Advance Access publication on 9 March 2014
doi:10.1093/iwc/iwu005

The Effects of Task Difficulty and Multitasking on Performance

RACHEL F. ADLER[1,2,*] AND RAQUEL BENBUNAN-FICH[3]

[1]*The Graduate Center, CUNY 365 Fifth Avenue New York, NY 10016, USA*
[2]*Northeastern Illinois University, 5500 North St. Louis Avenue, Chicago, IL 60625-4699, USA*
[3]*Baruch College, CUNY One Bernard Baruch Way, Box B11-220, New York, NY 10010, USA*

**Corresponding author: r-adler@neiu.edu*

Multitasking is prevalent during computer-mediated work. Users tend to switch between multiple ongoing computer-based tasks either due to a personal decision to break from the current task (self-interruption) or due to an external interruption, such as an electronic notification. To examine how different types of multitasking, along with subjective task difficulty, influence performance, we conducted a controlled experiment using a custom-developed multitasking environment. A total of 636 subjects were randomly assigned into one of the three conditions: *discretionary*, where they were allowed to decide when and how often to switch tasks; *mandatory*, where they were forced to switch tasks at specific times; and *sequential*, where they had to perform tasks in sequence, without switching. The experimental environment featured a primary problem-solving task and five secondary tasks. The results show that when the primary task was considered difficult, subjects forced to multitask had significantly lower performance compared with not only the subjects who did not multitask but also the subjects who were able to multitask at their discretion. Conversely, when the primary task was considered easy, subjects forced to multitask had significantly higher performance than both the subjects who did not multitask and the subjects who multitasked at their discretion.

RESEARCH HIGHLIGHTS

- The study compares the effects of different types of multitasking and subjective task difficulty with an experiment.
- It uses a custom-developed multitasking environment with three conditions.
- Compares performance scores of mandatory, discretionary and no multitasking.
- Those forced to multitask performed the worst when the task was deemed difficult.
- However, when the task was deemed easy, those forced to multitask performed the best.

Keywords: user studies; laboratory experiment; empirical studies in HCI

Editorial Board Member: Marta Kristín Lárusdóttir

Received 27 November 2012; Revised 17 November 2013; Accepted 16 January 2014

1. INTRODUCTION

Multitasking is a prevalent behavior when using personal computers or mobile platforms. Both at home and in the workplace, people are frequently switching tasks to check their email, their social networking site, or another website. Studies report that computer users have multiple applications open, and switch between them frequently (Crook and Barrowcliff, 2001; Czerwinski *et al.*, 2004). In fact, managing one's email alone involves a lot of multitasking (Bellotti *et al.*, 2005). Multitasking is defined as the performance of several tasks at once (Rubinstein *et al.*, 2001). However, depending on how tasks and times are defined, several interpretations are possible (Benbunan-Fich *et al.*, 2011). Some researchers reserve the term multitasking for simultaneously conducted activities (Meyer and Kieras, 1997), while others define it in terms of task switching (Czerwinski *et al.*, 2004). More encompassing definitions accommodate all possible cases. For example, Salvucci and Taatgen (2011) defined a multitasking continuum based on the average time spent on one task before switching to another. At one extreme, there are tasks that involve highly frequent and sometimes imperceptible switching, such as talking while driving. At the other extreme, there are tasks that involve longer spans between switches, such as writing a paper and reading email.

Prior research has identified two different drivers of multitasking: *external* interruptions and *internal* decisions to stop ongoing tasks (Gonzalez and Mark, 2004; Mark *et al.*, 2005; Miyata and Norman, 1986). An external interruption occurs when an event in the environment forces a user to switch tasks, while an internal interruption comes from one's self, i.e. self-initiated, when a user decides to switch tasks at his/her discretion (Miyata and Norman, 1986). Self-initiated interruptions occur just as often as external interruptions (Gonzalez and Mark, 2004).

While multitasking has been examined in the human–computer interaction (HCI) literature, there is still ample opportunity to extend research on this topic (McCrickard *et al.*, 2003c). There are at least two areas where additional research might be fruitful. One is the study of voluntary task switching. The work of Payne *et al.* (2007) established that people switch away from tasks that are no longer rewarding. Related research studies by Janssen *et al.* (2011) and Duggan *et al.* (2013) have incorporated explicit payoff structures (rewards) to investigate in more depth the determinants of voluntary task interleaving. The second area that could benefit from additional research is the study of how multitasking affects performance. The existing literature in this regard is somewhat fragmented. Some studies have examined how users' performance is impacted when receiving external interruptions (Bailey and Konstan, 2006; McFarlane, 2002; Speier *et al.*, 2003). Other studies have focused on the relation between discretionary multitasking and the resulting performance. The findings suggest that although some amount of multitasking may not be detrimental for performance (Davidson, 2011; Palladino, 2007), intensive multitasking, characterized by high frequency switching and a large number of ongoing tasks, tends to degrade performance (Adler and Benbunan-Fich, 2012; Bailey and Konstan, 2006; Hembrooke and Gay, 2003).

The present study seeks to systematically compare different types of multitasking and its effects on performance. Using a laboratory experiment, three types of multitasking resulting from mandatory interruptions (Bailey and Iqbal, 2008), discretionary self-interruptions (Payne *et al.*, 2007) and sequential execution where all tasks are performed in succession (Salvucci and Bogunovich, 2010) are compared. These alternative multitasking scenarios are examined in conjunction with subjective task difficulty to investigate the effects on performance. An analysis of the differences in performance among these scenarios will enable HCI researchers to have a better understanding of the positive or negative impacts of multitasking.

