Computers in Human Behavior 63 (2016) 906–915

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Interrupt me: External interruptions are less disruptive than self-interruptions

Ioanna Katidioti*, Jelmer P. Borst, Marieke K. van Vugt, Niels A. Taatgen

University of Groningen, Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 21 December 2015
Received in revised form
14 June 2016
Accepted 17 June 2016

Keywords:
Interruption
Pupil dilation
Multitasking

ABSTRACT

Interruptions are part of everyday life and are known to be disruptive. With the current study we investigated which kind of interruption is more disruptive: external interruptions or self-interruptions. We conducted two experiments, one behavioral experiment (28 participants) and one in which pupil dilation was measured (21 participants). In both experiments, self-interruptions made participants complete the main task slower than external interruptions (occurring at similar moments in the task as the self-interruptions). However, there was no difference between the two kinds of interruptions in the time needed to resume the main task (resumption lag). Instead, the pupil dilation data revealed that the decision to self-interrupt takes about 1 s, resulting in slower performance overall.

1. Introduction

It is hard to imagine a working day without interruptions. Telephones ringing, colleagues walking into the office, and the constant checking of email and social media are part of everyday life for most people. Various observational studies reveal the frequency of interruptions in different kinds of environments: office workers switching tasks every 3 min (Gonzalez & Mark, 2004) and students interrupting their tasks every 6 min mostly to engage in social media (Rosen, Carrier, & Cheever, 2013) are some of the most impressive, but also representative, results. About half of these interruptions are initiated by an external source (external interruptions) and the other half are initiated internally (self-interruption; e.g., Czerwinski, Horvitz, & Wilhite, 2004; Gonzalez & Mark, 2004; Mark, Gonzalez, & Harris, 2005).

Interruptions can be considered as a form of multitasking (Salvucci & Taatgen, 2011), since the interrupted person has to deal with more than one task. The common notion of the term multitasking is usually limited to concurrent multitasking: performing two tasks simultaneously, meaning that the changes from one task to the other are very quick. An example of concurrent multitasking is driving and talking on the phone or driving and texting. Salvucci and Taatgen (2011) claim that what they call sequential multitasking is governed by the same rules as concurrent multitasking, but on a longer time scale. Working on a paper while answering emails and participating in meetings, watching a movie while having to check the food in the oven or even reading two different books at the same time can all be considered sequential multitasking. Salvucci, Taatgen, and Borst (2009) have demonstrated the theoretical similarities of both forms of multitasking by creating cognitive models that use the same underlying mechanisms. In the current study, we focus on interruptions.

Interruptions can be described as follows (Fig. 1): one is engaged in a *main task*, which is interrupted by an *interrupting task*. The time between the moment of the interruption and the beginning of the interrupting task is called the *interruption lag*. After the interrupting task is finished, the main task is resumed. The time from the end of the interrupting task until the resumption on the main task is called *resumption lag*.

This representation of an interruption timeline is accurate for external interruptions. However, we believe that it is incomplete when it comes to self-interruptions. In a previous study, we found an increase in pupil dilation before a self-interruption, which was not present before an external interruption (Katidioti, Borst, & Taatgen, 2014). In addition, there was the unexpected result of external interruptions being less disruptive than self-interruptions. We interpreted this finding by postulating that self-interruptions require a decision or preparation phase before the interruption, which is not the case for external interruptions. In this paper we will perform two experimental studies in order to complete the possible missing pieces of the self-interruption timeline and to find

* Corresponding author. Department of Artificial Intelligence, University of Groningen, 9747 AG Groningen, The Netherlands.
E-mail address: i.katidioti@rug.nl (I. Katidioti).

