DECISION SCIENCES INSTITUTE

Smart Multitasking with Smartphones

Deena Rubin
FactSet Research Systems, Inc.
E-mail: deenagrubin@gmail.com

Rachel F. Adler
Northeastern Illinois University
E-mail: r-adler@neiu.edu

ABSTRACT

Research has shown that people use smartphones in a way that encourages constant multitasking. While previous studies have determined that multitasking with technology has a negative impact on performance of a primary task, we examine how constant task switching specifically with mobile devices affects users. Our experiment examines discretionary task interleaving with smartphones and the effects this has on performance. The results demonstrate that while any amount of multitasking negatively affects performance on a primary task, these effects are lessened when switching at a breakpoint between subtasks rather than during a subtask.

KEYWORDS: Multitasking; Smartphones; Breakpoints

INTRODUCTION

Multitasking can be categorized in two different ways: internal versus external. External interruptions cause someone to switch tasks due to an interruption in the environment (Miyata & Norman, 1986), such as receiving a text message. In contrast, internal interruptions, occur when someone is interrupted due to their own thought processes (Miyata & Norman, 1986), for example someone remembers to respond to an outstanding email. An internal interruption is "a self-initiated switch away from a task prior to its completion" (Jin & Dabbish, 2009). These are also referred to as "voluntary, discretionary task interleaving between independent tasks" (Payne, Duggan, & Neth, 2007). External interruptions and internal discretionary task switching occur at about the same frequency (Gonzalez & Mark, 2004).

People are constantly multitasking whether at work or at school. Employees switch tasks often, and experience internal and external interruptions at about the same rate (Czerwinski, Horvitz, & Wilhite, 2004; Gonzalez & Mark, 2004; Dabbish, Mark, & González, 2011). Dabbish et al. (2011) found that in the workplace the organizational environment, individual differences, and external interruptions explain self-interruptions. Studies focusing on multitasking with technology have also largely focused on students. Surveys of college students have found that multitasking with technology including emailing, texting, and social networking is widespread during classroom and study time (Jacobsen & Forste, 2011; Junco, 2012; Junco & Cotten, 2012). These results have also been supported through direct observation. Despite the presence of an observer, middle school, high school and college students were found to remain on task for an average of ten minutes total out of fifteen, and were on task without a break for an average of just 5.61 minutes at a time (Rosen, Carrier, & Cheever, 2013).

Recent studies have shown that the timing of external interruptions may have different effects on performance of the primary task depending on when the interruption occurs (Adamczyk & Bailey, 2004; Iqbal & Bailey, 2005; 2007; 2008). Some studies have tried to mitigate the negative impacts of interruptions by placing them at breakpoints (Iqbal & Bailey, 2008; Okoshi, Ramos, Nozaki, Nakazawa, Dey, & Tokuda, 2015a).

Although there has been recent research examining the effects and prevalence of multitasking with technology in general, fewer studies have specifically studied multitasking with mobile devices. However, people are using these devices and they are being used in a way that encourages multitasking behaviors (Oulasvirta, Tamminen, Roto, & Kuorelahti, 2005; Böhmer, Hecht, Schöning, Krüger, & Bauer, 2011; Oulasvirta, Rattenbury, Ma, & Raita, 2012). Multitasking on mobile devices has received some research attention in recent years (Ho & Intille, 2005; Okoshi, Ramos, Nozaki, Nakazawa, Dey, & Tokuda, 2015b; Okoshi et al., 2015a). Users download numerous apps on their smartphones and some are constantly interrupting users (Ho & Intille, 2005). Performance effects of multitasking on mobile devices may not be the same as computer-based multitasking. Adepu and Adler (2016) found that users performed the same task better on desktop computers than on smartphones, though they preferred their smartphones for that task. Larger screen sizes can enhance performance.

In addition to receiving interruptions which force a user to multitask, often people choose to multitask from one app to another. However, research in the area of discretionary multitasking has received less attention. Similarly, while computer-based multitasking has been studied, less research has focused on mobile-based multitasking and its performance implications. This research examines the timing and performance effects of discretionary task interleaving on smartphones.

LITERATURE REVIEW

Performance Effects of Multitasking

Hembrooke and Gay (2003) found that multitasking with laptops during a classroom lecture negatively impacted performance. A later study expanded on that concept and examined how multitasking with different technologies affected an assessment of a series of three typical lectures (Wood, Zivcakova, Gentile, Archer, De Pasquale, & Nosko, 2012). This study found that social based technologies, such as Facebook and instant message, had the most negative impact.

The negative effects of multitasking with some technologies may fluctuate with their popularity. For example, one study found instant messaging to be a popular form of multitasking which led to negative effects (Levine, Waite, & Bowman, 2007). Later studies performed after instant messaging was no longer popular did not find the same negative effects, but found multitasking with other forms of technology that were used more often, and for social purposes to be harmful. For example, surveys of college students have found that multitasking with Facebook or texting is negatively correlated with GPA (Jacobsen & Forste, 2011; Junco, 2012; Junco & Cotten, 2012). Similarly, another study found that students who texted or accessed Facebook at least once during the fifteen minutes they were observed were more likely to have a low GPA (Rosen et al., 2013).

Bowman et al. (2010) did not find that multitasking with technology negatively impacted the performance of a primary task. However, they did find that multitasking increased the length of time to complete the task. These findings were contradicted by Adler and Benbunan-Fich (2012). Their study found that medium levels of multitasking helped participants finish a larger amount of the primary task, but that any amount of multitasking negatively impacted accuracy on the primary task. This may be due to the fact that in the study by Adler and Benbunan-Fich (2012) the primary task was timed, while the experiment performed by Bowman et al. (2010) was untimed.