2. RELATED WORK

2.1. Mandatory multitasking

In HCI, multitasking has been examined from the perspective of external interruptions and the role of notification systems (McCrickard *et al.*, 2003a,b; McFarlane, 2002; McFarlane and Latorella, 2002; Oulasvirta and Saariluoma, 2004, 2006; Trafton *et al.*, 2003). Interruptions tend to have negative effects on performance (Oulasvirta and Saariluoma, 2004, 2006). In particular, being interrupted with a secondary task can impact performance on the primary task because of the extra time and effort needed to recall the primary task when it is resumed.

Interruptions are a frequent occurrence in computer-mediated work (Iqbal and Horvitz, 2007). Bailey and Konstan (2006) found that when interrupted, users needed more time to finish their primary task, made more errors in both tasks and had more annoyance and anxiety than those who were not interrupted. A substantial body of research has examined the disruptive effects of interruptions and has documented that increased complexity in the interrupting task leads to slower resumption times (Hodgetts and Jones, 2006) and lower primary task accuracy (Gillie and Broadbent, 1989). Cades and colleagues found that interruption complexity, defined by the number of mental operators required to complete a task, reduces the opportunity for rehearsal in the primary task, leading to an increase in the disruptiveness of the interruption (Cades *et al.*, 2007, 2010).

Of the four different types of interruptions (immediate, negotiated, mediated and scheduled) identified by McFarlane (2002), immediate is the most detrimental for performance. The other types are not as detrimental because the user has some level of control, as he/she can decide whether or not to respond immediately (negotiated), a middle agent determines whether the interruption will occur (mediated) or the interruptions occur at predetermined intervals (scheduled).

Receiving interruptions during a task and being forced to respond at that moment is disruptive and causes users to lose their thought process and control during the performance of a task (Altmann and Trafton, 2002). Bogunovich and Salvucci (2011) discuss the concept of cognitive load interruptibility and argue that forced interruptions are less disruptive when the cognitive load is low. A key determinant of cognitive load is the level of difficulty of a task, which can be assessed through an objective measure of task complexity or through a subjective perception of complexity (Maynard and Hakel, 1997). From an objective perspective, task difficulty can be determined by task designers based on an estimation of the amount of mental resources required to complete a task. In contrast, subjective task difficulty refers to the perception that some tasks seem harder due to an intuitive sense of difficulty (Cades *et al.*, 2008).

The impact of interruptions are contingent upon the level of difficulty of the task being performed (Gillie and Broadbent, 1989). For example, Speier *et al.* (2003) found that interruptions helped improve performance on simple tasks but hurt performance on more complicated tasks. When users are interrupted during complex tasks, their cognitive ability is impaired and task performance suffers. During a complex task, a distraction can interrupt the user's concentration and therefore can cause negative results (Altmann and Trafton, 2002). However, during simple tasks, where users do not have to invest a substantial amount of cognitive resources on the task at hand, interruptions can actually help them focus their attention (Speier *et al.*, 2003), thereby improving their performance.

2.2. Discretionary multitasking

Multitasking also occurs when users decide at their own volition to interrupt the current task to pursue another one. Jin and Dabbish (2009) identified seven categories of internal interruptions. These categories explain why a user would switch to another task: adjustment, break, routine, wait, inquiry, trigger and recollection. A user may need to take a *break* when frustrated or tired, or multitask due to a *trigger* or *recollection* when recalling a related or completed new task. People also multitask due to *routine*, such as checking one's email out of habit, or they may multitask due to necessary *adjustment*s of the working environment. Other causes of multitasking include a *wait*, which involves filling downtime during a task, or an *inquiry* to receive necessary information that will help complete the task.

Discretionary multitasking has also been examined in the psychology literature. Payne *et al.* (2007) conducted a set of experiments designed to investigate different types of multitasking. In their second experiment, participants were performing two similar computer-based tasks and were allowed to switch between these tasks at will. The results of this experiment indicated that people switched either because tasks were no longer rewarding or because they finished a sub-goal and decided to take a break from the current task by attending to another. In fact, when given a choice, people prefer to switch at low cognitive load points (Bogunovich and Salvucci, 2011) because workload decreases upon the completion of a subtask and the disruptive effects of interruptions are minimized at natural breaking points (Bailey and Iqbal, 2008).

In terms of task difficulty, Czerwinski *et al.* (2004) found that complex tasks were more difficult for subjects to resume. However, given a set of tasks, the level of difficulty affects which tasks subjects decide to pursue, the order in which these are executed and the extent to which they are interleaved (Yeung, 2010). When faced with multiple tasks, people can strategically control their allocation of attention to maximize their payoffs and meet specific performance goals (Duggan *et al.*, 2013; Janssen and Brumby, 2010; Janssen *et al.*, 2011).

2.3. Sequential task completion

While conceptually different, both discretionary and mandatory multitasking are theoretically important, particularly when compared with sequential task performance, which is free from interruptions. The most important difference is that the user controls the pace and timing of self-interruptions in discretionary multitasking, but does not control them in the mandatory interruption scenario. The sequential scenario, where tasks are performed consecutively and without interruptions, serves as a control condition to systematically compare different types of multitasking. In sequential execution (also called serial or mono-tasking), only one task is executed at a time from beginning to end. Although multiple tasks are completed in a time frame, there is no task interference and no switching. Thus, this mode is widely used to establish a baseline condition for performance.