http://dx.doi.org/10.1016/j.chb.2016.06.037
0747-5632/

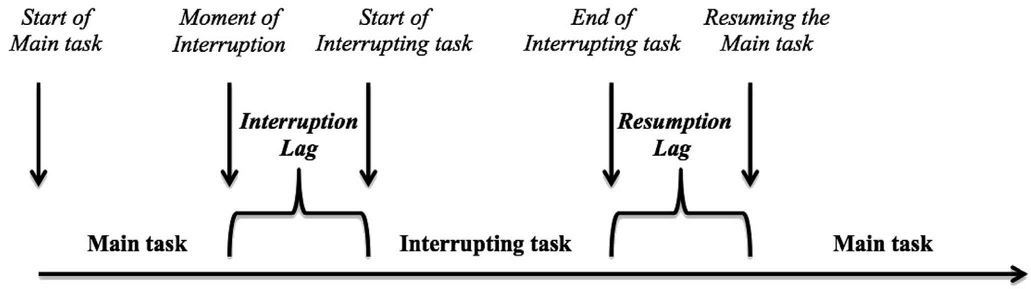


Fig. 1. Timeline of an interruption (based on Trafton et al., 2003).

out which kind of interruptions are less disruptive: self-interruptions or external interruptions.

1.1. The disruptiveness of interruptions

Interruptions typically affect performance on the main task negatively. This negative effect is apparent in several different ways. First, it takes longer to complete the main task when interrupted. Mark et al. (2005) found that 22.7% of the interrupted tasks in an office environment were not even resumed on the same day. Iqbal and Horvitz (2007) showed that, after sending answers to emails that interrupted them, people engaged in other unrelated tasks and ended up resuming their main tasks only after 10–15 min. Even if people return to the main task immediately, there are time costs to interruptions: being interrupted makes people slower on their main tasks (e.g., Allport & Wylie, 2000; Monk, Boehm-Davis, & Trafton, 2008; Trafton, Altmann, Brock, & Mintz, 2003). In addition, errors in the main task are also more frequent when interrupted (e.g., Brumby, Cox, Black & Gould, 2013; McFarlane, 2002). Although it is clear that interruptions affect main task performance, there are several factors that determine the disruptiveness of an interruption.

The Memory for Goals theory (Altmann & Trafton, 2002; also supported by the model of; Salvucci et al., 2009) claims that each task has a goal with an activation level. If a task is interrupted, its goal is stored and starts decaying. Therefore the longer an interruption is, the more the goal of the main task decays and the harder it is to resume it. Two experiments by Monk, Trafton, and Boehm-Davis (2008) suggested that a longer interruption is more disruptive than a short one, resulting in longer resumption lags (see also Borst, Taatgen, & van Rijn, 2015; Hodgetts & Jones, 2003). However, if the interruptions were too long (more than 23 s in the setup Monk and colleagues used), the disruptiveness stopped increasing with time. In addition, more cognitively difficult interruptions are more disruptive than simple interruptions (e.g., Borst et al., 2015; Cades, Boehm-Davis, Trafton, & Monk, 2007; Monk et al., 2008) and an interruption that is relevant to the main task is less disruptive than an unrelated one (e.g. Czerwinski, Cutrell, & Horvitz, 2000; Gould, Brumby, & Cox, 2013).

Not only the properties of the interruption itself determine its disruptiveness, but also the moment of the interruption in the main task is important. Several studies have shown that interruptions on low-workload moments (typically between subtasks) are less disruptive than interruptions on high-workload moments (mid-subtask; e.g. Iqbal & Bailey, 2005; 2006; Katidioti & Taatgen, 2014; Monk, Boehm-Davis, & Trafton, 2004). For instance, interrupting participants on high-workload moments was more disruptive than interrupting them on low-workload moments in the data entry task of Gould et al. (2013), the VCR programming task of Monk et al. (2004), the task in which people have to combine e-mailing and chatting by Katidioti and Taatgen (2014) and the three tasks (video editing, route planning and document editing) that Iqbal and Bailey (2006) used. In some of these tasks, high-workload interruptions meant that participants were interrupted at moments where they had to retain information in their working memory, for example a product name that had to be typed (Katidioti & Taatgen, 2014) or information about a show that had to be recorded (Monk et al., 2004). In these tasks, low-workload moments were considered those where working memory was free.

