Journal of Experimental Psychology: Applied

What Happens When We Switch Tasks: Pupil Dilation in Multitasking

Ioanna Katidioti, Jelmer P. Borst, and Niels A. Taatgen

Online First Publication, October 27, 2014. http://dx.doi.org/10.1037/xap0000031

CITATION

Katidioti, I., Borst, J. P., & Taatgen, N. A. (2014, October 27). What Happens When We Switch Tasks: Pupil Dilation in Multitasking. *Journal of Experimental Psychology: Applied*. Advance online publication. http://dx.doi.org/10.1037/xap0000031

Journal of Experimental Psychology: Applied

1076-898X/14/$12.00 http://dx.doi.org/10.1037/xap0000031

What Happens When We Switch Tasks: Pupil Dilation in Multitasking

Ioanna Katidioti, Jelmer P. Borst, and Niels A. Taatgen
University of Groningen

Interruption studies typically focus on external interruptions, even though self-interruptions occur at least as often in real work environments. In this article, we therefore contrast external interruptions with self-interruptions. Three multitasking experiments were conducted, in which we examined changes in pupil size when participants switched from a primary to a secondary task. Results showed an increase in pupil dilation several seconds before a self-interruption, which we could attribute to the decision to switch. This indicates that the decision takes a relatively large amount of time. This was supported by the fact that in Experiment 2, participants were significantly slower on the self-interruption blocks than on the external interruption blocks. These findings suggest that the decision to switch is costly, but may also be open for modification through appropriate training. In addition, we propose that if one must switch tasks, it can be more efficient to implement a forced switch after the completion of a subtask instead of leaving the decision to the user.

Keywords: multitasking, interruption, pupil size, decision making

Self-interruptions are a very common occurrence. Students interrupt their studying to check social media, office employees stop working to check the news online, professors suspend their writing to get another cup of coffee. Several observational studies have provided a scientific background for these everyday experiences. For example, Gonzalez and Mark (2004) observed information workers in an office environment and found that they switch between tasks on average every 3 min, and Chisholm and colleagues reported that physicians in an emergency department were interrupted on average about 50 times in 180 min (Chisholm, Collison, Nelson, & Cordell, 2000). Observing students in their home environments, Rosen, Carrier, and Cheever's (2013) found that they studied on average 6 min before they interrupted themselves, usually to text or engage in social media.

Besides interrupting ourselves, interruptions can also have an external source (i.e., a pop-up message, a phone ringing, another person walking in, etc.). Although most experiments on interruptions focus on external interruptions (e.g., Hodgetts & Jones, 2003, 2006; Monk, Boehm-Davis & Trafton, 2004; Monk, Trafton, & Boehm-Davis, 2008), observational studies show that people interrupt themselves as often as they are interrupted by external events. Czerwinski, Horvitz, and Wilhite (2004) analyzed a week-long multitasking diary of information workers and found that 40% of the interruptions reported were internal. Mark, Gonzalez, and Harris (2005) observed information workers and found that 52% of the interruptions were internal, and Gonzalez and Mark (2004) report that percentage to be 49%. Gould, Cox, and Brumby (2013) conducted an online experiment and found that 80% of the participants switched from the experiment at least once, although they were warned that switching would result in a reduction in remuneration.

Because self-interruptions are so prevalent, and it is well known that at least external interruptions lead to a considerable decrease in performance (e.g., Monk et al., 2008), it is important to gain more knowledge on self-interruptions. Although external interruptions can be very disruptive, there are ways to minimize them: turn off your cell phone, disable e-mail and instant message pop-ups, and lock the office door. Self-interruptions are more difficult to manage. With the current series of three experiments we aim to learn more about self-interruptions, specifically comparing them with external interruptions and investigating the effects of a lag between the start of the interruption and the start of the secondary task. To create a precise cognitive time course of what happens when people self-interrupt—and how that contrasts with external interruptions—we used pupil dilation as a psychophysiological measure in our study. Pupil dilation is known to reflect changes in cognitive processing and has a continuous nature that allows the creation of a cognitive time course. It does not disrupt the user and it is more natural and less intrusive than other methods (such as EEG). Before we describe our study, we will first provide a background on effects of interruptions and on the use of pupil dilation.

Interruption Effects

Interruptions are a form of sequential multitasking (see Salvucci & Taatgen, 2011, for a review on types of multitasking). Interruptions can be defined as follows (Trafton, Altmann, Brock, & Mintz, 2003): People are engaged in a *primary* task, which is interrupted by a *secondary* task. The interruption can be either an external interruption or a self-interruption. After completing the secondary task, the primary task is resumed. Sometimes there is an alert before the secondary task (e.g., a sound). The time between

---

Ioanna Katidioti, Jelmer P. Borst, and Niels A. Taatgen, Department of Artificial Intelligence, University of Groningen.

This research was funded by European Research Council (ERC) ERC-StG Grant 283597 awarded to Niels Taatgen.

Correspondence concerning this article should be addressed to Ioanna Katidioti, Department of Artificial Intelligence, University of Groningen, 9747 AG Groningen, The Netherlands. E-mail: i.katidioti@rug.nl

the alert or the interruption moment and the beginning of the secondary task is called *interruption lag*. The time between completing the secondary task and returning to the primary task is called *resumption lag*. Figure 1 shows the time course of an interruption. The resumption lag is one of the indicators of the negative effects of interruptions on task performance: It is time that would not be lost without the interruption. Even switching between the simplest tasks creates such a resumption lag (e.g., Allport & Wylie, 2000; Rogers & Monsell, 1995; Trafton et al., 2003).

One of the major cognitive theories on interruptions is Altmann and Trafton's (2002) memory for goals theory. According to memory for goals, a person's primary task goal is suspended and starts to decay when an interruption occurs, and the secondary task goal is activated. When the person returns to the primary task, the goal must be resumed. This resumption process will take time and is a major cause of the resumption lag. Much of the interruption research has focused on factors that affect the resumption lag. The main factors are the timing of the interruption (i.e., when it happens in the primary task), the duration of the interruption, whether or not there was an alert before the secondary task began, if there is time for rehearsal of the primary task goal, and the difficulty of the secondary task (e.g., Iqbal & Bailey, 2005; Monk et al., 2008; Trafton et al., 2003). In the current study we will mainly focus on the timing of the interruptions and on the effects of a delay before the beginning of the secondary task.

Although most studies agree that interruptions have a negative effect on the primary task (not just in time costs but also in more errors, e.g., Brumby, Cox, Back, & Gould, 2013), the timing of the interruption can make it more or less disruptive. Iqbal and Bailey (2005) interrupted their participants at low-workload moments (when a subtask was finished), high-workload moments (while performing a subtask), or at random moments. Results showed that being interrupted at a predictable low-workload moment caused a smaller resumption lag. Monk et al. (2004) interrupted the participants in one of their experiments midsubtask and after subtasks and found that an interruption midsubtask was more disruptive. Thus, being interrupted at a high-workload moment seems to be more disruptive than being interrupted at a low-workload moment. The majority of this type of studies focused on external interruptions. However, the same effects can be found in self-interruptions: If people interrupt themselves on a high-workload moment they are more negatively affected than if they interrupt themselves on a low-workload moment (Katidioti & Taatgen, 2014).

Another line of research on interruptions focuses on the effect of alerts and delays before the beginning of the secondary task. The presence or absence of an alert or a delay before the secondary task begins can affect the duration of the resumption lag. Trafton, Altmann, Brock, and Mintz (2003) performed an experiment where participants either had an alert before the secondary task began, followed by an 8-s delay or not. Their hypothesis was based on the memory for goals theory (Altmann & Trafton, 2002) and specifically the idea that a longer interruption lag gives people more time to prepare for the interruption and thus facilitates retrieval of information during resumption. Results showed that participants were much faster in resuming their primary task when they were alerted that they would switch tasks compared to not being alerted. In a similar study, Hodgetts and Jones (2003) found that participants resumed their primary task significantly slower when the switching occurred immediately after the completion of a subtask compared to when they faced a 3-s delay.

Monk, Trafton, and Boehm-Davis (2008) studied the effects of rehearsal of the primary task goal when faced with an interruption. They hypothesized that rehearsing the primary task while performing the secondary task could minimize the resumption lag, because it helps to avoid the decay of the goal (Altmann & Trafton, 2002). Results showed that a less-demanding secondary task lead to better performance and a later computational model (Salvucci, Monk, & Trafton, 2009) explained these effects by assuming rehearsal during the secondary task.

Although the majority of interruption studies focused on external interruptions, there are some more high-level studies that attempted to find the reasons behind self-interruptions. Because self-interruptions are an important issue in real-life situation and because they are hard to study in an experimental setting, most of these studies are observational. Dabbish, Mark, and Gonzalez (2011) analyzed the self-interruption observational data of Gonzalez and Mark (2004) and found that individual differences (in habits) and working in an open office were the most important factors of self-interruption, although the nature of work (dealing with clients or not) and time of the day also had an effect. In another observational study, Jin and Dabbish (2009) shadowed 13 people working with a computer and separated the self-interruptions into seven categories (adjustment, break, inquiry, recollection, routine, trigger, and wait), including positive and negative consequences of every category. Three of these categories (break, recollection, and routine) are caused by the person's cognitive state and the other four are caused by the environment or the person's physical state (e.g., there is a pause in the task, person is not sitting comfortably, etc.).

Self-interruption is not easy to study in an experimental environment because the reasons behind self-interrupting vary between people and are not easy to manipulate. However, some attempts have been made to study the mechanisms of self-interruption. Payne, Duggan, and Neth (2007) let participants allocate their time between two tasks freely and their results indicated that they self-interrupted either to temporarily abandon a task that is no longer rewarding or because of the tendency to switch to an

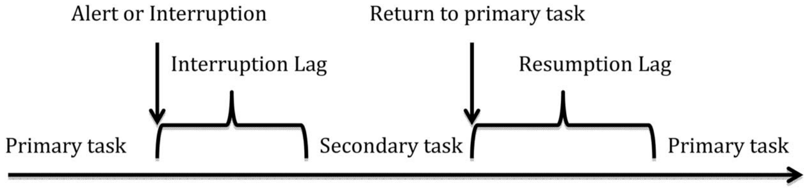

Figure 1. Time course of an interruption (based on Trafton et al., 2003).

unrelated task when a subtask is completed. Salvucci and Bogunovich (2010) gave participants freedom in choosing when to switch to the secondary task in an Internet and chat environment and found that they did not switch on high-workload moments but waited until the end of a subtask. However, Katidioti and Taatgen (2014) found that this rationality in self-interruption disappeared when participants were forced to wait while an Internet browser was loading. In this setting, participants preferred to switch tasks instead of waiting, in spite of being midsubtask. These results suggested that people having total control of the interruption does not mean that they will make optimal decisions.