2.4. Performance effects and task complexity

The relation between multitasking and performance can be explained from the perspective cognitive skills/abilities or with other factors, such as personality traits or psychological states. The level of arousal is one of the factors that has been used to explain the effects of multitasking on task performance (Oswald *et al.*, 2007). Complex tasks produce higher levels of mental workloads and lead to higher arousal than easier tasks. Therefore, the level of difficulty of a task imposes mental workload demands on the performer that interacts differently with task interruptions. At low levels of workload, performance is compromised due to inattention and lack of stimulation, while at high levels, performance also suffers due to the cognitive inability to deal with overload. Optimum performance is in the middle, where there is the right combination of workload and attention. This inverted-U relationship between workload and performance is known as the Yerkes–Dodson law (Yerkes and Dodson, 1908). According to this law, easy tasks produce low levels of arousal, and performance can improve when the user faces additional stimuli (Teigen, 1994). Therefore, receiving interruptions during an easy task may help performance. In contrast, because difficult tasks already require substantial cognitive resources for their performance, extra interruptions further increase the overload, and performance is impaired (Altmann and Trafton, 2002).

3. HYPOTHESES

Prior research has examined performance differences considering objective task difficulty (Payne *et al.*, 2007; Speier *et al.*, 2003), usually in the context of a single multitasking scenario. In discretionary switching, Payne *et al.* (2007) found that time allocation is sensitive to the level of difficulty of each task as participants seek to optimize performance. In an interruption scenario, Speier *et al.*'s (2003) comparative study of simple and complex tasks found that interruptions during a task helped performance with simple tasks, but hurt the execution of more complicated ones.

Complex tasks require more cognitive effort than easier tasks and task performance is impacted. However, the same task can be difficult for one person and easier for another. Maynard and Hakel (1997) indicate that task performance depends not only on objective task complexity but on subjective perceptions of task difficulty as well. As a result, we propose that depending on the subjective difficulty of the task at hand, performance will be affected differently as a result of discretionary or mandatory interruptions. Furthermore, performance will be impacted for both the primary task and the interrupting tasks.

Deciding to take a break or being forced to take a break can affect a user's performance in different ways. The negative effects of mandatory interruptions are due to the cognitive costs associated with switching between ongoing tasks at times that are beyond the control of the user (Altmann and Trafton, 2002). When people multitask at their discretion, they can decide when and how often to switch among ongoing tasks. During a complex task, receiving unplanned interruptions can impact performance more than planning an interruption at a suitable breaking point. Therefore, we propose:

> H1a. During a task they consider harder, those who are forced to multitask will perform worse on all tasks than those who do not multitask.

> H1b. During a task they consider harder, those who are forced to multitask will perform worse on all tasks than those who multitask at their discretion.

In contrast, during an easier task, less cognitive resources are used. Based on the Yerkes–Dodson law, easy tasks are associated with low arousal levels. Any increase in the level of arousal can improve performance. For example, an unexpected interruption will raise the levels of arousal, and performance will improve.

Performance of those who are forced to multitask will also be different from performance of those who choose to multitask at their discretion. When multitasking at one's discretion a user may choose to switch tasks after a sub-goal has been completed (Payne *et al.*, 2007), depending upon their priorities (Janssen and Brumby, 2010; Janssen *et al.*, 2012). Because the user can decide when to switch and it is not unexpected, these self-interruptions are known and do not increase arousal. Given that receiving unexpected interruptions can provide greater stimulation, those forced to multitask can improve their performance more than those who multitask at their discretion. Based on these arguments, we formulate the following hypotheses:

> H2a. During a task they consider easier, those who are forced to multitask will perform better in all tasks than those who do not multitask.

> H2b. During a task they consider easier, those who are forced to multitask will perform better in all tasks than those who multitask at their discretion.

4. MATERIAL AND METHODS

4.1. Participants

Six hundred and thirty-six subjects (334 male and 302 female) were recruited from a large urban college in the Northeast USA. About half (307) received $10 monetary compensation and the other half (329) received course credit. Subjects performed the computer-based experiment in a laboratory setting. Participants were equally distributed in the three conditions (212 subjects in each).

4.2. Design

We developed an experimental multitasking environment in Microsoft Visual C++. In this environment, we conducted a controlled experiment where participants had to perform a primary task and a set of secondary tasks. Participants were randomly assigned into one of three multitasking conditions: *discretionary, mandatory* and *sequential*.

(i) *Discretionary*: In the discretionary condition, all tasks were presented at once, in different tabs and subjects were able to choose when to complete each task. Subjects in this condition were allowed to switch tasks at any point. The interface kept track of when subjects were switching and how often.

(ii) *Mandatory*: In the mandatory condition, the secondary tasks appeared while subjects were in the middle of completing the primary task. In this condition, subjects were interrupted at different intervals of time with pop-up windows that forced them to complete other tasks. The interrupting task had to be completed before the user was able to resume the primary task. In this instance, one of the visual exercises covers the screen and subjects have to answer as many answers as they can before time for the time for this task expires. Once the time limit was reached, subjects were brought back to the primary task screen.

(iii) *Sequential*: In the sequential condition, the secondary tasks were displayed as pop-up windows only after the primary task was completed (i.e. the total allotted time on task had elapsed).

4.3. Tasks

The experimental environment presented six game-like tasks for participants in all three conditions. The *primary task* was a Sudoku puzzle.[1] The goal of a Sudoku puzzle is to fill in all the boxes in a 9 × 9 grid, so that each column, row and 3 × 3

---

[1] The Sudoku puzzle was taken from http://puzzles.about.com/od/sudokupuzzles/qt/Sudoku-Puzzles-To-Print.htm. This site provides easy, medium and hard Sudoku puzzles. The selected puzzle was taken from the easy puzzles (Easy Sudoku Puzzle #07).

box have the numbers 1– 9 without any of those numbers being repeated.