1.2. External interruptions and self-interruptions

Interruptions can be separated into two kinds: external interruptions, which are initiated by an external source, and self-interruptions, which are initiated internally. A phone ringing or a colleague walking into the office are external interruptions. Deciding to check social media or getting up to go for a walk are self-interruptions. Studies show that both kinds of interruptions are roughly equally frequent. In the study of Mark et al. (2005), 52% of the interruptions in an office environment were self-interruptions and in the study of Czerwinski et al. (2004), 40% of the interruptions were initiated internally.

It is not easy to name the causes of self-interruptions. Some studies suggest that if a task it too easy or too difficult, the person is more likely to self interrupt because of either boredom or frustration (Adler & Benbunan-Fich, 2013). Other studies suggest that people self-interrupt or are distracted when the cognitive resources for the interrupting task are available. Taatgen, Katidioti, Borst, and Van Vugt (2015) have performed an experiment where participants were more distracted by a video when their visual resources became more available, regardless of the difficulty level of the task they were performing. Taatgen et al. created a cognitive model that explains these results. Research on cyberloafing (surfing the Internet during work hours) is also trying to uncover the reasons behind self-interrupting. The most common notion is that cyberloafing is taking a break using a computer (e.g. Blanchard & Henle, 2008; Lim & Teo, 2005). According to the study by Wagner, Barnes, Lim, and Ferris (2012), sleep deprived participants were more likely to cyberloaf. Theory of Planned Behavior (people form intentions before they behave, influenced by social norms and others' view on this behavior, Ajzen, 1991) is also proposed as an explanation of cyberloafing. There are many theories that try to discover the reasons behind self-interrupting, but there is no unified theory yet.

One of the reasons for that is that self-interruptions are still challenging to study in an experimental setup. In order to find the basic mechanisms behind self-interruptions, research should focus on comparing them to external interruptions, which are very well-studied (e.g. Dindar & Akbulut, 2016; Monk et al. 2004; 2008). Although it seems intuitive that self-interruptions are less disruptive than external interruptions – because people are free to choose the moment they will self-interrupt – only few studies have compared the disruptiveness of external interruptions and self-interruptions.

In one of the studies that compared external interruptions and a form of self-interruptions (external interruptions turned to self-interruptions), participants had to play a simple videogame while they were being interrupted by a simple matching task in four different ways: immediate (random external interruptions), mediated (external interruptions occurring on low-workload moments), scheduled (external interruptions occurring every 25 s) or negotiated interruption (McFarlane, 2002). Negotiated interruptions resembled self-interruptions: participants were interrupted for 150 ms by a flashing interrupting task, but could choose when to act on that interruption. McFarlane's results were mostly in favor of the negotiated interruption. However, it is debatable whether the negotiated interruption can indeed be considered a self-interruption, since participants were initially externally interrupted for 150 ms.

In contrast, Mark et al. (2005), Panepinto (2010) both used 'real' self-interruptions, in the sense that there was no external notification. Mark et al. (2005) did not find any difference in their observational study between self-interruptions and external interruptions. Panepinto (2010) conducted a task-switching study, where participants had 30 min to complete a Sudoku and correct a document. Some of the participants were free to choose when to switch between the two tasks and some were being forced to switch. There was no significant difference in performance or reaction times between these two groups. However, it should be noted that the forced task-switching occurred at random moments (which could be high-workload moments), while participants in the self-interruption condition could choose opportune moments to switch.

Finally, Katidioti et al. (2014) performed two experiments comparing self-interruptions with external interruptions in a memory game with mathematical equations. In their first experiment there was no significant difference between the two different kinds of interruptions. However, the external interruptions occurred mid-subtask (while participants were solving a mathematical equation), while the self-interruptions occurred between subtasks. As is known from other studies (e.g. Katidioti & Taatgen, 2014; Salvucci & Bogunovich, 2010), people are rational when they self-interrupt and do so mostly on low-workload moments. As a result, participants self-interrupted on low-workload moments but were externally interrupted on high-workload moments. Given that there was no significant difference in performance in favor of self-interruption (on the contrary, there was a tendency to being faster in the external interruption condition) a possible implication was that external interruptions are less disruptive. To test this, they set up their Experiment 2 so that external interruptions were mirroring the self-interruptions: they occurred at the same low-workload moments. Results showed that participants were faster to complete the main task when they were externally interrupted compared to when they were self-interrupted.