A person's perception of task importance can also affect performance when multitasking (Coens, Reynvoet, & Clarebout, 2011). Participants in this study who were told that the secondary task was more important performed better on the secondary task than on the primary task. Participants who were told that both tasks were equally important performed better on the secondary task than participants who were told that the primary task was more important. Janssen et al. (2012) also found that people perform differently depending upon the priority of the tasks.

When Do People Multitask and its Performance Implications

Many studies that have examined when people multitask have not focused on discretionary multitasking, which involves task switches prompted by the user, but rather they focus on external interruptions (Dabbish et al., 2011). However, some experiments involving external interruptions can help predict when people are most likely to multitask when given a choice.

McFarlane (2002) compared four methods of interruption: immediate, negotiated, mediated, and scheduled. Participants in the immediate category received interruptions immediately when they occurred. Participants in the negotiated category were given control over when they would respond to the interruptions. Participants in the mediated category were interrupted at times of lower workload as judged by the application. Participants in the scheduled category received interruptions every 25 seconds. The negotiated category was the most successful category. Although the timing of the task switches was not recorded, the fact that it led to the best performance suggests that people waited for an opportune time to respond to interruptions.

Salvucci and Bogunovich (2010) determined that people tended to respond to interruptions at a moment of lower workload when given a choice. A time of lower workload is a point in the primary task when the participant would not need to remember any information about the primary task when returning after an interruption.

In one study where participants were allowed to multitask at their own discretion, people preferred to switch at subtask boundaries (Payne et al., 2007). Participants were presented with two similar tasks with hard and easy versions. People in this experiment preferred to switch tasks at the completion of a subtask.

Several experiments have been done to determine the most opportune time to interrupt people who are using technology. These studies have all demonstrated that interruptions at breakpoints are the least disruptive to the user by comparing the cost of disrupting participants at different points in a task (Adamczyk & Bailey, 2004; Iqbal & Bailey, 2005; 2008). These experiments have involved users completing a task on a computer, and being disrupted at predetermined best times (at breakpoints between subtasks), at predetermined worst times (during a subtask), and at random times.

Adamczyk and Bailey (2004) report that participants experienced more frustration and annoyance when the interruptions were at random and worst times. Participants also rated these disruptions as less respectful of their primary task than disruptions at best times. However, differences in resumption lag (the time between responding to the interruption and resumption of the primary task) were not significant. Iqbal and Bailey (2005) contradict the previous findings about resumption lag. In their experiment, resumption lag was significantly higher for random and worse cases than for best cases.

Iqbal and Bailey (2007) divide breakpoints into three categories: coarse, medium, and fine. For example, in a task involving document editing, “Fine may be switching paragraphs; medium may be switching documents; and coarse may be switching to an activity other than editing.” Iqbal and Bailey (2008) interrupted participants at coarse, medium, and fine breakpoints with interruptions related to the primary task and interruptions unrelated to the primary task. Participants reported preference for related material at fine and medium breakpoints and unrelated material at coarse breakpoints. They also reported that resumption lag differed based on the relevance of the interruption content. Furthermore, they found that scheduling notifications at breakpoints, rather than having them occur immediately, reduced frustration and reaction time.

Other research has found that the timing of an interruption can affect the length of time it takes to complete the interrupting task. Fischer et al. (2011) interrupted users using their phones in natural environments throughout a two week period and found that interrupting tasks take longer to acknowledge and complete when participants were interrupted at a random time rather than at the end of another task (such as after hanging up from a phone call). Their findings did contradict Iqbal and Bailey (2008) in that their participants did not give random interruptions a worse rating than interruptions at the end of another task.

How Mobile Devices are Used

Several studies have demonstrated that people use mobile devices in a way that is habit forming and facilitates multitasking. When navigating typical urban environments, people pay attention to their mobile devices in short bursts of four to eight seconds (Oulasvirta et al., 2005). Smartphone use throughout the day is more spread out, with one study finding participants spending an average of 59.23 minutes on their mobile devices per day and each session lasted an average of 71.56 seconds (Böhmer et al., 2011).

Smartphones are also used for a longer portion of the day than laptops. Oulasvirta et al. (2012) found that people use smartphones for an average of 160 minutes per day as opposed to 87 minutes per day on a laptop. They also found that unlike laptops, smartphone usage is habit forming. This study looked for smartphone usage sessions that qualify as Short Duration and Reward Based (SIRB). This is defined as a session that is less than 30 seconds in duration that provides some emotional reward. Smartphone sessions were found to be significantly more likely to qualify as SIRB than laptop sessions, which lead to habit forming behaviors. Oulasvirta et al. (2012) suggest that this is because smartphones are more readily available than laptops. They found several characteristics of habitual smartphone use. These habits increase the overall time spent using smartphones because they lead to other smartphone based activities.

More people are now becoming distracted with their smartphones. This can occur in many situations, such as in the classroom (Bradstreet Grinols & Rajesh, 2014). With interruptions from phones becoming normal occurrences, designers need to think about ways to manage these interruptions (Tolmie, Crabtree, Rodden, & Benford, 2008). Böhmer et al. (2014) found that a multiplex design, where people can be alerted to a call through a small partial-screen notification while working on their primary task worked best. Nagata (2003) found that anticipated interruptions led to better performance in web-based tasks, particularly on a mobile device.

Many websites that were once used solely on computers now have mobile versions or apps. Kalkbrenner and McCampbell (2011) conducted a survey comparing smartphone use vs. standard phone use and found that 88% of respondents felt that smartphones led to an increase in their productivity compared with 42.8% of regular phone owners. Having many apps on one's phone increases people's multitasking. As computers became faster, cheaper, and more powerful, user multitasking occurred more frequently. Similarly, smartphones are become more commonplace and people can open multiple apps simultaneously and task switch.