There were five *secondary tasks*: one textual task, two visual tasks and two number series tasks. The *textual* task consisted of unscrambling a series of letters to find up to 20 words. The *visual* tasks required subjects to select the shape that best fit the pattern. Subjects were shown four shapes and had to choose the shape that did not belong. There were ten visual multiple-choice problems and there were two of these visual tasks (i.e. two sets of ten visual exercises).[2] The *Number Series* tasks involved subjects guessing the missing number in the series of numbers presented. Subjects had two number series exercises to complete, each with ten questions.[3]

Sudoku was chosen as the primary task as it requires more time and concentration to complete than the secondary tasks. In addition, when subjects are performing other tasks and return to the primary task, they need to remember their thought process. While multitasking may not be as disruptive when dealing with multiple tasks on different modalities, such as one auditory task and a separate visual task, having multiple tasks in the same modality is more disruptive (Wickens, 2002). Although the chosen tasks were unique in that they used different skills (visual, textual or numeric), they all required significant cognitive resources for their successful performance.

The goal was to implement tasks that required different skills and durations in order to emulate an actual computer usage session. Generally, users work on a primary task, which requires more time and concentration. They might be interrupted by an instant message alert, which will not require as much time or thought to respond to. Or, perhaps they receive an email message that requires a little more time than the IM alert, but less than their primary task. Our experiment tried to mimic this by providing different types of tasks with different durations.

The time to complete each task was limited and determined based on prior pilot testing. The time for each task was intentionally shorter than the time subjects needed to complete the task in order to avoid subjects being idle. For the primary task (Sudoku), the maximum time limit was set to 18 min. For the secondary tasks, the time for the word task was set to 1.5 min, while the time for the two visual and two numeric series tasks was set to 48 s. Since time allotted for each task was the same for every subject, we were able to compare the performance results across all the three different conditions, ruling out the potential influence of time on task.

4.4. Procedure

Upon arrival to the lab, each subject was randomly assigned to one of the three conditions and given specific instructions according to their condition. After signing a consent form, subjects started to use the multitasking environment. They were presented with a pre-test questionnaire, which included demographic questions (i.e. age and gender) as well as questions about usage of and comfort with a computer, and prior experience with Sudoku. After this questionnaire session, participants had a practice round of Sudoku to familiarize themselves with the Sudoku as well as the interface for this task. Next, the system presented a reminder of the game instructions and the tasks were presented according to the condition (one at a time in sequential, all at once in discretionary or through interruptions in the mandatory condition). Once the time for all the tasks expired, subjects were brought to the post-test questionnaire. The results of the tasks and questionnaires were automatically written to a unique log file generated for each participant.

4.5. Measures

Sudoku performance was calculated as the number of correct answers as a percentage of the total answers required. Specifically, in the Sudoku task there were 49 empty spaces that needed to be filled out with the appropriate numbers. The score was the number of correct values entered divided by the total number of squares that had to be filled during the session (49).

Secondary task performance was computed by averaging the performance scores of all five secondary tasks. For the word task, there were 20 acceptable words that could be generated from unscrambling the letters. The percent correct is the number of correct responses out of 20. The same method was applied to calculate the visual and number series tasks' scores.

Overall performance was calculated as the average of all six tasks (Sudoku and secondary tasks).

Subjective task difficulty: While the Sudoku puzzle chosen was from an online selection of puzzles in the easy category, not all subjects may find it easy. Therefore, we measured subjective task difficulty in the post-test questionnaire by asking the subjects to rate the level of difficulty of the primary task (from$1 =$ easy to$5 =$ hard).

5. RESULTS

Table 1 presents the basic statistics of the demographic and skills characteristics (age, computer skills and Sudoku experience) of the sample.

To ensure that randomization worked and to rule out alternative explanations, the demographic characteristics of participants were first checked for possible pre-existing differences among conditions. None of the continuous pre-test

---

[2] Taken or modified from http://www.mathworksheetscenter.com/mathskills/shapes/shapenotbelong/shapenotbelongK2begles.pdf, http://www.intelligencetest.com/questions/visualization.htm, http://www.iqtestexperts.com/visual-sample.php, http://www.didax.com/newsletter/pdfs/mental_math_215297.pdf and http://www.syvum.com/iq/.

[3] Number Series task problems were taken from http://www.funbrain.com/cracker/index.html.

Table 1. Descriptive statistics ($n = 636$).

Pre-test variables	Mean	SD	Min	Max
Age[a]	22.44	4.69	18	54
Computer skills[b]	3.70	0.80	1	5
Sudoku experience[c]	1.55	1.43	0	5

[a] One subject typed an invalid age and was omitted from this analysis.

[b] Computer skills was measured with a five-point scale from 1 = poor to 5 = excellent.

[c] Sudoku experience was measured with a 0-to-5 scale similar to computer skills, but with a 0 for those who had never played Sudoku before.

questionnaire variables showed a systematic variation. Separate ANOVAs were performed using age ($\text{Mean}_{\text{Discretionary}} = 21.98$; $\text{Mean}_{\text{Sequential}} = 22.53$; $\text{Mean}_{\text{Mandatory}} = 22.80$; $F(2, 632) = 1.71$ ns), computer skills ($\text{Mean}_{\text{Discretionary}} = 3.71$; $\text{Mean}_{\text{Sequential}} = 3.70$; $\text{Mean}_{\text{Mandatory}} = 3.70$; $F(2, 633) = 0.02$ ns) and Sudoku experience ($\text{Mean}_{\text{Discretionary}} = 1.54$; $\text{Mean}_{\text{Sequential}} = 1.57$; $\text{Mean}_{\text{Mandatory}} = 1.53$; $F(2, 633) = 0.04$ ns) as dependent variables. A separate χ^2 analysis was conducted for gender. The results showed that male and female participants were equally distributed across conditions ($\chi^2 = 1.68$; $P = 0.43$ ns). The demographic variables (age, gender, computer skills and Sudoku experience) were similarly distributed across conditions. In the subsequent statistical analyses, the previous experience with Sudoku will be used as a control given its potential effect to explain differences in Sudoku performance.