As the goal of Katidioti et al. (2014) was not to compare self-interruptions and external interruptions, their experimental setup was not created to make this comparison and the results have to be interpreted with care. In the current study we therefore compare self-interruptions and external interruptions in a more controlled manner. Based on the results of Katidioti et al. (2014), our hypothesis is that external interruptions are less disruptive than self-interruptions — even though that seems counter-intuitive. In our second experiment of the current study, we used pupil dilation as a psychometric measure, in order to further investigate differences between self-interruptions and external interruptions.

2. Experiment 1

2.1. Methods

2.1.1. Design

The main and the interrupting task of this experiment were based on the ones used by Salvucci and Bogunovich (2010), Katidioti and Taatgen (2014). The experimental setup resembles a client service work environment of an electronics company. The main task was an email-answering task, in which participants had to answer emails that asked about the prices of fictional products, by looking up the information in a simulated internet browser. The interrupting task was a chat-answering task, in which participants had to answer personal questions. The windows used in the experiment are shown in Fig. 2. In the actual experiment, the windows were overlapping and could not be moved. The participant had to click on a window in order to bring it on the foreground and view it. This forced the participants to memorize the information in windows that were not currently visible.

The steps of the main task are shown in Fig. 3. First, the participant opened an email (Mail Select) by clicking on it and reading the question (e.g. "What is the price of laptop Zanium A-63?"). In order to force the participants to memorize the product, we inserted a 3-s loading time for the email, with the intention of discouraging participants from easily returning to the mail window to read it again. After reading the email, the participant had to first click on the simulated browser window in order to bring it in focus (Browser Focus), next click on the Home button if necessary[1] (Browser Home), click on the product category (Link 1), then the product name (Link 2) and finally the product code (Link 3). After a 2-s delay the price of the product loaded, the participant could read it, return to the email window (Mail Focus) and press the "Reply" button. Then the composer window would appear (Composer Focus) and the participant had to type the price, press the "Send" button (Composer Sent), making the composer window disappear, and finally drag and drop the answered email in the Archive folder (Mail Move).

As can be seen Fig. 3, the email task has both low and high-workload moments. High-workload moments are those where the participant has to retain information in working memory, either the product name or the product price. Low workload moments are those where the participant's working memory is free, which includes the moment before reading the product name in the email (Mail Select), after finishing the search and before reading the product price (Link 3) and after typing the email response (Compose Send and Mail Move).

The interrupting task simulated a casual chat conversation. Chat questions were in the form of "What is your favorite … ?" (e.g., color, restaurant, cartoon, book). In order to make the interrupting task more natural and engaging, one in four questions was a follow-up question, asking "Which is your least favorite?", referring to the previous question. The interrupting task is irrelevant to the main task, requires some time and can be interesting for the participant. The difficulty of the interrupting task (open questions versus yes/no questions) did not create any results in previous research (Katidioti & Taatgen, 2014).

2.1.2. Conditions

Two factors were varied in the experiment: the kind of

[1] The "Home" button could be also pressed after participants finished the browser search. However, in the analysis of the interruption moments, we only included the "Browser Home" that occurred before the browser search, which was by far the most common.