Some studies investigated the differences between external interruptions and self-interruptions. Mark et al. (2005) observed the switching behavior of office workers and separated the data into external interruptions and self-interruptions. Tasks that were externally interrupted did not need more time to be completed than tasks that were self-interrupted, suggesting that the two kinds of interruptions do not create performance differences. There was a slight trend showing that an external interruption led to a faster resumption of people's work when compared with a self-interruption, but that was not enough to make them generally faster. Panepinto (2010) used forced (i.e., external interruption) and voluntary (i.e., self-interruption) task-switching between a Sudoku and a document proofreading task, expecting that forced task-switching would be more disruptive. However, she found no performance difference between these conditions. McFarlane (2002) conducted a study on external interruptions, using six different conditions. One of them (immediate) was a forced interruption that happened at a random moment, a typical external interruption. The condition in his study that was more like a self-interruption was the negotiated condition, which gave participants control over when they could handle the interruptions. He compared all conditions in many different performance criteria, but overall the negotiated condition was the best method to handle an interruption and the immediate condition was the worst. In general, data is inconclusive on whether self-interruptions, and thus control over interruptions, are better than external interruptions.

To sum up, although self-interruption is a very important everyday matter, there are only few experiments studying or comparing self-interruption and external interruption, leaving a gap in research on interruptions. There are also limited studies (but see Iqbal et al., 2004) that have any psychometric data on self-interruption. In the current study we plan to investigate both kinds of interruptions in the same experimental setup and contrast them using task performance and pupil dilation. In the next section we introduce pupil dilation as a psychometric measure and explain why we chose it for our study.

Pupil Dilation

The measurement of pupil dilation has been used in cognitive science at least since the 1960s (for an overview, see Beatty & Lucero-Wagoner, 2000). Apart from changes in light, the dilation of the pupil also reacts to a number of cognitive processes and can therefore be used to create a time frame of the cognitive system's reactions to certain tasks. It should be noted that the pupillary response to an event is not instantaneous: It peaks approximately 1 s after a stimulus (e.g., Beatty & Lucero-Wagoner, 2000; Hoeks & Levelt, 1993; Steinhauer & Hakerem, 1992; Wierda, van Rijn, Taatgen, & Martens, 2012).

There are many studies that use changes in pupil dilation to measure mental workload. Kahneman and Beatty (1966) aurally presented strings of digits (from three to seven different digits) to their participants and asked them to repeat them. Results showed that the pupil diameter increased with each digit spoken, reaching baseline after the last digit. Peavler (1974) conducted a similar experiment with the same results, although he increased the number of digits and found that the pupillary response reached an asymptote at the seventh or eighth digit. In another study Kahneman, Tursk, Shapiro, and Crider (1969) asked their participants to add 0, 1, or 3 to a number, with the pupil dilating more when they had to add 3 and less when they had to add 0. Iqbal, Adamczyk, Zheng, and Bailey (2005) used pupil dilation to measure cognitive workload changes during task execution. Analysis of the pupil dilation showed that the size of the pupil increased during a subtask and decreased when the subtask finished. Their conclusion was that these results reflect the effect of workload in pupil dilation, because workload decreases when the subtask is finished. These studies show that the pupil size increases as the mental workload increases. Pupil dilation is also used to study other forms of cognitive effort, such as Stroop effects (Laeng, Ørbo, Holmlund, & Miozzo, 2011) the complexity of tasks (e.g., Moresi, Adam, Rijcken, van Gerven, Kuipers, & Jolles, 2008; Prehn, Heekeren, & van der Meer, 2011) and the difficulty of retrieving information from memory (van Rijn, Dalenberg, Borst, & Sprenger, 2012).

Although cognitive effort is one of the main areas in which pupil dilation is used as a measure, increase in the size of the pupil can also reflect other kinds of cognitive processing. For instance, Richer and Beatty (1985) reported that pupil dilation increased more when participants had to press a "heavy" button (activated with a load of 1,250 g) than when they had to press a "light" button (activated with a load of 100 g), concluding that it is not the response but the move itself that creates the increase in pupil dilation. In Chiew and Braver's (2013) study, participants showed a greater increase in pupil dilation when they had to suppress the urge to press a button, that is, inhibit a response. Satterthwaite, Green, Myerson, Parker, Ramaratnam, and Buckner (2007) used a simple gambling task and found that the more uncertain the participant was about the result, the more the pupil dilation increased. If a result was uncertain, decision making was harder and that reflected on the pupil.

Current Experiment

In the current study we will use pupil dilation to investigate what happens when people switch between tasks, and to see if there are differences between external interruptions and self-interruptions. There are a number of cognitive processes and actions that may take place when people are interrupted or interrupt themselves, which can cause a change in the dilation of the pupil:

1. Decision to switch tasks: The process of deciding to switch from the primary to the secondary task. This is the main difference between a self-interruption and an external interruption.

2. Suspension of the current goal: According to memory for goals theory (Altmann & Trafton, 2002), the goal of the primary task (e.g., the sentence a person is currently writing) is suspended when an interruption occurs (e.g., the phone rings) and the person's attention is shifted to the secondary task. When the secondary task is over, the goal of the primary task must be resumed. Suspending the current goal might include processes like rehearsal, to make sure that the goal is still available when returning to the primary task.

3. Click of the mouse button: In our experimental setup, participants that decide to self-interrupt can switch from the primary to the secondary task by clicking the mouse button. Clicking the mouse button is known to produce a small increase in pupil dilation (Richer & Beatty, 1985).

4. The actual switch from the primary task goal to the secondary task goal: The moment when attention shifts from the primary to the secondary task. This is the same process for self-interruptions and external interruptions.

5. Preparation for starting the secondary task: For example, initiating the goal of the secondary task, setting up mental resources (e.g., working memory; Borst, Buwalda, Van Rijn, & Taatgen, 2013; Borst, Taatgen, & Van Rijn, 2010) or trying to remember the instructions.

By manipulating the type of interruption and the interruption lag we will investigate the processes behind interruptions. In addition, because pupil dilation is a continuous measure, it will provide us with a time course of cognitive processes around interruptions.

Experiment 1

Method

Design. The main task of the experiment was a variation of a children's memory game, which is usually known as Concentration or Memory. Typically, the game consists of a deck of cards, containing pairs of matching items (usually images). At the start of the game, all cards are arranged face down in random positions. The players open the cards in pairs: They first open one card and after inspecting this card they choose a second one. If the cards match they stay open and the player scores a point, if not they are closed again. The goal of the game is to find all pairs (one-player version) or more pairs than your opponent (two-player version).

For the current experiment we altered this game in several ways (cf. Anderson, Fincham, Schneider, & Yang, 2012). In our instantiation there are 16 cards (eight pairs) with equations on them (in the form $a * X + b = c$, where X is the unknown variable and a, b, and c are integers, with a and b being in the range of 2 to 9), arranged in a 4 × 4 matrix (see Figure 2). Two cards are said to match when they have the same value for X (solutions were integers from 2 to 9). For example the cards "$2 * X + 2 = 12$" and "$3 * X + 4 = 19$" are a match, because $X = 5$ in both of them.

A second difference from the typical memory game was that cards are opened one at a time instead of in pairs. A match is made when the opened card matches the previously opened card. When there is no match, the previously opened card closes again but the last card remains open. In the classic game the first card remains open after the player opens a second card, after which both cards are closed. In this way we reduced the complexity of the game to make the analysis more straightforward (given that there is only one player it did not affect the way the game is played).

The secondary task of the experiment was a working memory task called n-back (Kirchner, 1958). In this task, participants see letters appearing one by one and have to judge if the letter they are seeing is the same as the n-th letter back on the list. In this experiment we used 2-back, which means participants had to judge if the letter they were seeing was the same as two letters back and respond accordingly. One of the reasons we chose 2-back as the secondary task was to eliminate rehearsal of the primary task during the secondary task. Monk et al. (2008) used three different levels of secondary task difficulty to facilitate or eliminate the rehearsing of the primary task. Their most difficult secondary task was 1-back, an easier version of the n-back task. By using 2-back,

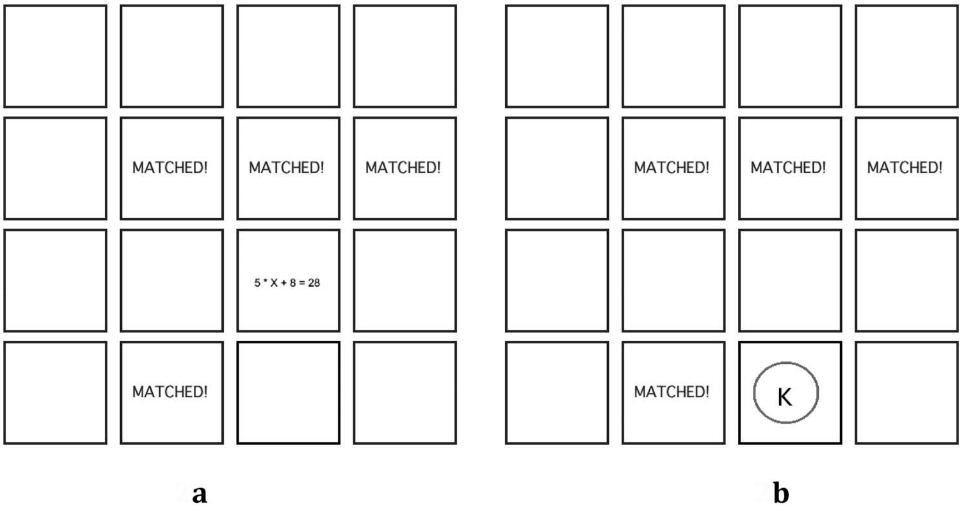

Figure 2. The interface of the primary task (memory game) (2a) and the secondary task (*n*-back) (2b).

we can be reasonably sure that there will be no rehearsal of the primary task while performing the secondary task.

The tasks we chose are not normal office tasks, but they share many common aspects with them. The memory game resembles many working environment or studying situations, where some mental work has to be performed and then memorized, for example, making a calculation using a calculator and memorizing the result while entering data on worksheets or doing homework. N-back is a typical working memory task that resembles some popular interrupting tasks such as simple online games or working environment tasks such as keeping something in the working memory in order type it in a browser.

Procedure. To perform the main task, participants had to click on a card with the left mouse button, mentally solve the equation, remember the value of X and continue by clicking on another card. If they clicked on a card that matched the one that was previously clicked, the word “MATCHED” appeared on the back of these cards and they could not be clicked again (Figure 2a). A move in this memory game could be either a “new card” (when participants opened a card for the first time), a “revisit” (when they opened a card they had opened before), or a “match” (when they opened a card that was a match to the previous card they had opened). A “match” was considered a “lucky match” when participants clicked on a card for the first time and it happened to match with the previous card they opened.

To switch to the secondary task, participants had to click with the right mouse button on an empty card. The word “n-back” appeared on the card for 0.5 s after which the n-back task started. The n-back task lasted 15 s and the probability of a letter being the same as two letters before was 50%. Each letter stayed on the screen for 1.5 s, waiting for a response from the participant. The participant pressed “z” if the letter was the same as two letters before and “x” otherwise. They were given feedback after every letter, either a green circle around the letter if the answer was correct, or a red circle if the answer was wrong or there was no answer within the time limit (Figure 2b). After the feedback, there was 0.5-s blank screen as an interval between two letters.