While multitasking can occur between smartphones and other media and multitasking is prevalent with smartphones when driving (Liu, Cao, Tang, He, & Wen, 2017), we concentrate exclusively on multitasking on mobile devices from one smartphone-based app to another in order to a) compare the results to computer-based multitasking and b) understand more about the ways in which people multitasking on a mobile devices and its effects on performance.

HYPOTHESES

When given a choice about when to respond to an interruption, people tend to choose a time in between subtasks when they will not need to remember information about the primary task (Salvucci & Bogunovich, 2010). People also experience frustration when forced to task switch during a subtask instead of in between subtasks (Adamczyk & Bailey, 2004). People perform better with negotiated interruptions (McFarlane, 2002). If given a choice, people will prefer to switch tasks at a time that is more convenient, such as during a breakpoint rather than during a subtask. Therefore, we hypothesize:

H1: People will have more task switches at breaks between subtasks than during subtasks on mobile devices.

Multitasking has been demonstrated in various studies to negatively impact performance on a primary task. Some studies have found a negative correlation between GPA and multitasking in class or while completing school work (Jacobsen & Forste, 2011; Junco, 2012; Junco & Cotten, 2012; Rosen et al., 2013). Other studies have found that multitasking has a direct impact on performance of the primary task (Hembrooke & Gay, 2003; Adler & Benbunan-Fich, 2012; Wood et al., 2012). Although these studies have not specifically focused on the effects of multitasking with mobile devices, other studies have shown that smartphones are used in a way that specifically encourages multitasking habits and behaviors (Oulasvirta et al., 2005; Böhmer et al., 2011; Oulasvirta et al., 2012). Therefore, we predict:

H2: People who multitask with mobile devices will have worse performance on the primary task than those who do not multitask.

As mentioned above, task switches negatively impact performance on a primary task. This impact can be affected by several factors, including the timing of the task switch. While using technology, interruptions during a subtask can be more detrimental than interruptions in between subtasks (Adamczyk & Bailey, 2004; Iqbal & Bailey, 2005; 2008). When people are interrupted at a subtask, they experience less frustration (Adamczyk & Bailey, 2004) and less time lag when returning to the primary task (Iqbal & Bailey, 2005). Similarly, with discretionary task interleaving there can be better and worse times to choose to interrupt oneself. Therefore, we propose:

H3: Task switches during a subtask on a mobile device will negatively affect performance on the primary task more than task switches in between subtasks.

MATERIALS AND METHODS

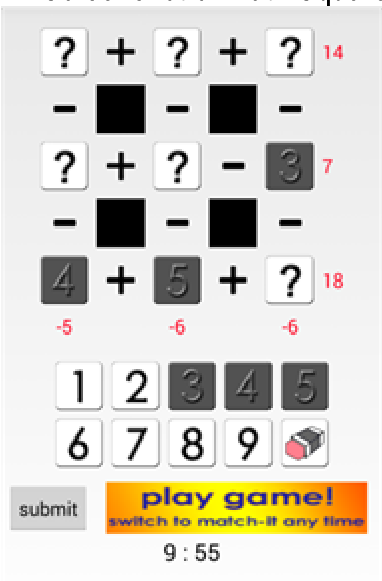
In order to examine when people task switch and its effects, we created a primary task and secondary task that participants were able to choose to switch to.

Primary Task

For the primary task we created an app-based math puzzle called Math Squares. The puzzle consisted of a grid of three rows and three columns of the numbers that formed equations using addition and subtraction. Each of the numbers one through nine was used in the grid exactly once. At the start of the puzzle, the answers to each of the equations were displayed at the end of each row and column. The goal of the task was to fill in the numbers on the grid to create accurate equations. Three numbers were displayed by default to help participants get started. See Figure 1 for a screenshot of the Math Squares puzzle.

The Math Squares puzzle was designed as the primary task for several reasons. The puzzle requires concentration to find the correct answer, but only involves basic skills like addition and subtraction. Multiple puzzles could be completed in the ten minute time limit so that participants would be able to task switch both in the middle of completing a puzzle and in between puzzles.

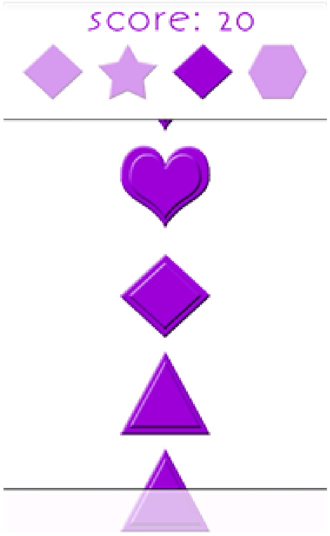
Figure 1: Screenshot of Math Squares Puzzle



Secondary Task

For the secondary task, we created an app-based game called Match-It. At the top of the phone screen, four shapes were displayed in a row. Other shapes scrolled down the middle of the screen. The goal of the game is to tap the scrolling shapes in the order of the pattern at the top of the screen. See Figure 2 for a screenshot of the Match-It game.

Figure 2: Screenshot of the Match-It game



The Math Squares puzzle was designed as the primary task since it requires concentration to find the correct answer similar to performing some kind of homework related task. The Match-It game was designed as the secondary task because of its similarity to popular games available in the Google Play store that can be played quickly. Our goal was to mimic the user who is working on something, such as homework, but switches to play a game as is common nowadays.