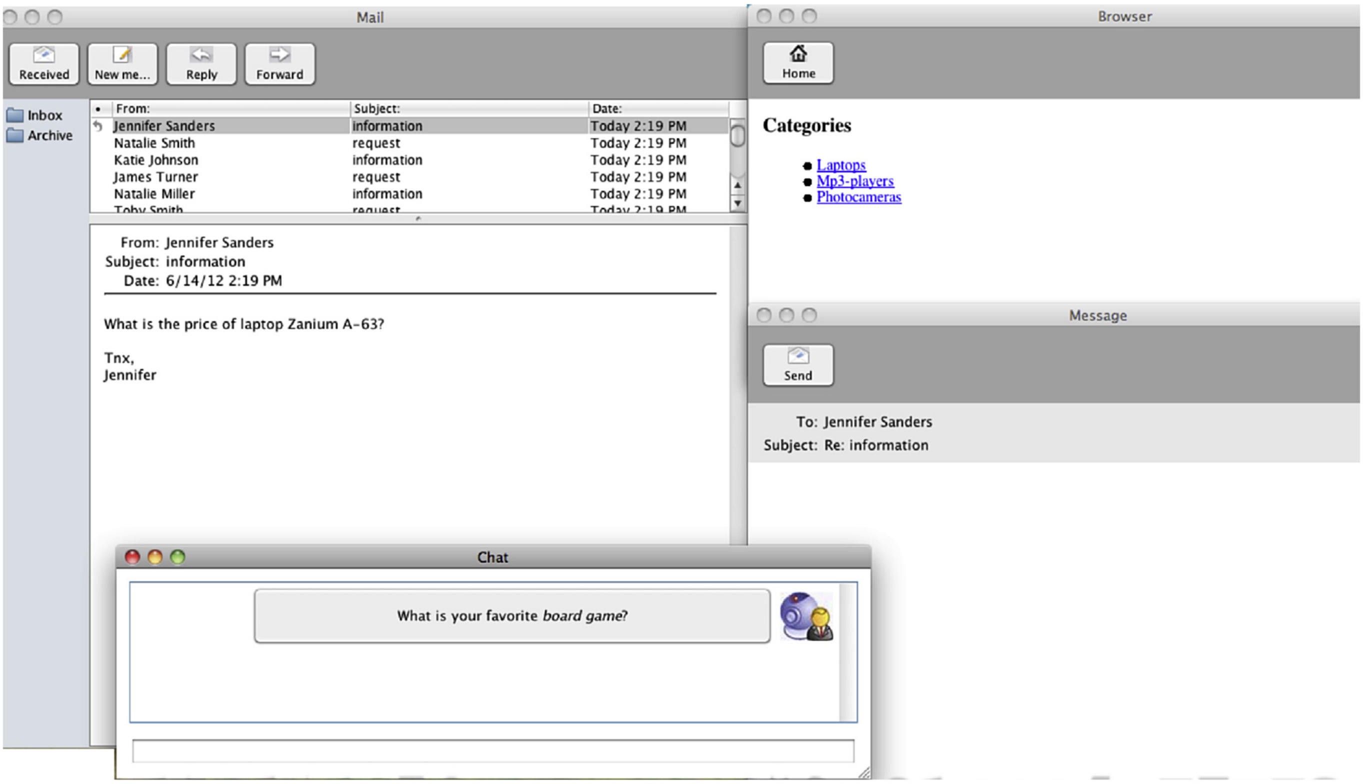


Fig. 2. All the windows used in the experiment. During the experiment, the windows were overlapping.

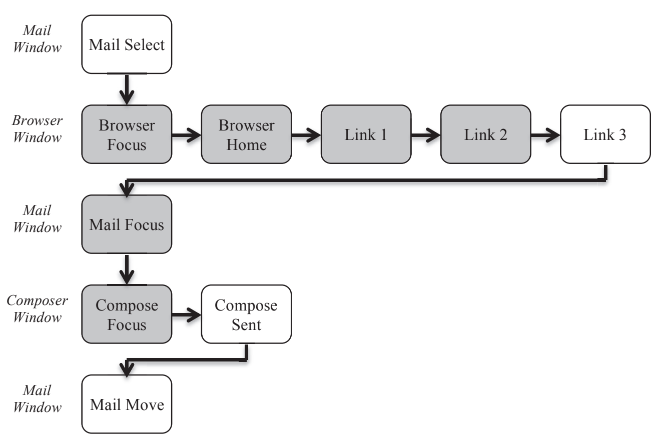


Fig. 3. The sequence of the main task. The high-workload moments are indicated with grey color and the low-workload moments with white.

interruption (Control, Voluntary or Forced) and the presence of a browser delay in the main task (Delay or No Delay condition). Therefore, there were six different conditions in Experiment 1: Control-No Delay, Control-Delay, Voluntary-No Delay, Voluntary-Delay, Forced-No Delay and Forced-Delay. The experiment finished after 12 blocks, two of each kind.