Participants could not switch to the n-back task before opening at least two memory cards at the beginning of a game or after returning from the n-back task. In addition, they could not switch to n-back if there were only two cards left in a game. With these restrictions (about which participants were informed before the experiment started) we created a clear sequence of performing a primary task, switching to the secondary task, returning to the primary task, switching again after a while and so forth. Two factors were manipulated in the experiment:

1. Voluntary/forced: In the voluntary condition, participants were free to choose when to switch to the secondary task by clicking on an empty card with the right mouse button, as described above. The n-back task appeared in this box (Figure 2b). In the forced condition, the n-back task occurred at an unexpected moment, while the participant had already opened a card and was looking at the equation on it. Thus, the voluntary condition is comparable with self-interruptions, whereas the forced condition measures the effect of external interruptions.
2. Delay/no delay: After the switch, the word “n-back” appeared on the card for 0.5 s. In the delay condition there was a 3-s delay after that and before the appearance of the first letter. In the no delay condition the first letter appeared immediately after “n-back.”

Participants had to complete 12 blocks (three blocks delay/forced, three blocks delay/voluntary, three blocks no delay/forced, and three blocks no delay/voluntary, in random order). Each block was finished when all the cards in a memory game were matched. The participants were instructed to switch to the n-back task three times within a block (in the voluntary condition), otherwise they would get a penalty of a 30-s delay after the block finished. Before each block started, there was a message showing what kind of block it was going to be (e.g., delay–voluntary). They were also informed that the experiment would finish after 12 blocks were completed (approximately 1 hr). Before the experiment started, participants completed two practice blocks in the presence of the experimenter to familiarize them with the task. In the first of these blocks they were asked to match cards with numbers, and in the second block they had to solve equations. This practice phase lasted about 5 min.

Participants. Twenty-six participants (12 females, mean age 22.32) participated in the experiment. They all had normal or corrected-to-normal vision and received monetary compensation of 10 euros for participating. One participant (female) was removed for not following the instructions (she performed zero switches in the voluntary blocks). Three additional participants were removed, as they did not seem to solve the equations, but clicked randomly (see below for details). The remaining 22 participants (10 females) had a mean age of 21.8.

Apparatus and setup. Participants were tested individually in a small windowless room. They were seated at a desk with a 20 in. LCD monitor with screen resolution of 1,024 × 768 pixels and screen density of 64 pixels/inch. Participants were asked to use a chin-rest during the blocks of the experiment. The eyetracker was an Eyelink 1000 from SR Research, positioned approximately 45 cm from the edge of the desk.

Monocular pupil dilation was measured with a sample rate of 250 Hz. Calibration and drift correction were performed before the experiment started. A calibration accuracy of 0.8° was considered acceptable. Before each block began, drift correction was performed and participants looked for 0.5 s at a fixation cross.

Measurement and preprocessing of pupillary data. Eyeblinks were removed from the results, starting 100 ms before the blink and finishing 100 ms after, and replaced them by a linear interpolation. The data were downsampled to 100 Hz.

We were interested in how the pupil reacts to switching from the primary task and how it reacts to clicking a card. For that reason, we isolated the pupil dilation 5 s before and 10 s after every switch and every click. Switches that had more than 20% of their data (20% of the 15 s we plotted) interpolated were removed, which resulted in removal of 2% of the switches. We calculated the percentage change in the pupil dilation from baseline, which was defined by a very slow lowess filter, that is, a smooth curve that follows the pupil dilation data, given by a weighted linear least squares regression over the span (with a smoother span of two thirds; for more details see Cleveland, 1981).

Results

Behavioral results. To remove participants that did not actually perform the task, we set a threshold of an average time of 3 s per card per block. We assumed people could not solve an equation on average in less than 3 s per card per block. If a block had an average time of less than 3 s, it was removed, assuming that participants did not solve the equations in those blocks, but simply clicked as fast as they could. For all three experiments combined, blocks that were not removed from the analysis had an average of 36.9 clicks per block as opposed to 50.9 for the blocks that were removed, reaffirming that they clicked randomly in those blocks (subjects needed a minimum of 16 clicks to finish a block). If a participant had half or more blocks under the threshold, that participant was removed. Three participants and six blocks from four other participants were removed. Removing these blocks and participants did neither alter the main behavioral results, nor the pupil dilation results.

The behavioral data in Table 1 show that the differences between conditions were minor. There was no statistically significant difference among conditions for any of these measures, as indicated by ANOVAs.

One of the findings in the literature is that self-interruptions are less disruptive than external interruptions (e.g., McFarlane, 2002). For the current experiment this predicts that subjects should have performed better in the voluntary condition than in the forced condition, for instance by taking less time or making fewer errors. Thus, there should be a difference between the forced and voluntary condition in the average time per memory game (time per block with the n-back time removed) or the number of revisits (i.e., the second and fourth row of Table 1). A high number of revisits could indicate that participants forgot many of the cards because of the interruption, therefore a difference between the two conditions could indicate that one kind of interruption was more disruptive than the other. However, in both measures there was no significant difference between the forced and the voluntary condition, with $t(21) = -1.15$, $p = .26$ and $d = 0.19$ for the time per memory game and $t(21) = 1.22$, $p = .24$ and $d = 0.19$ for the number of revisits.

Several studies indicated that subjects typically switch at low-workload moments (e.g., Salvucci & Bogunovich, 2010). In our experiment that would mean that subjects switch after a match, as that reduces the number of cards to keep available in (working) memory. Analyzing the switches in the voluntary condition, we found that participants strongly preferred to switch after they made a match: 14.53% of the switches were made after opening a new card, 8.55% after revisiting a card and 76.92% after a match. Switching after a match indeed occurred significantly more often, according to a repeated-measures ANOVA, $F(2, 42) = 45.96$, $p <$ .001, $\eta_p^2 = 0.69$. In the forced condition a switch occurred while participants were looking at a card. Switches therefore never occurred after a match, instead 76.39% of the switches happened when subjects were looking at a new card and 23.61% when they were revisiting a card.

In the delay condition participants had 3 s to prepare for the secondary task. However, having a 3-s delay before being forced to switch to another task did not improve performance in the primary task. Time per memory game (with the time of the delays and the n-back task removed) in the forced/delay condition was 190.45 s and in the forced/no delay condition 181.78 s, the difference of 8.67 s was not significant, $t(21) = 0.77$, $p = .45$, $d = 0.17$.

To obtain more specific information on the effects of the interruptions, we calculated the resumption lag. In the current experimental set up, resumption lag is the time after the end of the n-back task and before the next click on a card. Because in the forced condition participants were interrupted while they were looking at an open card and after the interruption returned to the same open card, it is not possible to calculate the resumption lag of this condition. Therefore, we can only compare voluntary/delay and voluntary/no delay conditions. There were in total 360 switches in the voluntary condition, but we removed 18 of them as outliers (the resumption lag was greater than 2 standard deviations from the average in these cases). The average resumption lag for the voluntary/delay condition was 3.07 s and 2.98 s for the voluntary/no delay condition, with that difference not being significant, $t(21) = 0.48$, $p = .63$, $d = 0.08$, confirming that delay did not have a beneficial effect on the resumption lag.

Pupil dilation results. A two-way ANOVA (voluntary/forced and delay/no-delay) was performed on every sample, followed by a False Discovery Rate (FDR) correction over all samples to correct for multiple comparisons (Benjamini & Hochberg, 1995). There was no interaction between conditions in increase of pupil dilation. Figure 3 shows the forced and voluntary conditions. Taking into account that pupil dilation has a one second delay when responding to an event, it is obvious that the increase in pupil dilation was significantly greater in the voluntary than in the forced condition some seconds before the switch and started declining before the switch was made. On the other hand, the pupil dilation in the forced condition showed no signs of increase before the switch and peaked suddenly at the moment of the switch. Some seconds later, there were again several fluctuations in pupil dila-

Table 1
Behavioral Data of Experiment 1 (Mean (SE))

		Condition			
	Average	Forced	Voluntary	Delay	No delay
Time per block (s)	263.38 (8.7)	259.89 (10.37)	266.9 (8.95)	261.73 (6.93)	257.75 (9.14)
Time per memory game (n-back time removed; s)	191.11 (8.57)	187.16 (9.9)	195.28 (8.73)	192.65 (8.98)	190.6 (9.45)
# of revisits per block	12.61 (1.74)	13.42 (2.16)	11.78 (1.52)	12.91 (1.75)	12.36 (1.9)
# of lucky matches per block	1.28 (0.06)	1.29 (0.09)	1.28 (0.09)	1.22 (0.08)	1.33 (0.06)
# of total clicks to complete a block	35.33 (1.75)	36.13 (2.22)	34.5 (1.49)	35.69 (1.77)	35.03 (1.9)
No. (x4) of switches to n-back per block	2.92 (0.05)	2.94 (0.03)	2.89 (0.09)	2.96 (0.06)	2.89 (0.06)

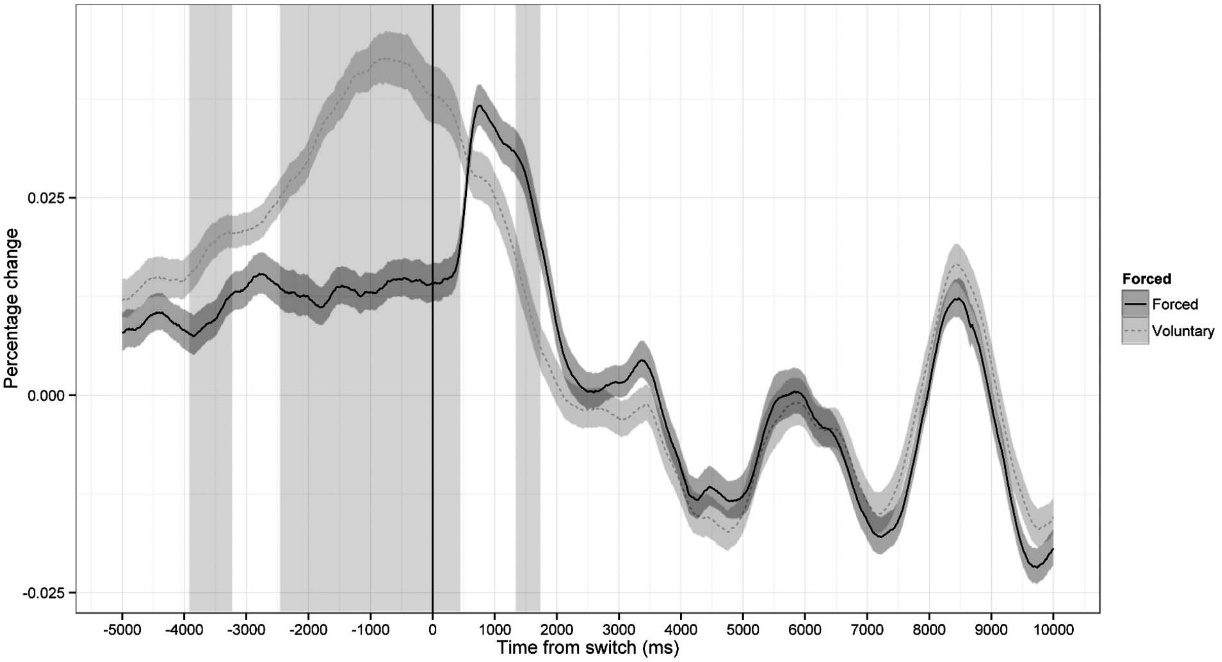

Figure 3. Average pupil dilation of all participants for the forced and voluntary conditions of Experiment 1 around the switch (0 is the switching point from memory game to the n-back task). Marked with dark gray color are the moments where there was a statistically significant difference ($p < .05$, FDR-corrected) between the forced and the voluntary conditions. Shading indicates a standard error.

tion for both conditions, indicating the participants' responses in the n-back task.