Participants

Participants were recruited through Amazon's Mechanical Turk, an online platform that allows task creators (known as requesters) to pay workers to complete online tasks known as Human Intelligence Tasks (HITs). Requesters are able to hire a large number of workers from around the world for relatively low pay. This allows researchers using Mechanical Turk to inexpensively recruit more participants from a diverse pool of candidates (Horton, Rand, & Zeckhauser, 2011; Berinsky, Huber, & Lenz, 2012; Mason & Suri, 2012). Research has shown that although workers are paid small amounts of money for doing short tasks, they are often intrinsically motivated and complete the HITs for personal enjoyment (Buhrmester, Kwang, & Gosling, 2011). The reliability of research using Mechanical Turk has been found to be on par with research conducted using traditional methods. In fact, Mechanical Turk participants performed better on online attention checks in three different studies when compared with an undergraduate subject pool of students (Hauser & Schwarz, 2015). Studies can also be done using a wider and more diverse pool of candidates than recruitment on college campuses (Buhrmester et al., 2011; Horton et al., 2011; Berinsky et al., 2012; Mason & Suri, 2012). For better accuracy of results, the requester can also specify qualifications for completing a HIT, such as the number of HITs a worker has previously completed, and their prior approval rating. Requesters can choose to reject a HIT and not pay the worker if the HIT was not completed accurately (Mason & Suri, 2012).

For this study, a HIT was created on Mechanical Turk. $1.00 payment was given for approximately twenty to twenty-five minutes of work on the experiment. To incentivize accurate results, the person with the highest score was given an additional reward of $5.00. Participants were given forty minutes to complete the HIT so that they would not feel rushed. In order to ensure that our participants would take the tasks seriously, workers had to have previously completed at least 500 HITS at 95% approval rate by the previous requesters. Because the directions were written in English, participation was also restricted to English speaking countries.

After accepting a consent form, participants were given a link to access the experiment via an Android phone. After installing the app and beginning, participants were given a questionnaire regarding their demographic information. Participants were then presented with tutorials for the primary and secondary tasks. Both tutorials included written instructions for how to complete the task as well as two minutes of practice time for each of the tasks.

After the tutorials, instructions were provided to explain the rules of the experiment. The directions were as follows:

1) You will be given ten minutes to complete the math puzzle as many times as you can.
2) After you submit each completed puzzle, a new puzzle will appear. You will not find out your score until the end of the experiment.
3) Your goal is to complete as many puzzles as you can accurately. The participant with the highest number of complete wins will receive an additional $5.00.
4) You may switch to the Match-It game at any time. You can play the Match-It game as often as you want.
5) Please do not use a calculator.
6) Please do not close this app, or engage in any other activities during the ten minutes. If you receive a phone call, please do not answer it. If this app is closed, you will not be able to continue with the experiment, and you will not be paid.

In the experiment itself, participants had ten minutes to complete a primary task and were given the opportunity to switch to a secondary task at any time. Once the experiment was completed, participants were given a post-questionnaire regarding their performance during the experiment.

One hundred and eighty three participants completed the experiment. Two participants were not able to complete the experiment due to technical difficulties with their phones and had to be removed. One additional participant was removed who had 53 switches. We chose to remove this participant, since the range of switches for all other participants was between 0 and 13, and 53 was way above the average and not representative of typical participants but an outlier. The data of the remaining 180 participants (108 male and 72 female) was analyzed.

Since the experiment was through Mechanical Turk and therefore remote (we could not see the participants), a few safeguards were put in place to ensure that participants were not switching to other apps on their phone and were completing the experiment with full concentration and commitment:

1) The app was designed so that if the participant attempted to navigate away from the app

on their phone, they were prompted to abandon the experiment or to continue. After abandoning the experiment, the app was made inaccessible.

2) The post-experiment questionnaire asked participants if they had used a calculator, left the experiment at any time and asked them to self-rate their effort on a scale of one to five. The directions in the questionnaire ensured participants that they would be paid regardless of their answers.

No participant reported using a calculator. Four participants (2%) reported leaving the experiment at some time, and six participants (3%) rated their effort as a one (from a scale from 1 – 5), which we felt indicated they did not put in much effort to try and complete the task with the best results possible. We considered removing these 10 participants, but removing them from the results did not change the significance of the analysis so we left them in.

We also allowed users to leave comments, many of which indicated that they enjoyed the experiment and put effort into completing it. Some comments include: “It was awesome playing the puzzle”, “I wanted to do as best I could on the math game so I could win the bonus.”, “Nice game. Hope you release it for Android.”, and “good hit..brain heating”.

Pre- and Post-Questionnaires

Participants were given a pre- and post-questionnaire on their smartphone before and after the experiment. Since the experiment as well as the questionnaires was completed on a smartphone we asked mostly closed-ended questions with choices to limit the amount of typing on the phone that would be necessary.

In the pre-questionnaire, participants were asked questions related to their frequency of game use and multitasking patterns on phones. In the post-questionnaire they were asked the questions described previously on their effort, whether they used a calculator, and left the experiment at any point.

Experiment

For ten minutes, participants played the Math Squares puzzle as many times as possible. Underneath the puzzle there was a button to submit a puzzle and receive a new one, a button allowing the participant to switch to the Match-It game, and a timer showing the remaining time. Each Math Squares session had to take a minimum of thirty seconds to keep participants from randomly filling in the grid and submitting. If the user attempted to close the app by clicking on the home or recent apps buttons, a popup would appear asking the user to confirm their desire to abandon the experiment. This was done to prevent people from using other applications on their phones instead of participating in the experiment. Participants could switch to the Match-It game at any time. During the Match-It game, the timer was stopped to ensure that everyone would spend ten minutes playing the Math Squares puzzle in order for performance to be compared accurately.

Measures

During the experiment the following data were captured:

1) *Complete Wins*: The number of Math Squares puzzles completed correctly, with all six numbers filled in accurately.

