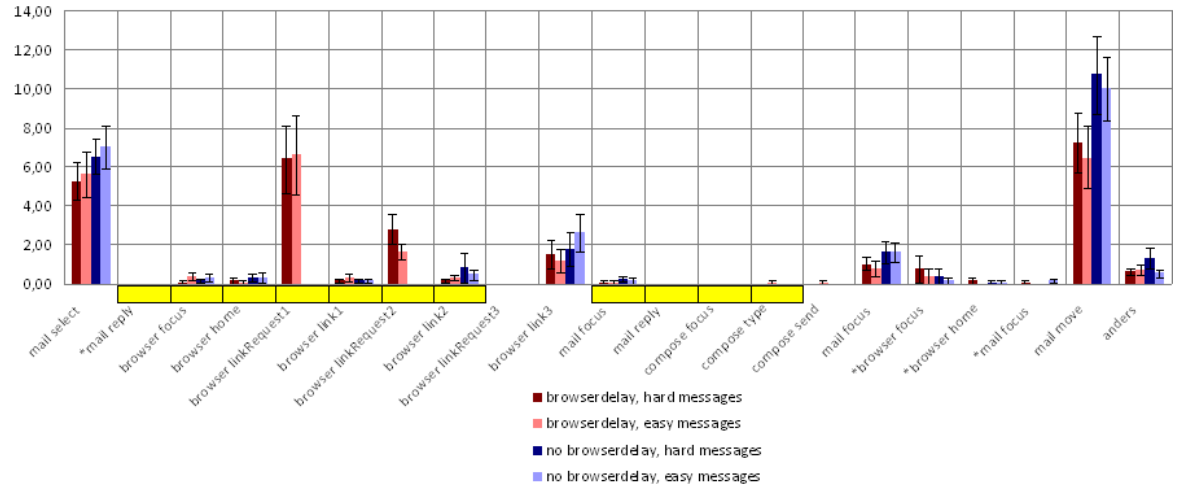

Figure 2: Number of switches for all action steps. The yellow marks highlight the moments of high workload. Asterisks(*) highlight alternative sequences.

or hard. Easy questions were about whether or not the participant had seen a certain movie (randomly chosen from a list of 50 Academy Award Best Picture nominees from 1999-2008) and could be answered with a simple yes or no. Hard questions were open questions about what the participant's favourite thing is of different categories (e.g. songs, sports, quotes, animals) and required a more creative answer. This changed the nature of the second task, possibly changing the overall multitasking behaviour. After half of the questions, a follow up question would appear. For easy questions, the message would ask if the participant liked the previous movie or would like to see that movie, without mentioning the movie again. For hard questions, the participant was asked for his least favourite thing in the afore mentioned category, without mentioning the category. Users were able to scroll within the chat window to see what the movie or category was, but the idea is for users to maintain a problem state for the chat task that could potentially interfere with the mail task.

None of the windows could be repositioned, nor was the use of keyboard shortcuts possible. By not allowing the repositioning of windows, participants were forced to actively switch between windows to be able to view the contents, allowing the program to register the switches between the tasks. The task environment was coded in Java Swing to match the appearance of standard Macintosh applications. During the experiment the software would write all user actions including timestamps to a logfile.

2.2 Procedure

To allow both national and international students to participate, both a Dutch and English version of the experiment was written. Participants were first asked to read detailed instructions on paper (including images) about the tasks. They were then given the chance to try out the tasks and ask questions about it before the actual experiment. In each of the four conditions, 24 chat messages had to be answered, including follow up questions. This resulted in at least 24 switches to the chat window. Between the conditions, participants were instructed to take a short break before resuming with the next condition. The duration of the sessions in the first experiment lasted from 50 to 85 minutes.

2.3 Participants

In this experiment 7 men and 6 women participated. The age of the participants ranged from 19 to 32 years with an average of 24 years. Participants were local and international students and gave informed consent for the experiment. As a reward participants could take a monetary reward (EUR 12) or course credit.

		Exp 1		Exp 2	
Measurement	Between Conditions	F(1,12)	P(>F)	F(1,10)	P(>F)
Percentage of lowload switches	Delay/No Delay	19.43	**<0.001**	22.80	**<0.001**
	Easy/Hard Chat	0.14	0.72	0.57	0.47
Average time on trials	Delay/No Delay	0.97	0.35	29.42	**<0.001**
	Easy/Hard Chat	0.22	0.65	4.04	0.07
Total number of switches	Delay/No Delay	6.02	**0.03**	1.38	0.27
	Easy/Hard Chat	1.85	0.20	2.27	0.16

Table 1: Overview of ANOVA Results

3 Results

The generated log files were first analysed by looking at what moments the participants switched from the first task to the second. Switches were classified by observing the last action steps taken before the switch occurred. Because after careful analysis some switches remained ambiguous, these switches had to be classified as other and could not be included in further analysis. All necessary steps of a trial with its corresponding number of switches are seen in Figure 2. In this figure, a different behaviour can be observed between the delay and no delay conditions. Users are more tempted to switch to the secondary task during high workload if there is a forced pause in the primary task. Repeated measures analysis of variance (ANOVA) showed that the different behaviour in delay and no delay conditions is indeed significant, but no significant difference was found between easy and hard chat messages. (See Table 1 for an overview of ANOVA results.)