In Figure 4 we plotted the delay/voluntary and no delay/voluntary[1] conditions and indicated statistically significant ($p < .05$; FDR-corrected) differences between them. There was no difference in pupil dilation increase before and around the switch. The only difference occurred after the n-back task started and is due to the phase difference of stimulus appearance between the delay and no delay conditions.

Pupil dilation also increased when people clicked on a card. This dilation was not the same for the different kinds of cards (see Figure 5). It was greater when participants made a match, smaller when they were revisiting a card they had opened before and even smaller when they opened a card for the first time.

Discussion

Switching tasks created an increase in pupil dilation (Figure 3 and Figure 4). But what is the cause of that increase? As we indicated earlier in the article, switching tasks may involve the following processes that, in turn, may affect the pupil size:

1. Decision to switch tasks (applies only in self-interruption condition).
2. Suspension of the current goal.
3. Click of the mouse button.
4. The actual switch from the primary task goal to the secondary task goal.
5. Preparation for starting up the secondary task.

Figure 4 shows no difference in the reaction of the pupil between the delay and the no delay conditions. This suggests that suspension of the current goal and preparation for starting up the secondary task have no effect on pupil dilation. According to memory for goals (Altmann & Trafton, 2002), the extra time in the delay condition can be used to rehearse the interrupted goal. Otherwise, that time can be used to prepare for the interrupting task. The pupil size does not show evidence for either process. Moreover, there is no difference in performance between the conditions (see Table 1). That means that if suspension or preparation happens at all, it does not produce any benefit.

The actual click of the mouse (or the preparation for it) can at most have a very small contribution to the pupil response, if at all. This is obvious in Figure 5, where a click to open a new card creates a much smaller increase in pupil dilation than the click to switch tasks (compare with Figure 3 and Figure 4). This suggests that there is something more to voluntarily switching tasks than clicking the mouse button.

There was a large difference between the forced and the voluntary conditions before the switch (see Figure 3). Making a self-interruption (voluntary condition) created a reaction in the pupil some seconds before the switch, whereas being externally interrupted (forced interruption) created a sudden peak in pupil dilation

---

[1] We did not plot the pupil dilation increase for the forced/delay and forced/no delay conditions, because whatever processes may happen there (rehearsal, preparation) will be even stronger when participants choose themselves to switch tasks, that is, the voluntary condition.

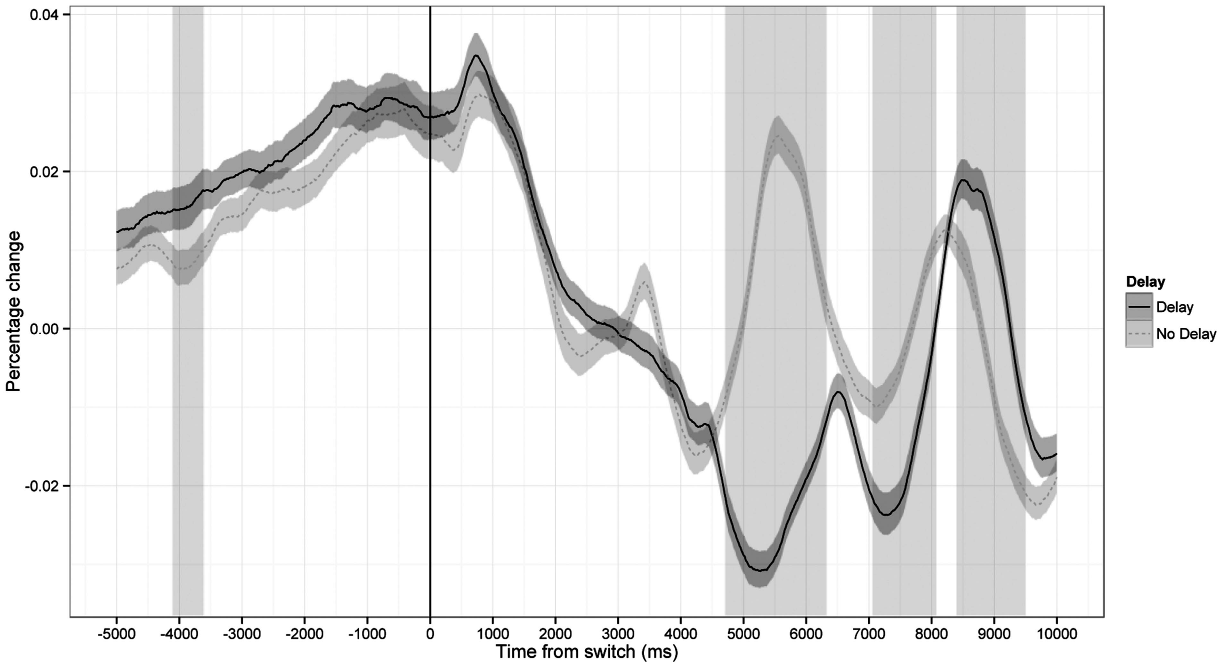

Figure 4. Average pupil dilation for all participants in the voluntary condition for the delay and the no delay conditions of Experiment 1 around the switch (0 is the switching point from memory game to the n-back task). Marked with dark gray color are the moments where there was a statistically significant difference ($p < .05$) between the delay and the no delay conditions. Standard error is shown with a lighter color around the line.

at the moment the interruption was made (assuming a 1-s delay between an event and the pupillary response). We assume that this peak in the forced condition happens because of the actual switching of goals, from primary task goal to secondary task goal. On the other hand, the pupil dilation increase in the voluntary condition is more extensive than the forced condition because it starts gradually increasing some seconds before the switch. We therefore must conclude that this increase includes some other process in addition to the actual switch of goals, which presumably is the decision process leading up to a voluntary switch, given that we have ruled out other explanations.

Is an external interruption (i.e., forced condition) more disruptive than a self-interruption (i.e., the voluntary condition)? In this experiment that would be the logical assumption, because the forced condition was an abrupt external interruption that happened midtask, sometimes while participants were solving an equation. However, the behavioral results revealed not only no significant difference in the time spent on the primary task between the voluntary and the forced conditions, but a trend in the other direction: Participants spent less time (8.12 s on average) to complete the primary task in the forced condition than in the voluntary condition (see Table 1). Why isn't a midtask interruption more disruptive than a posttask interruption? A possible explanation is that the decision process of switching tasks itself has a cost that is substantial enough to overcome the costs of a forced midtask interruption.

There was no statistically significant difference in the resumption lag between the delay and the no delay in the voluntary condition. The resumption lag is an indicator of how disruptive an interruption is (e.g., Trafton et al., 2003). The fact that there was no difference between these two conditions indicates that a delay before the secondary task started did not make the interruption less disruptive—which is surprising, given previous results (Hodgetts & Jones, 2003; Trafton et al., 2003).

Although the results of the experiment were clear, there were additional differences (on top of self-interruptions vs. external interruptions) between the forced and the voluntary conditions. Switching in the forced condition was different than switching in the voluntary condition, because participants clicked to switch in the voluntary condition, choosing a low-workload moment, whereas in the forced condition the switch occurred while they were looking at an equation, at a high-workload moment. Furthermore, participants decided to switch mostly after a match in the voluntary condition, while in the forced condition the switching moments were picked completely randomly and because they appeared while they were looking at a card, there was technically no switch that happened immediately after a match. In order to see if a self-interruption is more disruptive than an external interruption, as the behavioral data suggest, both interruptions should happen at a low-workload moment (because participants choose to self-interrupt themselves at low-workload moments). Also, it was impossible to calculate the resumption lag in the forced condition with the current experimental setup. For these reasons, we conducted a second experiment, in which the forced condition was as similar to the voluntary condition as possible.

Experiment 2

The second experiment was very similar to Experiment 1. We used the same tasks and the same conditions: delay/no delay and

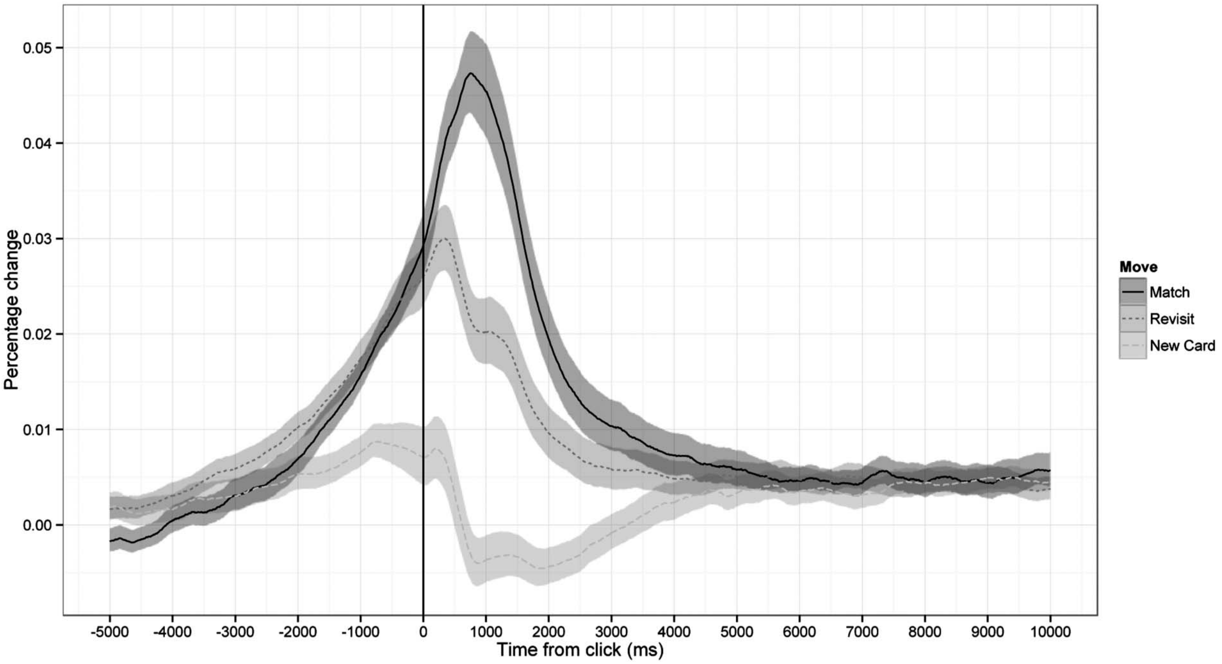

Figure 5. Average pupil dilation for all participants of Experiment 1 in all conditions graphed according to what kind of move they make after 0 (0 is the time of clicking on a card). Standard error is shown with a lighter color around the line.

voluntary/forced. However, the forced condition was different from Experiment 1. In Experiment 2, the switches to the n-back task in the forced condition happened at the moment participants clicked on a card to open it, whereas in the Experiment 1 the switches happened while a card was already open and participants had started solving the equation.