To examine whether the experimental condition had any influence on the subjective task difficulty, we compared subjective task difficulty across conditions and found no variation. Thus, the subjects' subjective task difficulty is independent of the multitasking condition to which they were assigned. In particular, Sudoku level of difficulty (mean $= 3.22$; 1.38 SD and median $= 3$), was not significantly different across conditions ($F(2, 633) = 0.54$ ns).

5.1. Test of hypotheses

In order to formally test our hypotheses, we examined whether there was an interaction effect between subjective task difficulty and experimental condition. To perform the analyses, we divided the sample into three categories: those who found the primary task (Sudoku) to be difficult (i.e. rated its difficulty with 1 or 2), those who found it easier (i.e. gave a rating of 4 or 5) and those who were neutral (i.e. chose the medium level 3). Three models were run, one for each dependent variable (overall performance, Sudoku performance and secondary task performance).

The results of each model are reported in Table 2. The name of the corresponding dependent variable is listed at the top of the table. The top portion shows the means of the nine conditions (3 modes, 3 levels of difficulty). The bottom portion of the table shows the F for the entire model and the corresponding percentage of variance explained (R^2). The main effects for the explanatory variables are listed below, indicating in each case the F-statistic, its significance and the eta square (η^2) to indicate the strength of the association or effect size.

As shown in Table 2, the results of the interaction effect between multitasking condition and subjective Sudoku difficulty are significant for the performance of both the primary task (Sudoku) and the secondary tasks, as well as the overall performance. In all cases, the interaction term was significant at $P < 0.05$, and the η^2 is of the order of 1%. These interaction effects are significant after controlling for the prior level of Sudoku experience, which plays a prominent role in explaining performance outcomes in this study. The scores indicate that the mandatory multitasking condition has the highest primary and secondary task performance when the primary task was considered easier, and the lowest performance when the task was considered more difficult.

To further analyze these results, we ran separate *post hoc* contrasts to test for significant differences among the groups. The results showed that overall performance in the mandatory condition when subjective task difficulty was easy was significantly higher than the performance results in the other groups ($F(9, 626) = 6.51$, $P = 0.0110$). In contrast, performance in the mandatory condition when the subjective task difficulty was hard was significantly lower than the performance results in the other groups ($F(9, 626) = 6.54$, $P < 0.0108$).

While it is intuitive that those who found Sudoku difficult have a negative correlation with performance in the primary task, in the mandatory multitasking condition, subjective task difficulty also negatively correlated with performance in the secondary tasks as well ($\rho = 0.18673$; $P = 0.0064$). This suggests that in difficult tasks, receiving interruptions can impact both primary and secondary tasks.

Figure 1 depicts the performance in the three conditions for the different difficulty levels. The medium difficulty level shows the most variability across multitasking modes. However, the results in the extreme categories clearly indicate that when the task is deemed easy, the best performance is achieved in the mandatory multitasking condition. In contrast, when the task is deemed difficult, the worst performance is found in the mandatory condition.

6. DISCUSSION

Our results support the proposed hypotheses. Those in the mandatory multitasking condition, who found the primary task to be difficult, performed worse than those in both the sequential and discretionary conditions. In addition, those in the mandatory condition, who found the primary task to be easy, performed

Table 2. Comparison of performance by multitasking condition and task difficulty.

Multitasking condition	Subjective Sudoku difficulty	Overall performance	Sudoku performance	Secondary task performance
Mandatory	Easy	44.44	77.65	37.80
Mandatory	Medium	38.23	69.54	31.96
Mandatory	Hard	34.59	46.15	32.28
Sequential	Easy	42.90	77.09	36.06
Sequential	Medium	43.52	74.63	37.29
Sequential	Hard	36.47	49.43	33.88
Discretionary	Easy	43.69	75.04	37.42
Discretionary	Medium	42.01	80.48	34.32
Discretionary	Hard	36.44	49.39	33.85

	Overall performance		Sudoku performance		Secondary task performance	
Model $F(9, 626)$	31.37***		57.49***		12.50***	
R^2	31.08%		45.25%		15.24%	
Explanatory variables	F(sig)	η^2	F(sig)	η^2	F(sig)	η^2
Condition	1.66 ns	0.004	1.65 ns	0.003	1.13 ns	0.003
Subjective Sudoku difficulty	3.49*	0.008	29.80***	0.052	1.77 ns	0.005
Condition × subjective Sudoku difficulty	3.34*	0.015	3.14*	0.011	2.51*	0.014
Sudoku experience	203.21***	0.224	284.87***	0.249	89.75***	0.122

Significance levels: $^{*}P < 0.05$; $^{***}P < 0.0001$.