2) *Total Switches*: The number of times that the user switched from the Math Squares puzzle to the Match-It game.
3) *Between Switches*: The number of times the user switched to the Match-It game in between finishing one Math Squares puzzle and starting the next puzzle.
4) *During Switches*: The number of times the user switched to the Match-It game after starting a Math Squares puzzle but before submitting the puzzle.
5) *Average Score per Match-It Game*: Participants were given one point each time they tapped the correct shape. The total number of points was recorded and averaged together.
6) *Average Seconds per Match-It Game*: The total number of seconds played for each Match-It game was recorded and averaged at the end.

RESULTS

Participants

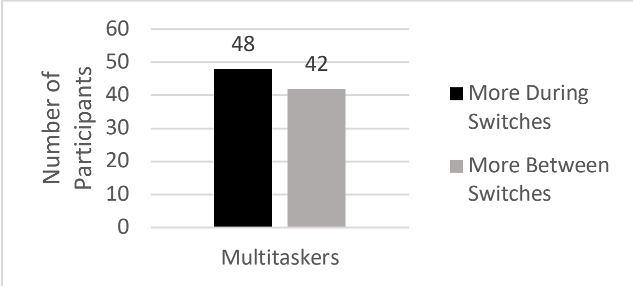
The 180 participants' data was analyzed for demographic information. In regards to ethnicity: 111 (61.67%) were Caucasian, 38 (21.11%) were Asian/Pacific Islander, 14 (7.78%) were Hispanic/Latino, 13 (7.22%) were Black/African American, 3 (1.67%) were Middle Eastern, and 1 (0.56%) identified as "other". The breakdown of age is as follows: 106 (58.89%) were 18-30, 54 (30%) were 31-40, 15 (8.33%) were 41-50, 2 (1.11%) were 51-60, and 3 (1.67%) were 61 or older.

Of the 180 participants, 78 (43.33%) of the participants were monotaskers, participants who did not switch during the experiment, while 102 (56.67%) were multitaskers, participants with at least one switch. The number of switches among multitaskers ranged from one to thirteen with a mean of 2.9. Of the multitaskers, 32 (31.37%) switched only one time while the remaining 70 participants (68.63%) had two or more switches. Participants completed an average of 5.8 Math Squares puzzles, and had an average of 3.07 Complete Wins. Participants spent an average time of 16.43 seconds on each Match-It game and 119.17 seconds on each Math Squares puzzle.

Test of Hypothesis 1

In order to determine whether participants would choose to task switch more frequently at breakpoints between subtasks rather than during a task, we compared Between Switches and During Switches for the 102 multitaskers. 42 participants had more Between Switches than During Switches, and 48 had more During Switches than Between Switches (see Figure 3). There were 12 participants who had an equal number of Between Switches and During Switches. A sign test shows that there were no significant differences between the number of Between and During Switches for participants ($z=0.63$, $p=0.53$), therefore H1 is not supported.

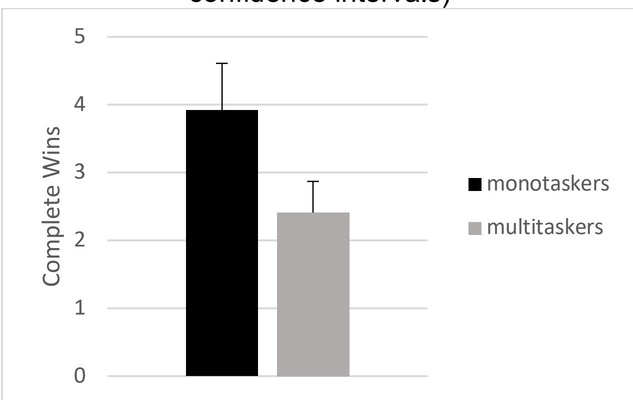
Figure 3: Comparison of participants with more Between Switches vs. more During Switches



Test of Hypothesis 2

To determine whether participants who multitask on mobile devices have lower performance on the primary task than those who do not multitask, we compared the performance on the Math Squares puzzle for multitaskers (102 participants who had at least one switch) and monotaskers (78 participants who had zero switches). The measurements used to gauge performance on the Math Squares puzzle (the primary task) was Complete Wins. The range of Complete Wins was zero to twelve, and the overall mean of Complete Wins was 3.067. As shown in Figure 4, there was a significant difference in the performance for monotaskers ($M=3.92$) over multitaskers ($M=2.41$), $t(178)=3.79$, $p=0.0002$. These results suggest that switches in general may negatively impact performance of a primary task.

Figure 4: Comparison of Complete Wins for monotaskers and multitaskers (error bars = 95% confidence intervals)



Furthermore, results of a Pearson correlation indicate that there is a significant and negative association between Total Switches and Complete Wins ($\rho=-0.22$, $p=0.003$). These results suggest that there is a negative relationship between people switching away from the primary task and their performance. Therefore, H2 was supported.

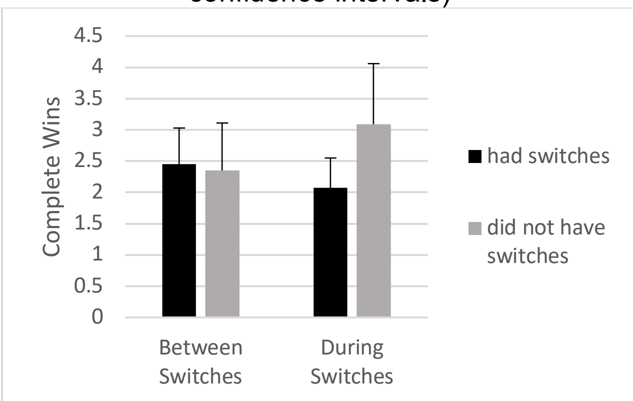
In addition to analyzing the effect of Total Switches on performance of the primary task, we also examined the amount of time spent on the Match-It game. A correlation shows a negative relationship between the average time spent playing the Match-It game and Complete Wins (ρ=-0.17, p=0.019). Even though the timer on the primary task was stopped and all participants spent the same ten minutes working on the primary task, the more time away from the task the lower the users performance. This suggests that the number of switches is not the only factor affecting performance on the primary task. The amount of time spent away from the primary task before returning may also impact the results.