In addition, we wanted the forced condition to be the as similar to the voluntary condition as possible with regard to the frequency of switches after each type of card click. Therefore, we attempted to make a forced block mirror the previous voluntary block. For instance, if in a voluntary block a participant chose to switch once after clicking on a new card and two times after a match, in the following forced block there would be one switch after opening a new card and two switches after a match. However, completely mirroring all blocks turned out to be very difficult. As we saw in Experiment 1, participants prefer to switch after a match. There are only seven matches in a block that could be a switching point (the eighth match finishes the game). It was difficult to make the switches in the forced condition happen at random moments, having three matches before the block ends and two or even three of them to happen after a match. Therefore, we had two “safe points” inserted in the forced condition, that is, two moments were a switch would occur by default in order to make the forced condition more similar to the voluntary condition.

The first safe point was the second-to-last (seventh) match: If there had not been three switches by that point, a switch occurred after the seventh match. That safe point ensured that there were going to be three switches per block. Because usually the switches happened after a match, we picked the seventh match (last match after which a switch could occur) as a safe point. A similar safe point was placed after the second-to-last (sixth) match. A switch occurred there if by the time the participant got to that point there was still one switch that should happen after a match, in order to mirror the pattern of the previous voluntary block.

Method

Participants. Twenty-five participants (17 females, mean age 21.04) participated in this experiment. They all had normal or corrected-to-normal vision and received monetary compensation of 10 euros for participating. Three participants were removed, as they did not seem to solve the equations, but clicked randomly (see below for details). The remaining 22 participants (15 females) had a mean age of 21.18.

Apparatus and setup. Identical to Experiment 1.

Procedure. Same as in Experiment 1, except that the switches in the forced condition happened when participants clicked to open a card and not while looking at a card and solving the equation as in Experiment 1. Furthermore, we tried to make a forced block mirror the previous voluntary block as much as possible by producing the switches in the forced condition after the same kind of move (opening a new card, making a revisit, or making a match) that participants chose to switch in the preceding voluntary block.

Measurement and preprocessing of pupillary data. Identical to Experiment 1. There were no switches with more than 20% of their data interpolated.

Results

Behavioral results. Three participants and 14 blocks from four other participants were removed because they did not pass the threshold of 3 s per card (see Behavioral Results of Experiment 1 for more details). Removing these blocks and participants did neither alter the behavioral nor the pupil dilation results.

The first safe point (after the seventh match) was used only in 2 blocks across all participants. The second safe point (after the sixth match) was used in 64 blocks. There were 58 blocks where no safe points were used.

In this experiment, participants spent significantly less time on a block in the forced condition than in the voluntary condition, spending on average 191.03 s per memory game (time per block minus the time spent on the secondary task) in the forced condition and 210.55 s in the voluntary condition, which was a significant difference of 19.52 s.; $t(21) = -3.53$, $p = .002$, $d = 0.32$.

In the voluntary condition, participants again preferred to switch mostly after a match: 32.48% of the switches were made after opening a new card, 15.38% after revisiting a card, and 52.14% after a match. Switching after a match happened significantly more than the other switch types according to a repeated-measures ANOVA, $F(2, 42) = 7.57$, $p = .002$, $\eta_p^2 = 0.27$. This time the forced condition produced switches mirroring the voluntary condition, with 30.06% of them happening after opening a new card, 9.52% after a revisit, and 60.42% after a match.

Having a 3-s delay before the beginning of the secondary task after a forced interruption again did not improve performance on the primary task. Time per memory game in the forced/delay condition was 198.82 s and in the forced/no delay condition was 182.21 s, with that difference not being significant, $t(21) = 1.86$, $p = .08$, $d = 0.28$.

Analyzing the resumption lag was possible for all conditions in this experiment (last line of Table 2). We removed 58 switches out of 679 as outliers (more than 2 standard deviations from the average). We performed a two-way ANOVA (voluntary/forced and delay/no delay) and results showed that the resumption lag was significantly lower in the forced condition than in the voluntary, $F(1, 20) = 5.4$, $p = .03$, $\eta_p^2 = 0.21$. There was no significant difference between the delay and the no delay conditions and no interaction.[2]

Pupil dilation results. A two-way ANOVA (voluntary/forced and delay/no delay) was performed on every sample, followed by an FDR correction over all samples to correct for multiple comparisons (Benjamini & Hochberg, 1995). There was no interaction between conditions. In Figure 6 the results of the forced and voluntary conditions are shown. As in Experiment 1, the pupil dilation increased significantly more in the voluntary condition before the switch was made and significantly more in the forced condition at the moment the switch was made.

In Figure 7 we plotted the delay/voluntary and no delay/voluntary conditions and where there is statistically significant ($p < .05$) difference between them. As in Experiment 1, there was no difference in pupil dilation increase around the switch and there was a difference after the n-back task started. There was a significant difference 3–4 s before the switch, but as we will explain later, we believe this is coincidental.

Discussion

There was a difference in the increase of pupil dilation between the forced and the voluntary conditions before the switch (see Figure 6), confirming our conclusions of Experiment 1. Although in both conditions the switch happened after a click, it was obvious that there is greater increase in pupil dilation in the voluntary condition before the switch. There was also a peak when the switch was made in the forced condition, whereas pupil dilation in the voluntary condition has already started decreasing by that moment. These results once more indicate that the decision to switch creates an increase in pupil dilation in the voluntary condition. The peak in the forced condition just after the switch probably occurred because participants had to switch suddenly from the primary task goal to the secondary task goal, also confirming the results of Experiment 1.

An obvious difference with the results of Experiment 1 is that the pupil dilation in the forced condition increased leading up to the switch, whereas there was no such effect in Experiment 1 (cf. Figure 3 and Figure 6). The difference between the forced conditions in the two experiments was that in Experiment 2 the forced switches were linked to clicking on a card (to make it more similar to the voluntary condition), while these switches occurred at random moments in Experiment 1—typically when participants were looking at a card. As Figure 5[3] demonstrates, clicking on a card also resulted in an increase in pupil dilation, especially when making a match. Given that most of the switches were made after a match, this effect probably explains the difference between the experiments. However, even though pupil dilation now also increased leading up to the switch in the forced condition, the increase was still significantly larger in the voluntary condition (starting at 2 s before the switch), confirming the main result of Experiment 1.

Figure 7 showed a significant difference between delay/voluntary and no delay/voluntary conditions 3–4 s before the switch, which was not present in the first experiment. These conditions were identical in both experiments (the difference of the two experiments is only in the forced condition), and there is therefore no reason to expect a difference between the conditions in this time window. We therefore think that this difference is coincidental. Apart from this, Figure 7 replicates the results of Experiment 1 and indicates once more that it is not the suspension of the current goal or the preparation of the next task that creates the increase in pupil dilation.

In this experiment, participants in the forced condition were interrupted on a low-workload moment, as they chose to do in the voluntary condition. They were not abruptly interrupted midtask as in Experiment 1, which minimized the switching costs and made both conditions more similar. Behavioral results (see Table 2) showed that in this experiment in contrast to Experiment 1, participants were significantly faster in the forced condition than in the voluntary condition. A part of this difference can be attributed to the decreased resumption lag, which was also smaller in the forced condition than in the voluntary condition, indicating that it

---

[2] One participant did not have any voluntary/delay blocks (they were removed because of random clicking). Therefore, we removed that participant in order to calculate the interaction.

[3] Experiment 2 had very similar pupil dilation results for clicking on a card.

Table 2
Behavioral Data of Experiment 2

		Condition			
	Average	Forced	Voluntary	Delay	No delay
Time per block (s)	269.73 (12.16)	259.22 (11.44)	280.35 (13.42)	258.21 (10.19)	265.62 (13.55)
Time per memory game (n-back time removed; s)	200.93 (12.96)	191.03 (11.66)	210.54 (14.56)	199.92 (12.62)	201.57 (14.13)
# of revisits per block	14.37 (1.99)	13.74 (1.98)	14.88 (2.21)	14.64 (2.17)	14.09 (2.07)
No. (x4) of lucky matches per block	1.26 (0.06)	1.35 (0.07)	1.17 (0.08)	1.26 (0.08)	1.25 (0.08)
# of total clicks to complete a block	37.11 (2.03)	36.39 (2.01)	37.71 (2.26)	37.38 (2.2)	36.84 (2.11)
# of switches to n-back per block	2.77 (0.07)	2.75 (0.07)	2.82 (0.09)	2.8 (0.09)	2.75 (0.08)
Resumption lag (s)	2.46 (0.12)	2.34 (0.12)	2.59 (0.15)	2.41 (0.14)	2.48 (0.12)

Note. Standard error in parenthesis.

was easier to resume the primary task after an external interruption than a self-interruption.

Panepinto (2010) found no difference in performance between a self-interruption and an external interruption and McFarlane's (2002) results suggest that an external interruption is worse than a self-interruption. The fact that self-interruptions turned out to be more disruptive is therefore unexpected. It is logical to expect that it would be more efficient to have control over the interruption than being interrupted at random moments. A possible explanation for this result is that decision making, in addition to creating an increase in pupil dilation, also has time costs. In the current experiment, the decision to make a self-interruption includes deliberate planning of switches, that is, thinking of things like "How many times more do I have to switch to have three switches before the block ends? Should I switch now or open one more card? Will it be an extra new card that I will have to remember before I switch or will it be a match?" In the forced condition participants could just focus on the primary task and do the secondary task when interrupted. They did not have to worry about planning their multitasking behavior in the most efficient way.

One of the goals of the current study was to give participants freedom of choice to switch tasks. Both Experiment 1 and Experiment 2 gave the participants freedom to switch whenever they chose. However, we believe that if they were completely free, they would rarely choose to switch to the n-back task, because that would only distract them from the memory game and make the experiment last longer (the experiment finished after 12 blocks were completed). For that reason, we instructed them to switch at least three times during a block and gave them a 30-s delay penalty at the end of the block if they switched less, which made the

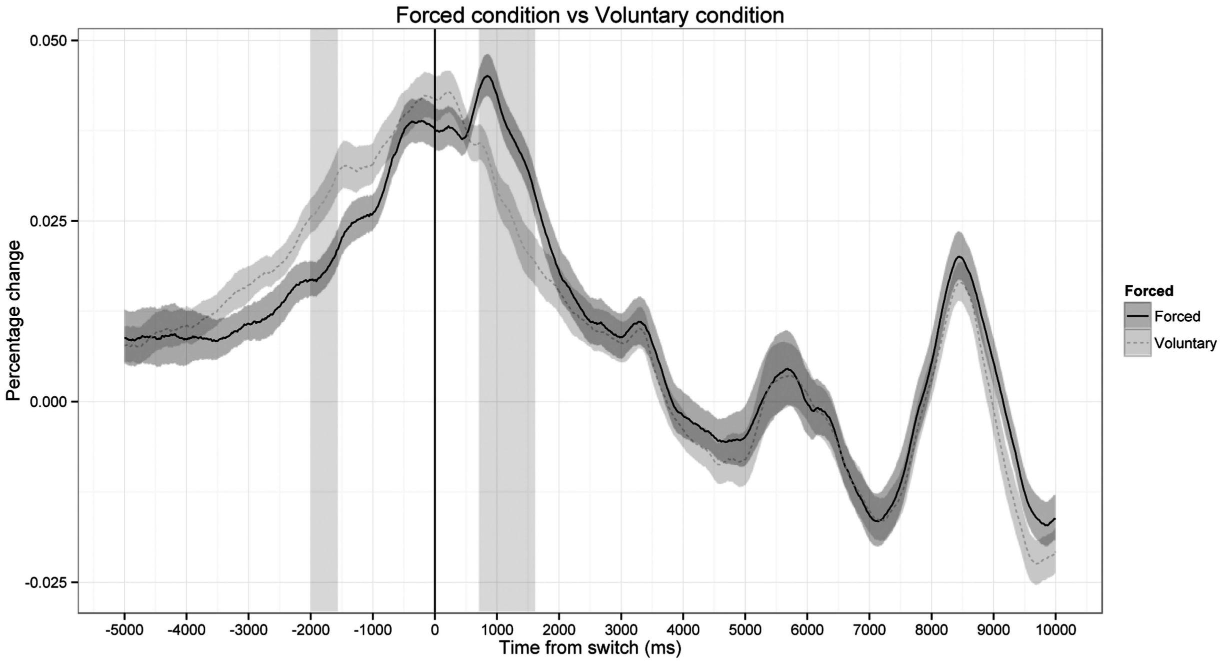

Figure 6. Average pupil dilation for all participants of Experiment 2 for the forced and voluntary conditions around the switch (0 is the switching point from memory game to the n-back task). Marked with dark gray color are the moments where there was a statistically significant difference ($p < .05$) between the forced and the voluntary conditions.