In the Control blocks, participants did not receive interruptions from the chat-answering task, to measure baseline performance on the mail task. In the Voluntary blocks, there was always a chat message waiting in the background and participants could read and answer it by clicking the chat window and bringing it into focus. They were free to choose when to answer a chat message, but in order to entice them to self-interrupt, they were informed at the beginning of the experiment that a Voluntary block ends after 10 emails and 15 chat messages are answered. When a chat message was answered, a new one appeared in the unfocused chat window when the participant resumed the email task. In the Forced blocks, the chat window appeared in the foreground when a chat message arrived and could not be unfocused until participants answered it.

Interruptions in the Forced blocks mirrored the interruptions of the last Voluntary block, taking into account the delay manipulation. Consequently, a Forced-Delay block mirrored the last Voluntary-Delay block and a Forced-No Delay block the last Voluntary-No Delay block. Because of the mirroring, the interruptions in the Forced blocks occurred at the same moments during each email when participants had chosen to self-interrupt in the Voluntary blocks. For example, if a participant chose to answer a chat message after clicking on Link 3 in the third email of the Voluntary-Delay block, the interruption on the third email of the next Forced-Delay block would also happen after clicking on Link 3. In contrast to external interruptions at random times, this ensures that effects on the mail task are not due to interruptions at different moments in the task, for instance high and low-workload moments. It should be noted that interruptions in the Forced blocks occurred during the same moments in the task (Fig. 3) as the ones in the Voluntary blocks, but they were not timed. This means that the participant could click to perform a main task move (e.g. Mail Select, Browser Focus etc.) and then wait 10 s before deciding to self interrupt in the Voluntary block. In contrast, in the Forced block the interruption would occur immediately after the click to perform the main task move.

In the Delay blocks a 3-s loading time was added when participants clicked on Link 1 and Link 2. In the No Delay conditions, Link 1 and Link 2 loaded immediately after clicking on them. The 2-s loading time in Link 3 was present in both Delay and No Delay blocks, in order to give participants some time before the price appears and make it a more clear low-workload moment. We used the delay manipulation in order to have longer high-workload moments. In Katidioti and Taatgen (2014) participants chose to switch on high-workload moments in the Delay blocks. If this is also the case in this experiment, we will be able to compare high-workload external interruptions with high-workload self-interruptions. However, note that the interrupting task is quite different in the current experiment than in Katidioti and Taatgen (2014). There, the chat arrived once or twice during one email sequence, by creating a sound and turning the chat window yellow. Furthermore, in that experiment participants were informed that a block finishes after they answered 24 chat messages, regardless of the number of emails. This kind of environment motivates participants to self-interrupt as soon as possible. In contrast, in the current experiment, participants know that there is always a chat message in the background (without sound or color changes) and that in order for a block to finish, both emails and chat messages must be answered.

The block order was semi-random. Since Forced blocks had to mirror a Voluntary block, the first block would be either Control or Voluntary. After a Voluntary-Delay or Voluntary-No Delay block finished, a Forced-Delay or Forced-No Delay block, respectively, could also be chosen randomly. When all six different kinds of blocks were finished, all block-types were presented once more, in random order. That way, in the second half of the experiment, a Forced block could appear first and therefore mirror the Voluntary block from the first half of the experiment. Control blocks finished after 10 emails were answered, Voluntary blocks after 10 emails and 15 chat messages were answered and Forced blocks were mirroring the Voluntary ones.

2.1.3. Participants

28 participants (14 female), with mean age 22.82, participated in this experiment. They all gave informed consent and received monetary compensation of 10 euros for their participation.

2.1.4. Procedure

The experiment lasted approximately one hour. Before the experiment started, participants completed 3 emails for each of the No-Delay conditions (Control-No Delay, Voluntary-No Delay and Forced-No Delay) as practice. They were instructed that the Voluntary blocks end after they completed 10 emails and answered 15 chat messages and that the chat messages in the Forced blocks should be answered immediately when they appeared. Every move the participant made and every change in the experiment was being recorded in a text file.