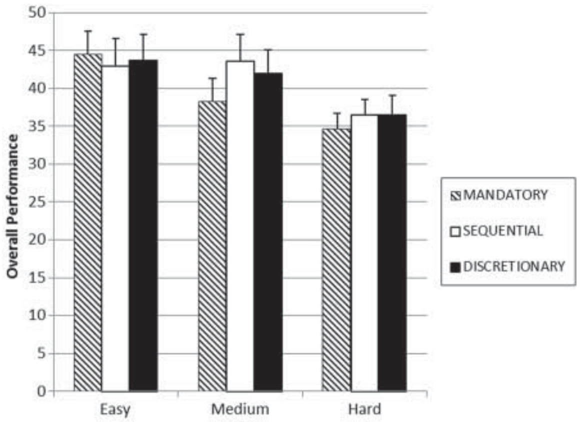

Figure 1. Mandatory, sequential and discretionary performance with regard to subjective task difficulty (error bars = 95% confidence intervals).

better than those in both the sequential and discretionary conditions. The interaction effect between interruption and subjective difficulty can be explained with the Yerkes–Dodson law (Yerkes and Dodson, 1908). Receiving interruptions during a hard task can cause too much arousal and an overload in mental workload and negatively affect performance (Altmann and Trafton, 2002). Remarkably, external disruptions have the opposite effect for easier tasks, where getting interrupted can actually help performance by increasing the amount of stimulation (Speier *et al.*, 2003; Teigen, 1994). This effect was found when comparing the mandatory condition to both the sequential and discretionary conditions.

Although we selected an easy Sudoku problem for all subjects, it did not present the same level of difficulty for all of them. The subjective level of difficulty was not affected by the multitasking condition and probably reflects some aspects of performance felt by the subjects. As such, it is a good measure of the cognitive challenges experienced by the subjects while they were solving the primary task. Based on the self-reported level of difficulty, we were able to analyze the interaction between multitasking and subjective task difficulty on performance. Owing to the stronger influence of prior Sudoku experience in performance, the effect size of the interaction term is comparatively small in this study. The smaller magnitude of the interaction effect does not diminish its importance as a novel contribution to literature. As Shadish *et al.* (2002) indicate, interactions can be difficult to detect, so large sample sizes are necessary when focusing on an interaction. The interaction effect could be much more prominent in other settings with a different choice of tasks. This is a worthwhile avenue of investigation for future research efforts.

6.1. Limitations

As with any other controlled experiment, there are some limitations that must be acknowledged. First, the subjects were college students, and the experiment was conducted in a laboratory environment. Thus, the results may not be applicable to different types of individuals or other settings. In non-experimental laboratory settings, multiple tasks might be competing for the person's attention. In these type of environments, there is some freedom to select which tasks to carry out, but the performance outcomes are not always easy to evaluate. Real-life tasks include decision-making, planning and generating new ideas and may not correspond to problem-solving (puzzle-like) exercises with demonstrable correct solutions. Nevertheless, having these tasks with different durations, objectives and demonstrable correct answers provided the means to objectively calculate performance scores for all subjects.

Sudoku experience is a significant covariate in the statistical analyses. Subjects who had more experience with Sudoku before had a better performance than subjects who had less experience. One way to minimize this effect would be to use a different primary task, where there is a more equal skill set, i.e. a task more subjects are familiar with or a completely new task that fewer subjects are familiar with. While the practice round tried to minimize some of the differences between those who never played and those who had played Sudoku before, one cannot really be experienced in playing Sudoku until one has played it at least a few times.

Finally, the laboratory experiment only created computer-based interruptions. The design of this experiment did not take into account interruptions that occur outside the computer, such as the ringing of a phone or a person approaching to chat. Other common computer interruptions such as notifications of incoming emails were also excluded. The effects of alerts announcing new email messages may be different from being interrupted by (or switching to) a problem-solving task, where correct answers need to be produced. Real-life interruptions can also occur with more frequency and, at many times, are related to each other. For example, having a back-and-forth instant message conversation while working on a project would be very different from the types of interruptions included in this experiment.

6.2. Implications for research

Although the limitations of this experiment suggest caution when generalizing the results to other populations and settings, the results provide the basis for analyzing the performance effects of different patterns of multitasking behavior. By intentionally using tasks that can be scored and time limits, the results are comparable for subjects working under different conditions. The use of a laboratory experiment allowed us to control for the variables of interest (tasks and times) and test the hypotheses in the absence of extraneous influences. In addition, having the experiment solely on a computer in a controlled environment provided an effective way of automatically recording the subjects' activities in a log file. This allowed an accurate analysis to be performed, which would not have been possible if there were other types of external, uncontrolled or non-computer-based, interruptions.

This study contributes to the multitasking literature by bringing together two different streams of research (interruption-driven and discretionary multitasking) and systematically comparing its effects on performance in a controlled environment in conjunction with subjective task difficulty. This comparison is completed by including a sequential (mono-tasking) condition. From a research standpoint, there are ample opportunities to further compare these three conditions in future studies by including different types of tasks.

A distinctive characteristic of this study is its focus on subjective task difficulty, as opposed to objective task difficulty, as one of the factors that influence performance. Subjective task difficulty gives an estimation of cognitive load experienced by participants and provides theoretical explanations for the differential effects on performance (Maynard and Hakel, 1997). Follow-up studies could systematically compare tasks with different degrees of objective difficulty.

6.3. Implications for interface designs

In modern personal computer environments, technological advances in computing power and notification capabilities are causing people to engage in multitasking. This brings up the challenge of creating better interfaces that make it easier for users to switch back and forth between projects (Czerwinski *et al.*, 2004). Multitasking platforms must be flexible enough to accommodate different ways of handling interruptions (McFarlane and Latorella, 2002).