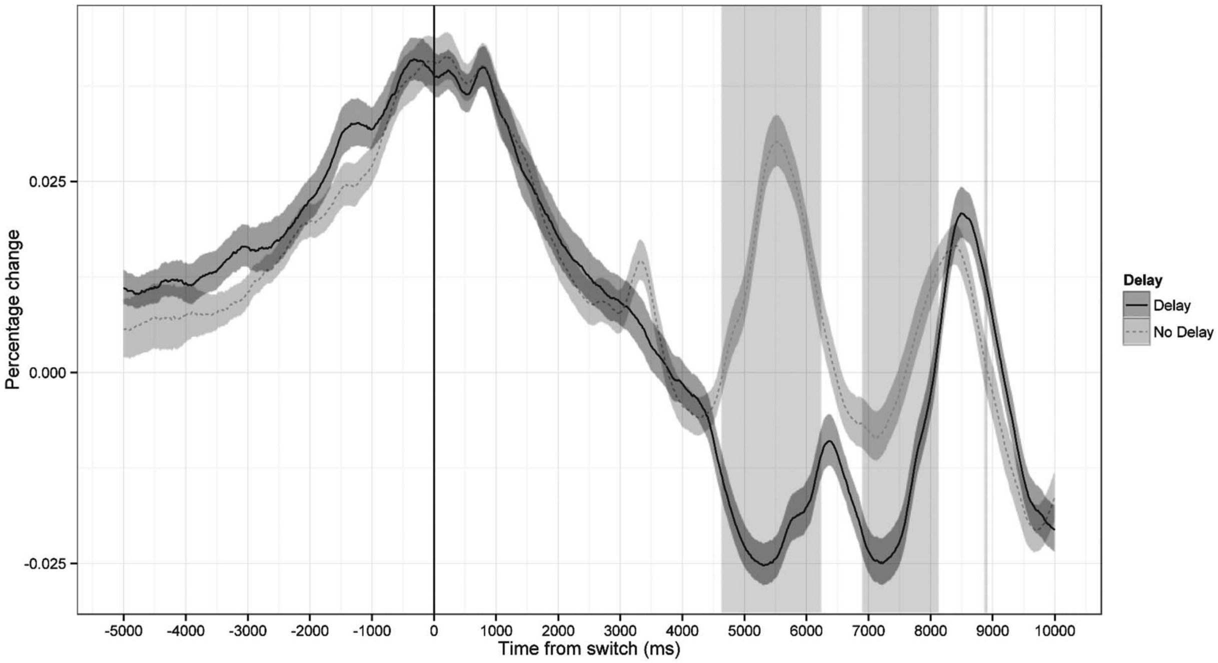

Figure 7. Average pupil dilation for all participants in the voluntary condition of the second experiment the delay and the no delay conditions around the switch (0 is the switching point from memory game to the n-back task). Marked with dark gray color are the moments where there was a statistically significant difference ($p <$ .05) between the delay and the no delay conditions.

voluntary condition not completely voluntary and also included the deliberate planning of switching mentioned in the previous paragraph. That deliberate planning could be partly responsible for the increase in pupil dilation seen in the voluntary condition (Figures 3 and 6) and for the unexpected time difference in memory game completion that is in favor of the forced condition. We were interested to see what would happen if participants had a completely free choice of switching tasks, which might also minimize the deliberate planning of switches. Would there be a time difference with the previous, restricted voluntary condition? To test this we conducted Experiment 3.

Experiment 3

In this experiment we tried to give participants complete freedom in switching to the secondary task. Because it would be hard to tempt them to switch to a task like n-back, we used a different, more "fun" secondary task. The new secondary task was a music quiz. The participants listened to 8 s of a song (the chorus) and were then given four options for either the artist's name or the song title and chose their answer by pressing 1, 2, 3, or 4. After the music fragment ended, participants had an additional 2 s to give their answer (the answer could also be given before the song was finished). Feedback was given by circling the correct answer: a green circle for correct answers, a red circle for incorrect responses. If there was no answer, the correct answer was circled with an orange circle. After the 2 s, a second music fragment played for 8 s plus 2 extra s for the answer and then they returned automatically to the memory game. Their score (percentage of correct answers) was shown on the top right corner of the screen while they were doing the memory game and was updated with the new results after every switch to the music quiz. We chose this task because it uses declarative memory, which is also being used by many office environment tasks.

Although this task was more fun and easy than n-back, participants still usually prefer to make experiments as short as possible. We expected that they would prefer not to switch at all and finish the memory games as quickly as possible. For that reason, Experiment 3 had a fixed duration of (roughly) 45 min, no matter how many blocks were completed. Participants were informed that the experiment would last 45 min and they could spend this time as they preferred: They could not switch to the music task at all if they preferred the memory game or they could switch more often (there was a limit of six switches per memory game but it was never reached).

To see the effects of complete freedom in switching compared with our previous self-interruption setup, we placed half of the participants in the music quiz version and the other half were placed in the n-back version. The n-back version was 45 min of the voluntary/no delay condition of Experiment 1 and Experiment 2 and participants were again instructed to switch three times to the n-back task within a block. The music version of this experiment is completely voluntary and therefore should not result in any deliberate planning of when to switch. The n-back version of this experiment includes the planning of making three switches before the block ends. We wanted to see if the results of Experiment 1 and 2 can be replicated with a different secondary task and an environment where switching is completely voluntary.

Method

Participants. Twenty-five participants (11 female, mean age 22.12) participated in the music quiz version. Two participants were removed, as they did not seem to solve the equations, but clicked randomly (see below for details). The remaining 23 participants (10 females) had a mean age of 22.13.

Twenty-five participants (nine female, mean age 22.08) participated in the n-back version. Two participants were removed, as they did not seem to solve the equations, but clicked randomly. The remaining 23 participants (eight females) had a mean age of 22.22.

All participants had normal or corrected-to-normal vision and received monetary compensation of 10 euros for participating.

Apparatus and setup. Identical to Experiment 1 and Experiment 2.

Procedure. As explained at the beginning of this section.

Measurement and preprocessing of pupillary data. Identical to Experiment 1 and Experiment 2. Removing switches that had more than 20% of their data interpolated resulted in removing 3% of the switches in the music quiz version and 0.35% of the switches in the n-back version.

Results

Behavioral results. Two participants and 19 blocks from six other participants were removed from the music quiz version because they did not pass the threshold of 3 s per card. Two participants and 10 blocks from four other participants were removed from the n-back version (see Behavioral Results of Experiment 1 for more details). Removing these blocks and participants did neither alter the behavioral results nor the pupil dilation results.

Table 3 reports the behavioral data. In the music quiz version, participants spent on average 171.65 s per memory game. They completed on average 11.3 blocks (including in this analysis the 19 blocks that were removed from six participants but not the two participants that were completely removed) in approximately 45 min. They switched to the secondary task on average 1.49 times per block. In the n-back version, participants spent on average 197.96 seconds per memory game. They completed on average 9.35 blocks (including in this analysis the 10 blocks that were removed from four participants but not the two participants that were completely removed) in approximately 45 min. They switched to the secondary task on average 3.21 times per block.

Participants again showed the same preference in switching points, with the switches after a match reaching the percentages of 57.55% (23.08% after a new card and 19.37% after a revisit) in the music quiz version and 75.96% (12.64% after a new card and 11.4% after a revisit) in the n-back version. However, that difference is not statistically significant for the music quiz version: A repeated-measure ANOVA showed that these differences approached significance, $F(2, 44) = 2.99$, $p = .06$, $\eta_p^2 = 0.12$. In the n-back version, switches after a match were still significantly more, as was indicated by a repeated-measure ANOVA, $F(2, 44) = 70.1$, $p < .0001$, $\eta_p^2 = 0.76$.

Participants made significantly fewer switches in the music version compared with the n-back version, $t(31.95) = -8.7$, $p < .001$, $d = 2.56$. That resulted in a significant difference in the number of blocks they completed, with more blocks being completed in the music quiz version (average 11.3 blocks) than in the n-back version (average 9.35). Switching less made the time per block significantly faster in the music quiz version, $t(37.9) = -4.04$, $p < .001$, $d = 1.19$. However, the effect disappeared when removing time spent on the secondary task, $t(36.87) = -1.95$, $p = .059$, $d = 0.57$. Resumption lag was not significantly different between the two conditions, with $t(42.23) = 0.19$, $p = .85$ and $d = 0.06$.

Pupil dilation results. We performed *t* tests between the music quiz and the n-back at every sample (10 ms), with an FDR correction. The results are shown in Figure 8. In both versions there is an increase in pupil dilation before the switch, significantly higher for the n-back version than for the music quiz version. After the switch, the pupil dilation in the n-back version followed the same patterns as in the previous experiments, and the music quiz created a greater increase in pupil dilation than the n-back.

Discussion

Although the music quiz was a more "fun" task than the n-back, the fact that participants were completely free to switch as many times as they wanted made them switch much less than in the n-back version, where they were instructed to switch three times within a block. Although doing the music quiz was more fun than solving equations for most people, participants more often preferred to stick to their primary task and not interrupt themselves. That had positive results, because the difference in time spent in the memory game was marginally significant in favor of the music quiz version.

Making on average 1.49 switches per block can explain why this time the switches after a match only approached significance. It was common in all three experiments that the first switch happened in the beginning of the block, especially in the first blocks.