Test of Hypothesis 3

To analyze whether task switches during a subtask negatively affect performance on the primary task more than task switches between subtasks, we examined the data for multitaskers (N=102) and explored the performance effects of During Switches and Between Switches.

For all multitaskers, a t-test showed that participants who had During Switches had significantly fewer Complete Wins than participants who had no During Switches, t(100)=2.13, p=0.036. However, performance for participants who had Between Switches was not significantly different from participants who did not have Between Switches, t(100)=-0.20, p=0.843, see Figure 5. Therefore, H3 was supported.

Figure 5: Performance differences between participants who had and did not have During Switches and participants who had and did not have Between Switches (error bars = 95% confidence intervals)



For multitaskers, we also computed a correlation comparing Complete Wins with During Switches and Between Switches. There is a significant negative relationship between Complete Wins and During Switches (ρ =-0.22, p=0.027), but no significant relationship between Complete Wins and Between Switches (ρ=0.06, p=0.545). These results suggest that switches during a subtask negatively affect performance on a primary task, while switches in between subtasks do not have a significant negative effect.

We also examined the difference in performance among those who had only Between Switches, only During Switches, both During and Between Switches, and no switches at all (Table 1). An ANOVA shows that there is no significant difference between those with no switches and those with only Between Switches (F(1, 110)=1.88, p=0.1728), but there is a significant difference between those with no switches and those with During Switches (F(1, 113)=7.80, p=0.0061). Furthermore, while there is no significant difference between those with only During Switches and those with both During Switches and Between Switches (F(1, 66)=1.63, p=0.2068), there is a significant difference between those with Between Switches and both During and Between

(F(1, 63)=5.77, p=0.0192). This suggests that the harmful effect of task switching is worse when those switches are During Switches, rather than Between Switches.

Table 1: Complete Wins and Switches

Complete Wins	N	Switches
3.92	78	No Switches
3.09	34	Only Between Switches
2.35	37	Only During Switches
1.74	31	Both During and Between Switches

DISCUSSION

In this experiment, there were no significant differences between the number of people who had more During Switches and the number of people who had more Between Switches. This seems to contradict previous studies that have determined that people prefer to task switch at breakpoints (Payne et al., 2007) and that when given a choice about when to respond to an interruption, people prefer to wait for a break between subtasks (Salvucci & Bogunovich, 2010). Similarly, Katidioti et al. (2014) found that participants preferred to switch at a moment of lower workload even in situations where they were able to switch completely at their own discretion. Our results also seem to contradict previous studies in the interruption literature that people prefer to task switch in response to an interruption when the interruption comes at a breakpoint (Adamczyk & Bailey, 2004; Iqbal & Bailey, 2005; 2008; Bogunovich & Salvucci, 2011; Borst, Taatgen, & Rijn, 2015).

There are a few possible reasons for this discrepancy. There may be factors that we are not accounting for such as the difficulty of the primary task. People may switch more/less depending on the difficulty level of the task. It is also possible that had we examined between switches not after completion of a puzzle but between-row or between-column switches vs. during-row and during-column switches we may have had different results. Finally, the low number of switches in our experiment may have led to the different results. Since participants were allowed to switch to the Match-It game as many or as few times as desired, some preferred to switch only once or not at all. Out of 180 participants whose data were analyzed, only 102 (56.67%) had any switches at all, and 32 (17.8%) had only one switch during the ten minutes. Leiva et al. (2012) found that people do not switch apps on their smartphones very often. It is possible that if participants were required to switch a fixed number of times (with the timing still at their discretion), that a clearer pattern would emerge.

Similar to previous studies (Hembrooke & Gay, 2003; Adler & Benbunan-Fich, 2012), we have found negative results when multitasking with technology. Specifically, our research demonstrates that these negative effects are also found when people multitask with smartphones. There was a strong negative relationship between primary task performance and any amount of switching.

In addition to the number of switches, we also found that time away from the primary task lowered performance. When participants in our experiment switched to the Match-It game the timer on the Math Squares puzzle was stopped, so all participants spent the same ten minutes on the primary task. Despite this, there was a significant negative relationship between time away from the primary task and number of Complete Wins. The more participants were away from the primary task, the lower their performance. This supports previous research by Hembrooke and Gay (2003) who found that time away from the primary task had more of a negative impact on performance than the content of the interruption.

The data in this research also supports Hypothesis 3 that switching in the middle of a subtask on mobile devices is more detrimental to performance on a primary task than switches in between subtasks. This suggests that the negative effects on performance can result from the amount of time it takes to remember information about the primary task after the end of an interruption. This supports previous research that found a time lag after interruptions from the primary task (Iqbal & Bailey, 2005). When the task switch occurs at a breakpoint there is no need to remember information and the primary task can be resumed immediately. In this research, participants who switched during the Math Squares puzzle would have to remember their thought process and strategy when returning to the task which negatively affected their performance.

While there were no significant differences between those who monotasked and those who switched in between subtasks, performance was significantly lower for those who switched during the subtask than those with no switches. This implies that some switching may not be detrimental if it is between subtasks and not during the task.