Based on our results, receiving interruptions during high cognitive load conditions resulting from a task that is perceived difficult can hurt performance. One way to prevent this situation is to design systems that allow subjects to indicate when they are performing complicated tasks and how to detect low cognitive load conditions. This information would help in determining the moments when interruptions are less disruptive. One such example is Iqbal and Bailey's (2010) Oasis system, which interrupts users during breakpoints.

7. CONCLUSIONS

This study compared the performance of three different conditions: discretionary (those allowed to choose whether and/or when to multitask), mandatory (those forced to multitask at allocated times designated by the system), and sequential (those who did not multitask) based on the subjects' subjective task difficultly. The results show that participants in the mandatory multitasking condition had the lowest performance. Thus, receiving interruptions during a difficult task has more detrimental effects than carrying out the task under discretionary multitasking or no multitasking at all. In the mandatory condition, subjects who considered the primary task easy had the best performance. Taken together, the results suggest that receiving interruptions during a hard task can lead to negative consequences, but positive consequences during easier tasks.

ACKNOWLEDGEMENTS

This work was part of the dissertation of the first author and the second author served as dissertation advisor. We are grateful to the members of the dissertation committee, Dr Matt Huenerfauth, Dr Richard Holowczak, and Dr Mark Silver for their insightful feedback in the developing stages of this research.

FUNDING

This work was partially supported by PSC-CUNY Research Grant # 62552-00 40 for subject compensation. In addition, a CUNY Graduate Center Doctoral Student Research Grant was used to compensate subjects during our pilot study.

REFERENCES

Adler, R.F. and Benbunan-Fich, R. (2012) Juggling on a high wire: multitasking effects on performance. Int. J. Hum.-Comput. St., 70, 156–168.

Altmann, E.M. and Trafton, J.G. (2002) Memory for goals: an activation-based model. Cogn. Sci., 26, 39–83.

Bailey, B.P. and Iqbal, S.T. (2008) Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. ACM Trans. Comput.-Hum. Interact., 14, 1–28.

Bailey, B.P. and Konstan, J.A. (2006) On the need for attention-aware systems: measuring effects of interruption on task performance, error rate and affective state. J. Comput. Hum. Behav., 22, 658–708.

Bellotti, V., Ducheneaut, N., Howard, M., Smith, I. and Grinter, R.E. (2005) Quality versus quantity: e-mail-centric task management and its relation with overload. Hum.-Comput. Interact., 20, 89–138.

Benbunan-Fich, R., Adler, R.F. and Mavlanova, T. (2011) Measuring multitasking behavior with activity-based metrics. ACM Trans. Comput.-Hum. Interact., 18, 1–22.

Bogunovich, P. and Salvucci, D. (2011) The Effects of Time Constraints on User Behavior for Deferrable Interruptions. In: Proc the SIGCHI Conf Human Factors in Computing Systems (CHI 2011), Vancouver, BC, Canada.

Cades, D.M., Davis, D.A.B., Trafton, J.G. and Monk, C.A. (2007) Does the Difficulty of an Interruption Affect Our Ability to Resume? In: Proc. the Human Factors and Ergonomics Society Annual Meeting, Baltimore, Maryland.

Cades, D.M., Werner, N., Boehm-Davis, D.A., Trafton, J.G. and Monk, C.A. (2008) Dealing with Interruptions Can Be Complex, but Does Interruption Complexity Matter: A Mental Resources Approach to Quantifying Disruptions. In: Proc. the Human Factors and Ergonomics Society Annual Meeting, New York, NY.

Cades, D.M., McKnight, P.E., Kidd, D.G., King, E.B. and Boehm-Davis, D.A. (2010) Factors Affecting Interrupted Task Performance: Effects of Adaptability, Impulsivity and Intelligence. In: Proc. the Human Factors and Ergonomics Society Annual Meeting, San Francisco, California.

Crook, C. and Barrowcliff, D. (2001) Ubiquitous computing on campus: patterns of engagement by university students. Int. J. Hum.-Comput. Int., 13, 245–256.

Czerwinski, M., Horvitz, E. and Wilhite, S. (2004) A Diary Study of Task Switching and Interruptions. In: Proc. the SIGCHI Conf Human Factors in Computing Systems (CHI'04), Vienna, Austria.

Davidson, C.N. (2011) Now You See It: How the Brain Science of Attention Will Transform the Way We Live, Work, and Learn. New York, NY: Penguin Group.

Duggan, G.B., Johnson, H. and Sørli, P. (2013) Interleaving tasks to improve performance: users maximise the marginal rate of return. Int. J. Hum.-Comput. St., 71, 533–550.

Gillie, T. and Broadbent, D. (1989) What makes interruptions disruptive? a study of length, similarity, and complexity. Psychol. Res., 50, 243–250.

Gonzalez, V. M. and Mark, G. (2004) 'Constant, Constant, Multi-Tasking Craziness': Managing Multiple Working Spheres. In: Proc the SIGCHI Conf Human Factors in Computing Systems (CHI '04), Vienna, Austria.

Hembrooke, H. and Gay, G. (2003) The laptop and the lecture: the effects of multitasking in learning environments. J. Comput. High. Educ., 15, 46–64.

Hodgetts, H.M. and Jones, D.M. (2006) Interruption of the tower of London task: support for a goal-activation approach. J. Exp. Psychol. Gen., 135, 103–115.

Iqbal, S.T. and Bailey, B.P. (2010) Oasis: A framework for linking notification delivery to the perceptual structure of goal-directed tasks. ACM Trans. Comput.-Hum. Interact., 17, 1–28.