Table 3
Behavioral Data of Experiment 3

	Music quiz version	n-back version
Average time per block (s)	202.32 (11.87)	289.97 (18.14)
Average time per memory game (secondary task time removed; s)	171.92 (46.45)	197.36 (61.95)
Average No. (x4) of revisits per block	13.55 (1.51)	17.2 (2.2)
Average # of lucky matches per block	1.21 (0.08)	1.26 (0.08)
Average # of total clicks to complete a block	36.33 (1.5)	39.95 (2.23)
Average # of switches to the secondary task per block	1.49 (0.18)	3.21 (0.09)
Resumption lag (s)	1.93 (0.08)	1.91 (0.07)

Note. Standard errors in parentheses.

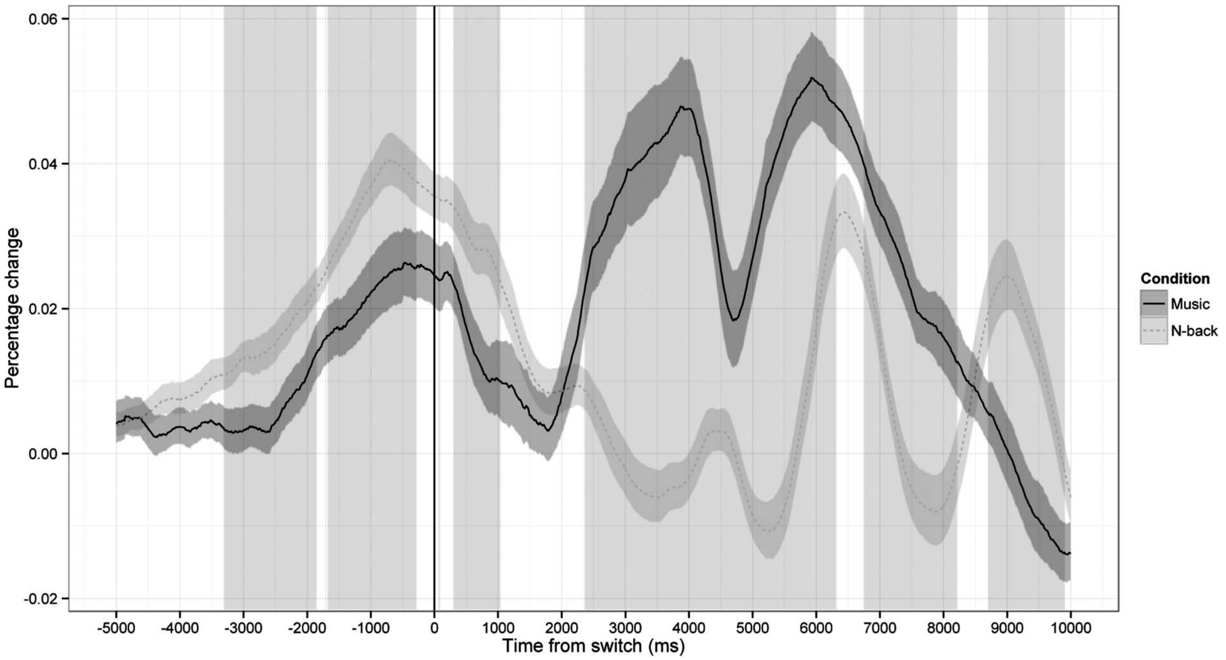

Figure 8. Average pupil dilation for all participants of the third experiment around the switch (0 is the switching point from memory game to the music quiz or the n-back task). Marked with dark gray color are the moments where there was a statistically significant difference ($p < .05$) between the music quiz version and the n-back version.

Because they often performed only a single switch per block, switching after a new card was more common (23% of the switches happened after a new card).

In the previous experiments and the n-back version of this experiment participants were instructed to switch to the n-back task three times within a block. This instruction might have resulted in a complicated process of mentally arranging the switching points within a block. Such a process could have been the explanation for the increase in pupil dilation before a voluntary switch in Experiments 1 and 2. However, participants had no limitations on switching in the music quiz version of Experiment 3. They could switch as many times as they wanted, even not switch at all. Results showed that the average number of switches per block was less than two, indicating that in many cases there was only one switch per block. Therefore, there was no deliberate planning of arranging when to switch. The fact that there was still an increase in pupil dilation before the switching point in this version of the experiment (see Figure 8) indicates that this increase was not only due to mentally planning the switch points.

Figure 8 also shows that pupil dilation before the switch was significantly greater for participants in the n-back task version of the experiment. This effect might indicate that trying to find the best multitasking strategy can create extra cognitive load—in the n-back task participants still had to switch three times. Another factor might be that the switches occurred later in the blocks in the n-back version, and therefore more often following a match. Given that matches are related to higher pupil dilation (see Figure 5), this might have resulted in greater pupil dilation on average in the n-back condition before a switch.

After the switch pupil dilation becomes immediately much larger on the music quiz version, which reflects participants listening to the music fragment. The second peak (starting at 4 s after the switch) probably reflects their answer or their decision making (deciding the correct answer). Therefore listening to music also creates an increase in pupil dilation and much greater than the increase created by doing a demanding task like n-back.

General Discussion

We conducted three multitasking experiments using pupil dilation as a psychometric measure in order to create a time-course of interruptions and to investigate the difference between self-interruptions and external interruptions. The results of Experiment 1 showed an increase in pupil dilation some seconds before a self-interruption, which only appeared at the interruption moment during an external interruption (see Figure 3). In Experiment 2 we refined the experimental setup in order to make the two kinds of interruptions more similar to each other. Pupil dilation results showed the same pattern as in Experiment 1 (see Figure 6) and, in addition, behavioral results showed that when self-interrupted, participants were slower in the primary task than when they were externally interrupted. In Experiment 3 we used a different secondary task and gave participants complete freedom in switching (as opposed to being instructed to switch three times within a block), in order to see if the increase in pupil dilation happened because of deliberate planning of multitasking behavior. There was again an increase in pupil dilation before the self-interruption (see Figure 8), showing that the deliberate planning is not the only reason behind the increase in pupil dilation before a switch.

From these three experiments we can conclude that self-interruptions produce a strong dilation of the pupil, which indicates that in the present experimental setup the decision process to switch tasks takes several seconds. Alternative explanations like pressing the key, rehearsal of the interrupted task or preparation for the interrupting task could be ruled out by the pupil dilation results of Experiment 1 and Experiment 2.

The decision to switch tasks in the voluntary conditions of Experiment 1 and Experiment 2 includes planning of multitasking behavior, that is, when to switch in order to fit three switches in a block while following all the requirements of the experiment (mainly not to make two consecutive switches). We tried to minimize the effect of this planning process by giving participants complete freedom to switch or not in the music quiz version of Experiment 3. It is obvious from Figure 8 that there is still an increase in pupil dilation before the switch, reflecting the decision to switch tasks without the possible extra addition of the deliberate planning of switches.

Although our primary interest in this study was the reaction of the pupil when switching tasks, there were also some interesting and surprising behavioral results. Participants preferred to switch tasks after making a match in the game, which was the moment when the number of items that were currently in their working memory decreased by two. This is a sign of rational multitasking behavior, because switching after a match is less disruptive due to decreasing of items in the working memory (see also Payne et al., 2007 and Salvucci & Bogunovich, 2010, in which participants preferred to self-interrupt after completing a subtask).

The most interesting behavioral result was that self-interruption (voluntary condition) on a low-workload moment turned out to be more disruptive than an external interruption (forced condition) on a low-workload moment (see Table 2). This was reflected in a greater resumption lag and more time spent in the primary task in the voluntary than in the forced condition. Although there is not much research that shows the opposite, one would expect that an unexpected interruption would be more disruptive than a self-interruption. The most relevant study is the one by McFarlane (2002) on external interruptions, where the immediate condition (in which participants were forced to switch tasks instantly on random moments) was the most detrimental to performance and the negotiated condition (which allowed participants to decide when to attend to the interrupting task and is similar to a self-interruption) was the best method to handle interruptions. Our results contradict these results, with participants being slower when they were given the freedom of choice over switching than when they were externally interrupted.

One difference between the two experiments is that the immediate condition in McFarlane's (2002) experiment interrupted participants at random moments, whereas in our Experiment 2 participants in the forced condition where interrupted at low-workload moments. However, that cannot explain the difference, because in the forced condition of Experiment 1 participants were interrupted on high-workload moments, without that making them slower than when they chose themselves when to switch tasks in the voluntary condition. This contradicts many studies (e.g., Iqbal & Bailey, 2005; Katidioti & Taatgen, 2014; Monk et al., 2004), which have suggested that a midsubtask interruption is more disruptive than an interruption after a subtask is finished (which is when participants chose to switch in the voluntary condition). Both Experiment 1 and Experiment 2 suggest that self-interruption is more disruptive than external interruption. An explanation for the contradictory results is that several factors play a role in the costs of interruption. As noted in the beginning of the article, the interruption lag and the resumption lag are considered the main indicators of these costs. Our experiments show there is a third source of costs: the decision to switch, which can take several seconds. In the voluntary conditions, subjects incur this extra cost. However, an appropriate choice of interruption point can reduce the interruption and resumption lag. In Experiment 1, the resumption lag was probably decreased because there were no interruptions in the middle of solving an equation, but this still did not result in a net benefit. In Experiment 2 the forced conditions mirrored the "good" decisions subjects made, so there was no benefit of voluntary choice, only the costs.

Although having a delay between the start of the interruption and the beginning of the secondary task is known to have positive effects on task performance (e.g., Hodgetts & Jones, 2003; Trafton et al., 2003), our results did not confirm this. Participants were not faster in the forced/delay than in the forced/no delay condition in both Experiment 1 and 2, even though the delay conditions included a 3-s lag between the start of the interruption and the secondary task. The resumption lag in the forced/delay condition of Experiment 2 was also not shorter than in the forced/no delay condition, contradicting the results of Hodgetts and Jones (2003). The delay also did not help for self-interruptions, because there were no significant differences either in time or resumption lag between the delay and no delay versions of the voluntary condition in both Experiment 1 and Experiment 2.

In memory for goals theory (Altmann & Trafton, 2002), rehearsing the primary goal during the performance of the secondary task helps minimize the resumption lag. We used a 2-back as a secondary task, which minimizes rehearsing (Monk et al., 2008). However, in the delay condition, participants had 3 s of idle time in which they could have rehearsed the primary task. If they did rehearse, it did not result in improvement on their performance, because they were not faster or had a shorter resumption lag in the delay relative to the no delay condition in both Experiments 1 and 2. In addition, we did not find any indication in the pupil dilation signal that they rehearsed, especially given that it is well known that memory retrieval results in a dilated pupil (e.g., Van Rijn et al., 2012). What this suggests is that the processes that Altmann and Trafton (2002) suggest may be specific to the nature of the primary and interrupting task, and are therefore part of a multitasking strategy as opposed to an automatic response of the cognitive system. In other words, while strategic rehearsal of the main task may occur in different tasks, there is no evidence for it in this experiment.

Interruptions are a serious problem for office workers, students, and generally people working with a computer (e.g., Gonzalez & Mark, 2004). Although there is considerable research on external interruptions (e.g., Hodgetts & Jones, 2003, 2006; Iqbal, Zheng, & Bailey, 2004, Iqbal, Adamczyk, Zheng & Bailey, 2005; Iqbal & Bailey, 2005; Monk et al., 2004, 2008; Trafton et al., 2003), there is limited research on the other half of the interruptions, the self-interruptions (Gonzalez & Mark, 2004; Mark, Gonzalez, & Harris, 2005).