Implications

Previous research has shown that there are negative consequences to multitasking with technology in general (Hembrooke & Gay, 2003; Adler & Benbunan-Fich, 2012). Since smartphones in particular are used in a way that facilitates multitasking that can become habit forming (Oulasvirta et al., 2005; Böhmer et al., 2011; Oulasvirta et al., 2012), it is important to understand the effects of multitasking with smartphones specifically. Previous research has shown that performance on smartphones can be lower than on desktops, yet people still prefer smartphones (Adepu & Adler, 2016). This research has demonstrated that similar to computer-based multitasking, multitasking with smartphones negatively affects performance on a primary task. In addition, the more time away from the primary task, the lower users' performance. Furthermore, this research suggests that there are smarter ways to multitask which can lessen these negative effects. While previous literature discusses breakpoints in terms of external interruptions, we found that task switches during breakpoints can minimize the negative effects of multitasking. People who task switch can wait for a natural breakpoint in their primary task to switch to a secondary task. This will allow them to continue their multitasking behavior while minimizing negative effects on performance of their primary task.

Limitations and Future Research

Although Mechanical Turk provides many benefits to researchers, there are some limitations as well. Because participants completed the experiments on their own, it is possible that they experienced other distractions in their environment other than the Match-It game. It is also possible that participants left the experiment for part of the ten minute time limit. However, these limitations may be balanced out by the benefits of Mechanical Turk, which allowed us to recruit a larger and more diverse group of participants than we could have recruited on a college campus.

Another limitation in our research was the low number of participants who chose to switch to the secondary task. It is possible that we would have gotten better results had there been more of an incentive to switch to the secondary task. Requiring participants to multitask a fixed number of times while still allowing them to switch at a desired time, might give a clearer picture of when people are more likely to multitask. Data from participants who only switched a couple of times may not provide enough information to determine when they may have switched when constantly multitasking.

Future research may also look at the amount of time spent away from the primary task in between subtasks and during a subtask. In this experiment, only the average amount of time per Match-It game was recorded, and a negative relationship between Average Time per Match-It game and performance on the Math Squares puzzle was found. It would be interesting to see if total time spent on the Match-It game in the middle of a puzzle had a worse effect than total time in between puzzles.

CONCLUSION

People multitask constantly in multiple environments such as the workplace and school. Multitasking with technology is very common, and is especially pervasive when people are using mobile devices. Prior research has established the negative effects of multitasking when using technology in general. This research focuses on smartphone-based multitasking and found negative consequences of multitasking on smartphones. Furthermore, the more time participants spent away from the primary task, the lower their performance. While this study did not find any difference between the number of people who multitasked in between subtasks and during a subtask, the difference in performance between these two groups was significant. Switches during a subtask had a significant negative effect on performance, while switches in between subtasks did not have a significant effect. This suggests that the negative effects of discretionary task interleaving can be lessened by the timing of the switch. Waiting to switch to a different task at a natural breakpoint can reduce the negative consequences of multitasking on mobile devices.

ACKNOWLEDGEMENTS

We would like to thank Northeastern Illinois University Committee on Organized Research grant for funding this research.

REFERENCES

Adamczyk, P. D., & Bailey, B. P. (2004). If not now, when?: The effects of interruption at different moments within task execution. *SIGCHI conference on Human factors in computing systems*. Vienna, Austria: ACM Press, 271-278.

Adepu, S., & Adler, R. F. (2016). A comparison of performance and preference on mobile devices vs. desktop computers. *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. 1-7.

Adler, R. F., & Benbunan-Fich, R. (2012). Juggling on a high wire: Multitasking effects on performance. *International Journal of Human-Computer Studies*, 70(2), 156-168.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. . *Political Analysis*, 20(3), 351-368.

Bogunovich, P., & Salvucci, D. (2011). The effects of time constraints on user behavior for deferrable interruptions. *SIGCHI Conference on Human Factors in Computing Systems (CHI 2011)*. Vancouver, BC, Canada: ACM, 3123-3126.

Böhmer, M., Hecht, B., Schöning, J., Krüger, A., & Bauer, G. (2011). Falling asleep with Angry Birds, Facebook and Kindle: A large scale study on mobile application usage. *13th International Conference on Human Computer Interaction with Mobile Devices and Services* ACM. 47-56.

Böhmer, M., Lander, C., Gehring, S., Brumby, D., & Krüger, A. (2014). Interrupted by a Phone Call: Exploring Designs for Lowering the Impact of Call Notifications for Smartphone Users. *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI 2014)*. Toronto, ON, Canada: ACM.

Borst, J. P., Taatgen, N. A., & Rijn, H. v. (2015). What Makes Interruptions Disruptive?: A Process-Model Account of the Effects of the Problem State Bottleneck on Task Interruption and Resumption. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Seoul, Republic of Korea: ACM, 2971-2980.

Bowman, L. L., Levine, L. E., Waite, B. M., & Gendron, M. (2010). Can students really multitask? An experimental study of instant messaging while reading. *Computers & Education*, 54(4), 927-931.

Bradstreet Grinols, A., & Rajesh, R. (2014). Multitasking With smartphones in the College Classroom. *Business and Professional Communication Quarterly*, 77(1).

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1), 3-5.

Coens, J., Reynvoet, B., & Clarebout, G. (2011). Mobile learning. Can students really multitask? *Journal of Educational Multimedia and Hypermedia*, 20(1), 5-20.

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. *SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. Vienna, Austria: ACM Press, 175-182.

Dabbish, L., Mark, G., & González, V. M. (2011). Why do I keep interrupting myself?: Environment, habit and self-interruption. *SIGCHI Conference on Human Factors in Computing Systems* ACM. 3127-3130.

Fischer, J. E., Greenhalgh, C., & Benford, S. (2011). Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*. Stockholm, Sweden: ACM, 181-190.

Gonzalez, V. M., & Mark, G. (2004). 'Constant, constant, multi-tasking craziness': Managing multiple working spheres. *SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. Vienna, Austria: ACM Press, 113-120.