Iqbal, S.T. and Horvitz, E. (2007) Disruption and Recovery of Computing Tasks: Field Study, Analysis, and Directions. In: Proc. the SIGCHI Conf Human Factors in Computing Systems, San Jose, California, USA.

Janssen, C.P. and Brumby, D.P. (2010). Strategic adaptation to performance objectives in a dual-task setting. Cognitive Sci., 34, 1548–1560.

Janssen, C.P., Brumby, D.P., Dowell, J., Chater, N. and Howes, A. (2011) Identifying optimum performance trade-offs using a cognitively bounded rational analysis model of discretionary task interleaving. Top. Cogn. Sci., 3, 123–139.

Janssen, C.P., Brumby, D.P. and Garnett, R. (2012) Natural break points: The influence of priorities, and cognitive and motor cues on dual-task interleaving. J. Cogn. Eng. Decis. Making, 6, 5–29.

Jin, J. and Dabbish, L.A. (2009) Self-Interruption on the Computer: A Typology of Discretionary Task Interleaving. In: Proc. the SIGCHI Conf Human Factors in Computing Systems (CHI '09), Boston, MA, USA.

Mark, G., Gonzalez, V.M. and Harris, J. (2005) No Task Left Behind?: Examining The Nature of Fragmented Work. In: Proc the SIGCHI Conf Human Factors in Computing Systems (CHI '05), Portland, Oregon, USA.

Maynard, D.C. and Hakel, M.D. (1997). Effects of objective and subjective task complexity on performance. Hum. Perform., 10, 303.

McCrickard, D.S., Catrambone, R., Chewar, C.M. and Stasko, J.T. (2003a) Establishing tradeoffs that leverage attention for utility: empirically evaluating information display in notification systems. Int. J. Hum.-Comput. St., 58, 547–582.

McCrickard, D.S., Chewar, C.M., Somervell, J.P. and Ndiwalana, A. (2003b) A model for notification systems evaluation-assessing user goals for multitasking activity. ACM Trans. Comput.-Hum. Interact., 10, 312–338.

McCrickard, D.S., Czerwinski, M. and Bartram, L. (2003c) Introduction: design and evaluation of notification user interfaces. Int. J. Hum.-Comput. St., 58, 509–514.

McFarlane, D.C. (2002) Comparison of four primary methods for coordinating the interruption of people in human–computer interaction. Hum.-Comput. Interact., 17, 63–139.

McFarlane, D.C. and Latorella, K.A. (2002) The scope and importance of human interruption in human–computer interaction design. Hum.-Comput. Interact., 17, 1–61.

Meyer, D.E. and Kieras, D.E. (1997) A computational theory of executive cognitive processes and multiple-task performance: part I. basic mechanisms. Psychol. Rev., 104, 3–65.

Miyata, Y. and Norman, D.A. (1986) Psychological issues in support of multiple activities. In Norman, D.A. and Draper, S.W. (eds) User Centered System Design, pp. 265–284. Hillsdale, NJ, USA: Lawrence Erlbaum.

Oswald, F.L., Hambrick, D.Z. and Jones, L.A. (2007) Keeping all the Plates Spinning: Understanding and Predicting Multitasking Performance. In Jonassen, D.H. (ed.), Learning to Solve Complex Scientific Problems. New York: Lawrence Erlbaum Associates.

Oulasvirta, A. and Saariluoma, P. (2004) Long-term working memory and interrupting messages in human–computer interaction. Behav. Inform. Technol., 23, 53–64.

Oulasvirta, A. and Saariluoma, P. (2006) Surviving task interruptions: investigating the implications of long-term working memory theory. Int. J. Hum.-Comput. St., 64, 941–961.

Palladino, L.J. (2007) Find Your Focus Zone: An Effective New Plan to Defeat Distraction and Overload. Free Press, New York.

Payne, S.J., Duggan, G.B. and Neth, H. (2007) Discretionary task interleaving: heuristics for time allocation in cognitive foraging. J. Exp. Psychol. Gen., 136, 370–388.

Rubinstein, J.S., Meyer, D.E. and Evans, J.E. (2001) Executive control of cognitive processes in task switching. J. Exp. Psychol. Human., 27, 763–797.

Salvucci, D.D. and Bogunovich, P. (2010) Multitasking and Monotasking: The Effects of Mental Workload on Deferred Task Interruptions. In: Proc. the SIGCHI Conf Human Factors in Computing Systems (CHI 2010), Atlanta, Georgia, USA.

Salvucci, D.D. and Taatgen, N.A. (2011). The Multitasking Mind. Oxford University Press, New York.

Shadish, W., Cook, T. and Campbell, D. T. (2002) Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton-Mifflin, Boston, MA.

Speier, C., Vessey, I. and Valacich, J.S. (2003) The effects of interruptions, task complexity, and information presentation on computer-supported decision-making performance. Decision Sci., 34, 771–797.

Teigen, K.H. (1994) Yerkes–Dodson: a law for all seasons. Theor. Psychol., 4, 525–547.

Trafton, J.G., Altmann, E.M., Brock, D.P. and Mintz, F.E. (2003) Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. Int. J. Hum.-Comput. St., 58, 583–603.

Wickens, C.D. (2002) Multiple resources and performance prediction. Theor. Issues Ergonomics Sci., 3, 159–177.

Yerkes, R.M. and Dodson, J.D. (1908) The relation of strength of stimulus to rapidity of habit-formation. J. Comp. Neurol. Psychol., 18, 459–482.

Yeung, N. (2010) Bottom-up influences on voluntary task switching: the elusive homunculus escapes. J. Exp. Psychol. Learn., 36, 348–362.