Our results show that self-interruption introduces the extra cost of decision. If these extra costs do not lead to a substantive reduction in the other costs of interruption, self-interruptions are more harmful than external interruptions. The fact that a decision to switch takes several seconds also means that such a decision can be influenced by external influences, or by changes in multitasking strategy.

This has several practical implications. Given a work situation in which interruptions are undesirable, the environment can be modified to influence the interruption decision process, for example by removing visual information from the screen that may cue self-interruption (e.g., incoming mail flags, chat windows, etc.), or to introduce artificial switch costs (e.g., requiring people to explicitly leave their current application instead of allowing easy back-and-forth switching). Additionally, it may be possible to train people on strategies to better deal with interruptions. In work situations where interruptions are unavoidable, it may be better to externalize the decision process to switch instead of leaving it up to the user. If possible in the task setting, switching tasks can be delegated to a dedicated scheduler, thereby taking the costs of decision away from the user.

Several existing applications or methods already implicitly support these guidelines (e.g., Iqbal & Bailey, 2006; McFarlane, 2002). Recent updates of operating systems allow applications to occupy the full screen, thereby removing interface elements that can prompt a self-interruption. Several applications allow people to block the Internet for a particular period of time (e.g., freedom[4]), thereby implicitly scheduling the next self-interruption when that period ends. Finally, methods can be developed to detect the switch decision, for example by monitoring the pupil size, possibly supported by other measurements (cf. Iqbal & Bailey, 2005).

---

[4] http://macfreedom.com/

References

Allport, A., & Wylie, G. (2000). Task-switching, stimulus–response binding, and negative priming. In S. & Monsell, J. S. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes* (pp. 35–70). Cambridge, MA: MIT Press.

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science, 26,* 39–83. doi:10.1207/s15516709cog2601_2

Anderson, J. R., Fincham, J. M., Schneider, D. W., & Yang, J. (2012). Using brain imaging to track problem-solving in a complex state space. *NeuroImage, 60,* 633–643.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 142–162). Cambridge, MA: Cambridge University Press.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B, 57,* 289–300.

Borst, J. P., Buwalda, T. A., Van Rijn, H., & Taatgen, N. A. (2013). Avoiding the problem state bottleneck by strategic use of the environment. *Acta Psychologica, 144,* 373–379. doi:10.1016/j.actpsy.2013.07.016

Borst, J. P., Taatgen, N. A., & Van Rijn, H. (2010). The problem state: A cognitive bottleneck in multitasking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 363–382. doi:10.1037/a0018106

Brumby, D. P., Cox, A. L., Back, J., & Gould, S. J. J. (2013). Recovering from an interruption: Investigating speed-accuracy tradeoffs in task resumption strategy. *Journal of Experimental Psychology: Applied, 19,* 95–107. doi:10.1037/a0032696

Chiew, K. S., & Braver, T. S. (2013). Temporal dynamics of motivation-cognitive control interactions revealed by high-resolution pupillometry. *Frontiers in Psychology, 4,* 15.

Chisholm, C. D., Collison, E. K., Nelson, D. R., & Cordell, W. H. (2000). Emergency department workplace interruptions: Are emergency physicians "interrupt-driven" and "multitasking"? *Academic Emergency Medicine, 7,* 1239–1243. doi:10.1111/j.1553-2712.2000.tb00469.x

Cleveland, W. S. (1981). LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician, 35,* 54. doi:10.2307/2683591

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. In E. Dykstra-Erickson & M. Tscheligi (Eds.), *CHI 2004 Proceedings* (pp. 175–182). New York: ACM Press. doi:10.1145/985692.985715

Dabbish, L., Mark, G., & Gonzalez, V. (2011). Why do I keep interrupting myself?: Self interruption, habit, and environment. In *CHI 2011 Proceeding* (pp. 3127–3130), New York: ACM Press. doi:10.1145/1978942.1979405

Gonzalez, V., & Mark, G. (2004). Constant, constant, multi-tasking craziness: Managing multiple working spheres. In E. Dykstra-Erickson & M. Tscheligi (Eds.), *CHI 2004 Proceedings* (pp. 113–120). New York: ACM Press.

Gould, S. J. J., Cox, A. L., & Brumby, D. P. (2013). Frequency and duration of self-initiated task- switching in an online investigation of interrupted performance. In B. Hartmann & E. Horvitz (Eds.), *Proceedings of the AAAI Conference on Human Computation & Crowdsourcing Works-in-Progress*. Palo Alto, CA: The Association for the Advancement of Artificial Intelligence (AAAI).

Hodgetts, H. M., & Jones, D. M. (2003). Interruptions in the Tower of London task: Can preparation minimize disruption? In *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1000–1004). Santa Monica, CA: Human Factors and Ergonomics Society.

Hodgetts, H. M., & Jones, D. M. (2006). Interruption of the Tower of London task: Support for a goal-activation approach. *Journal of Experimental Psychology: General, 135,* 103–115.

Hoeks, B., & Levelt, W. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, 25,* 16–26. doi:10.3758/BF03204445

Iqbal, S. T., Adamczyk, P. D., Zheng, S. X., & Bailey, B. P. (2005). Towards an index of opportunity: Understanding changes in mental workload during task execution. In G.v.d. Veer & C. Gale (Eds.), *CHI 2005 Proceedings* (pp. 311–320). New York, NY: ACM Press. doi:10.1145/1054972.1055016

Iqbal, S. T., & Bailey, B. P. (2005). Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. In G.v.d. Veer & C. Gale (Eds.), *CHI 2005 Proceedings* (pp. 1489–1492). New York: AMC Press doi:10.1145/1056808.1056948

Iqbal, S. T., & Bailey, B. P. (2006). Leveraging characteristics of task structure to predict the cost of interruption. In *CHI 2006 Proceedings* (pp. 741–750). New York, NY: ACM.

Iqbal, S. T., Zheng, X. S., & Bailey, B. P. (2004). Task- evoked pupillary response to mental workload in human-computer interaction. In E. Dykstra-Erickson & M. Tscheligi (Eds.), *CHI 2004 Proceedings* (pp. 1477–1480). New York, NY: ACM Press. doi:10.1145/985921.986094

Jin, J., & Dabbish L. A. (2009). Self-Interruption on the computer: A typology of discretionary task interleaving, In *CHI 2009 Proceedings* (pp. 1799–1808). New York: AMC Press. doi:10.1145/1518701.1518979

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load of memory. *Science, 154*(3756), 1583–1585. doi:10.1126/science.154.3756.1583

Kahneman, D., Tursk, B., Shapiro, D., & Crider, A. (1969). Pupillary, heart rate and skin resistance changes during a mental task. *Journal of Experimental Psychology, 79,* 164–167. doi:10.1037/h0026952

Katidioti, I., & Taatgen, N. A. (2014). Choice in multitasking: How delays in the primary task turn a rational into an irrational multitasker. *Human Factors, 56,* 728–736. doi:10.1177/0018720813504216

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55,* 352–358. doi:10.1037/h0043688

Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2011). Pupillary Stroop effects. *Cognitive Processing, 12,* 13–21. doi:10.1007/s10339-010-0370-z

Mark, G., Gonzalez, V., & Harris, J. (2005). No task left behind? Examining the nature of fragmented work. In G.v.d. Veer & C. Gale (Eds.), *CHI 2005 Proceedings* (pp. 321–330), New York: ACM Press. doi:10.1145/1054972.1055017

McFarlane, D. C. (2002). Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction, 17,* 63–139. doi:10.1207/S15327051HCI1701_2

Monk, C. A., Boehm-Davis, D. A., & Trafton, J. G. (2004). Recovering from interruptions: Implications for driver distraction research. *Human Factors, 46,* 650–663. doi:10.1518/hfes.46.4.650.56816

Monk, C. A., Trafton, J. G., & Boehm-Davis, D. A. (2008). The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology: Applied, 14,* 299–313. doi:10.1037/a0014402

Moresi, S., Adam, J. J., Rijcken, J., Van Gerven, P. W. M., Kuipers, H., & Jolles, J. (2008). Pupil dilation in response preparation. *International Journal of Psychophysiology, 67,* 124–130. doi:10.1016/j.ijpsycho.2007.10.011

Panepinto, M. P. (2010). Voluntary versus forced task switching. In *Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 453–457). Santa Monica, CA: Human Factors and Ergonomics Society.

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,* 370–388. doi:10.1037/0096-3445.136.3.370

Peavler, W. S. (1974). Pupil size, information overload and performance differences. *Psychophysiology, 11,* 559–566. doi:10.1111/j.1469-8986.1974.tb01114.x

Prehn, K., Heekeren, H. R., & Van der Meer, E. (2011). Influence of affective significance on different levels of processing using pupil dilation in an analogical reasoning task. *International Journal of Psychophysiology, 79,* 236–243. doi:10.1016/j.ijpsycho.2010.10.014

Richer, F., & Beatty, J. (1985). Pupillary dilations in movement preparation and execution. *Psychophysiology, 22,* 204–207. doi:10.1111/j.1469-8986.1985.tb01587.x

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124,* 207–231. doi:10.1037/0096-3445.124.2.207

Rosen, L. D., Carrier, L. M., & Cheever, N. A. (2013). Facebook and texting made me do it: Media induced task-switching while studying. *Computers in Human Behavior, 29,* 948–958. doi:10.1016/j.chb.2012.12.001

Salvucci, D. D., & Bogunovich, P. (2010). Multitasking and monotasking: The effects of mental workload on deferred task interruptions. In *CHI 2010 Proceedings* (pp. 85–88), New York: ACM Press. doi:10.1145/1753326.1753340

Salvucci, D. D., Monk, C. A., & Trafton, J. G. (2009). A process-model account of task interruption and resumption: When does encoding of the problem state occur? In *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting* (pp. 799–803). Santa Monica, CA: Human Factors and Ergonomics Society.

Salvucci, D. D., & Taatgen, N. A. (2011). *The multitasking mind*. New York, NY: Oxford University Press.

Satterthwaite, T. D., Green, L., Myerson, J., Parker, J., Ramaratnam, M., & Buckner, R. L. (2007). Dissociable but inter-related systems of cognitive control and reward during decision making: Evidence from pupillometry and event-related fMRI. *NeuroImage, 37,* 1017–1031. doi:10.1016/j.neuroimage.2007.04.066

Steinhauer, S. R., & Hakerem, G. (1992). The pupillary response in cognitive psychophysiology and schizophrenia. *Annals of the New York Academy of Sciences, 658,* 182–204. doi:10.1111/j.1749-6632.1992.tb22845.x

Trafton, J. G., Altmann, E. M., Brock, D. P., & Mintz, F. E. (2003). Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies, 58,* 583–603. doi:10.1016/S1071-5819(03)00023-5

van Rijn, H., Dalenberg, J. R., Borst, J. P., & Sprenger, S. A. (2012). Pupil dilation co-varies with memory strength of individual traces in a delayed response paired-associate task. *PLoS ONE, 7,* e51134. doi:10.1371/journal.pone.0051134

Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 8456–8460. doi:10.1073/pnas.1201858109

Received March 20, 2014
Revision received August 13, 2014
Accepted August 18, 2014 ■