Hauser, D. J., & Schwarz, N. (2015). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, 48(1), 400-407.

Hembrooke, H., & Gay, G. (2003). The laptop and the lecture: The effects of multitasking in learning environments. *Journal of Computing in Higher Education*, 15(1), 46-64.

Ho, J., & Intille, S. S. (2005). Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Portland, Oregon, USA: ACM, 909-918.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The Online Laboratory: Conducting Experiments in a Real Labor Market. *Experimental Economics*, 14(3), 399-425.

Iqbal, S. T., & Bailey, B. P. (2005). Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. *CHI '05 Extended Abstracts on Human Factors in Computing Systems*. Portland, OR, USA: ACM, 1489-1492.

Iqbal, S. T., & Bailey, B. P. (2007). Understanding and developing models for detecting and differentiating breakpoints during interactive tasks. *SIGCHI conference on Human factors in computing systems (CHI '07)*. San Jose, California, USA: ACM, 697-706.

Iqbal, S. T., & Bailey, B. P. (2008). Effects of intelligent notification management on users and their tasks. *twenty-sixth annual SIGCHI conference on Human factors in computing systems, CHI '08*. Florence, Italy: ACM, 93-102.

Jacobsen, W. C., & Forste, R. (2011). The wired generation: Academic and social outcomes of electronic media use among university students. *Cyberpsychology, Behavior, and Social Networking*, 14(5), 275-280.

Janssen, C. P., Brumby, D. P., & Garnett, R. (2012). Natural break points: The influence of priorities, and cognitive and motor cues on dual-task interleaving. *Journal of Cognitive Engineering and Decision Making*, 6, 5-29.

Jin, J., & Dabbish, L. A. (2009). Self-interruption on the computer: a typology of discretionary task interleaving. *SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. Boston, MA, USA: ACM Press, 1799-1808.

Junco, R. (2012). In-class multitasking and academic performance. *Computers in Human Behavior*, 28(6), 2236-2243.

Junco, R., & Cotten, S. R. (2012). No A 4 U: The relationship between multitasking and academic performance. *Comput. Educ.*, 59(2), 505-514.

Kalkbrenner, J., & McCampbell, A. (2011). The Advent of Smartphones: A Study on the Effect of Handheld Electronics on Personal and Professional Productivity. *Journal of Applied Global Research*, 4(8), 1-9.

Katidioti, I., Borst, J. P., & Taatgen, N. A. (2014). What Happens When We Switch Tasks: Pupil Dilation in Multitasking. *Journal of Experimental Psychology: Applied*, 20(4), 380-396.

Leiva, L., Bohmer, M., Gehring, S., & Kruger, A. (2012). Back to the app: the costs of mobile application interruptions. *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*. San Francisco, California, USA: ACM, 291-294.

Levine, L. E., Waite, B. M., & Bowman, L. L. (2007). Electronic media use, reading, and academic distractibility in college youth. *CyberPsychology & Behavior*, 10(4), 560-566.

Liu, X., Cao, J., Tang, S., He, Z., & Wen, J. (2017). Drive Now, Text Later: Nonintrusive Texting-While-Driving Detection Using Smartphones. *IEEE Transactions on Mobile Computing*, 16(1), 73-86.

Mason, W., & Suri, S. (2012). Conducting Behavioral Research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44(1), 1-23.

McFarlane, D. C. (2002). Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction*, 17(1), 63-139.

Miyata, Y., & Norman, D. A. (1986). Psychological issues in support of multiple activities. In D. A. Norman, & S. W. Draper (Eds.), *User Centered System Design*. Hillsdale, N.J., USA: Lawrence Erlbaum, 265-284.

Nagata, S. F. (2003). Multitasking and interruptions during mobile web tasks. *47th Annual Meeting of the Human Factors and Ergonomics Society*, 1341-1345.

Okoshi, T., Ramos, J., Nozaki, H., Nakazawa, J., Dey, A. K., & Tokuda, H. (2015a). Attelia: Reducing user's cognitive load due to interruptive notifications on smart phones. *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 96-104.

Okoshi, T., Ramos, J., Nozaki, H., Nakazawa, J., Dey, A. K., & Tokuda, H. (2015b). Reducing users' perceived mental effort due to interruptive notifications in multi-device mobile environments. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Osaka, Japan: ACM, 475-486.

Oulasvirta, A., Rattenbury, T., Ma, L., & Raita, E. (2012). Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing,*, 16(1), 105-114.

Oulasvirta, A., Tamminen, S., Roto, V., & Kuorelahti, J. (2005). Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. *SIGCHI conference on Human factors in computing systems*: ACM. 919-928.

Payne, S. J., Duggan, G. B., & Neth, H. (2007). Discretionary task interleaving: Heuristics for time allocation in cognitive foraging. *Journal of Experimental Psychology: General*, 136(3), 370-388.

Rosen, L. D., Carrier, L. M., & Cheever, N. A. (2013). Facebook and texting made me do it: Media-induced task-switching while studying. *Comput. Hum. Behav.*, 29(3), 948-958.

Salvucci, D. D., & Bogunovich, P. (2010). Multitasking and monotasking: the effects of mental workload on deferred task interruptions. *SIGCHI Conference on Human Factors in Computing Systems (CHI 2010)*. Atlanta, Georgia, USA: ACM, 85-88.

Tolmie, P., Crabtree, A., Rodden, T., & Benford, S. (2008). "Are you watching this film or what?": interruption and the juggling of cohorts. *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. San Diego, CA, USA: ACM, 257-266.

Wood, E., Zivcakova, L., Gentile, P., Archer, K., De Pasquale, D., & Nosko, A. (2012). Examining the impact of off-task multi-tasking with technology on real-time classroom learning. *Computers & Education*, 58(1), 365-374.