By analysing the average time spent on the trials (answering emails), efficiency could be tested. Because in all conditions the nature of the tasks were different, some corrections had to be made to the total time spent in one condition. The conditions with a delay had all the delays subtracted from the total time. Also, the time spent on the secondary task was subtracted from all conditions, because the hard questions need more time to be answered. Finally the total time was divided by the number of answered emails. No significant difference was measured in the remaining time of the conditions, indicating that participants did not gain or lose efficiency in the first task if the chat messages are of a different nature or if participants changed their switching behaviour as a result of a browser delay.

The number of switches per condition proved to be significantly higher for conditions with a browser delay. Because the number of chat messages is static, this could only mean that participants often switched more than one time to the secondary task. Further inspection of the data suggested that participants chose to read the chat message during a browser delay and reply at a later time, probably when they had come up with an answer. The data also showed that participants almost always had to look up the price or product again after answering a chat message, indicating that the participant almost never remembered problem state information required for the primary task. Because we want to discourage people from forgetting problem state information and from switching to the chat message twice, we did some modifications to the first experiment to create a second experiment.

4 Experiment 2

4.1 Setup

In the second experiment the same tasks had to be done with the same conditions. To discourage participants from switching twice to answer one chat message and encourage them to remember problem state information, a penalty was given if the participants switched back to the primary task. Switching to the mail window now resulted in a delay of 2000 milliseconds with 35% noise.

In this experiment 4 men and 7 women participated. The age of the participants ranged from 19 to 26 years with an average of 22 years.

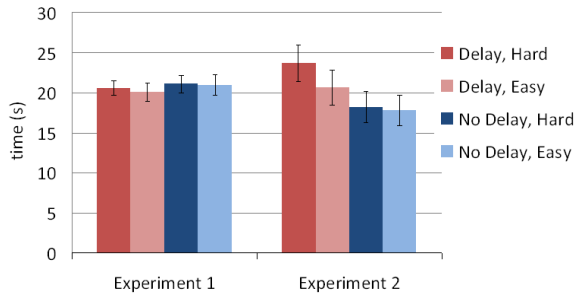

Figure 3: Average time per trial

4.2 Results

We first tested the number of switches per condition by analysis of variance. Unlike the first experiment, the difference between the conditions was insignificant. This could mean that the penalty for switching back to the email task caused participants to no longer switch to the chat task more often than necessary. The percentage of low workload switches was also increased for the condition with a browser delay as opposed to the first experiment. Despite the increase in low workload switches the difference between delay or no delay was still significant. The efficiency of the participants was tested the same way as before, but now the time penalty for switching to mail and browser windows was also subtracted from the total time. Unlike the first experiment, the efficiency was now significantly lower in conditions with a browser delay (see Figure 3). Also, the efficiency of conditions with hard chat messages proved almost significantly lower.

5 Discussion

The results of the two experiments are consistent with previous studies of Salvucci and Bogunovich (2010) and Timmerman (2011). Users choose to defer task interruptions to a point of lower workload. By adding a forced pause, users can be tempted to switch to a secondary task at higher workload. During the first experiment, the efficiency of the participants was stable under all conditions. But during the condition with hard questions and a browser delay users adapted their strategy. When a browser delay occurred, they would quickly read the chat message, switch back to the browser and answer the message at a later time, possibly when they had thought of an answer. Participants also generally forgot problem state information for the primary task, so they had to switch back to the mail window very often after answering a chat message.

In the second experiment, a penalty for switching to the mail window was added that would discourage participants to switch to this window more than necessary. Participants no longer switched windows any more than they should, but efficiency significantly decreased if there was a browser delay. We assume that this decrease in efficiency is caused by the interruption lag and resumption lag as previously discussed. With the addition of a penalty for switching to the mail window, participants are more encouraged to maintain problem state information. Therefore, if a browser delay is encountered, participants are less likely to switch to the interrupting task (See FIGURE). They are also more likely to take time to rehearse and recall problem state information if they do choose to switch to the interrupting task causing them to be less efficient.

As opposed to what we suspected before the experiments the variance of easy and hard chat questions had little or no effect on the multitasking behaviour of the participants.

The results of these experiments show that people are generally very adaptive to minor changes to the multitasking conditions. The adaptive behaviour is most likely caused by the limitations of the problem state resource, in further detail described as the cognitive bottleneck by Borst, Taatgen and Van Rijn (2010). It is likely that people adapt their behaviour to find an optimal way for using the time and mental resources they think they have. To get a better understanding of the problem state resource, further experiments varying the problem state information of the email task may be interesting.

References

J. R. Anderson. *How can the human mind occur in the physical universe?* New York: Oxford University Press, 2007.

J. P. Borst, N. A. Taatgen, and H. Van Rijn. The problem state: A cognitive bottleneck in multitasking. *Journal Of Experimental Psychology-*

Learning Memory And Cognition, 36(2):363–382, 2010.

D. D. Salvucci and P. Bogunovich. Multitasking and monotasking: The effects of mental workload on deferred task interruptions. *CHI2010*, 1:85–88, 2010.

D. D. Salvucci, N. A. Taatgen, and J. P. Borst. Toward a unified theory of the multitasking continuum: From concurrent performance to task interruption, and resumption. *CHI2009*, 10:1819–1828, 2009.

J. Timmerman. Multitasking: The effects of pause as a task interruption. Bachelor's thesis, University of Groningen, Department of Artificial Intelligence, 2011.

J. G. Trafton, E. M. Altmann, D. P. Brock, and F. E. Mintz. Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies*, 58:583–603, 2003.