2.1.5. Analysis

In order to analyze how much time participants needed to complete an email, we removed all delays from all blocks (time from clicking on a link until the link loaded completely). We also removed the time spent in the chat window, from the moment they clicked on it until the moment they pressed the enter button to send their response. As resumption lag we considered the time from pressing the enter button to send the response to the chat message until clicking on another window to resume the email task.

In this experimental setup, participants are free to click on any window at anytime. This makes the analysis difficult when it comes to classifying a move as high- or low-workload. For example, if an interruption occurred after “Mail Focus”, that could be a high-workload “Mail Focus” (if the participant click on the email window to re-read the product name while searching) or a low-workload “Mail Focus” (if the participant clicked on the email window during a low-workload moment). When analyzing the number of interruptions that occurred on each step of the email task, for simplification reasons we did not include ambiguous steps. Therefore, we only included the “Browser Focus” and “Browser Home” clicks that happened after “Mail Select” (as in Fig. 3) and the “Mail Focus” that happened after a “Link 3”. Due to this intervention, all steps could be clearly classified as high- or low-workload moments.

2.2. Results

There were very few switches during the delay moments in the Delay condition – not nearly as many as in Katidioti and Taatgen (2014). The differences in experimental setup probably made the interrupting task seem less urgent and participants preferred not to self-interrupt during high-workload moments. Therefore, we collapsed over Delay and No Delay for the remaining analyses.

First, we assessed at what moment in the task the interruptions occurred. Participants chose to self-interrupt mainly on low-workload moments (Fig. 4). In the Voluntary condition, 94.72% of the interruptions occurred on low-workload moments - or on average 15.49 (SE = 0.36) self-interruptions per block occurred on low-workload moments. In contrast, only 5.28% of the self-interruptions occurred on high-workload moments. Due to the mirroring, those percentages are 94.1% and 5.92% respectively in the Forced condition.

Furthermore, Fig. 4 shows that interruptions in the Forced blocks were mirroring the self-interruptions on the Voluntary blocks successfully, since the interruptions in the Forced condition occurred in the vast majority of cases at the same moments as in the Voluntary condition. Small differences between Voluntary and Forced blocks exist because of the design of the experiment, which gave participants a lot of freedom. For example, participants could go back to the mail window to see the product code, make a mistake in the search and start over, forget the product price while typing and check it again. It is therefore impossible in this setup to mirror the interruption moments with perfect accuracy.

Next, we analyzed how performance was affected by different types of interruptions, by comparing the time needed to complete an email for each condition (delays and time spent on the chat window removed, as described above). Participants were fastest to complete an email in the Control condition (20.79 s, SE = 0.18), followed by the Forced condition (23.36 s, SE = 1) and the Voluntary condition (24.83 s, SE = 0.72). An ANOVA showed an effect of interruption type ($F(2,54) = 27.03$, $p < 0.001$, $\eta_p^2 = 0.5$) and follow-up t-tests confirmed that all conditions differed significantly from one another (all *p*s < 0.005, Bonferroni-Holm correction for multiple comparisons).

We then investigated whether the time to resume the task after an interruption differed between self-interruptions and external interruptions. The resumption lag in the Voluntary condition lasted on average 1.65 s (SE = 0.063) and in the Forced condition 1.64 s (SE = 0.67). A pairwise *t*-test showed no significant difference ($t(27) < 1$) between the two conditions.

To analyze accuracy, we first looked at mistakes made in the email task. However, participants hardly ever gave the wrong answer to the simulated client. Another performance measure is the number of times participants forgot the product they were supposed to look up and had to turn back to the mail window, face again the 3-s loading time and read the email again. There were on average 0.11 (SE = 0.02) revisits to the mail window per block in the Control condition, 0.14 (SE = 0.026) in the Voluntary condition and 0.12 (SE = 0.026) in the Forced condition. An ANOVA showed no significant difference in the number of revisits between the three conditions ($F(2,54) = 1.45$, $p = 0.24$, $\eta_p^2 = 0.05$).

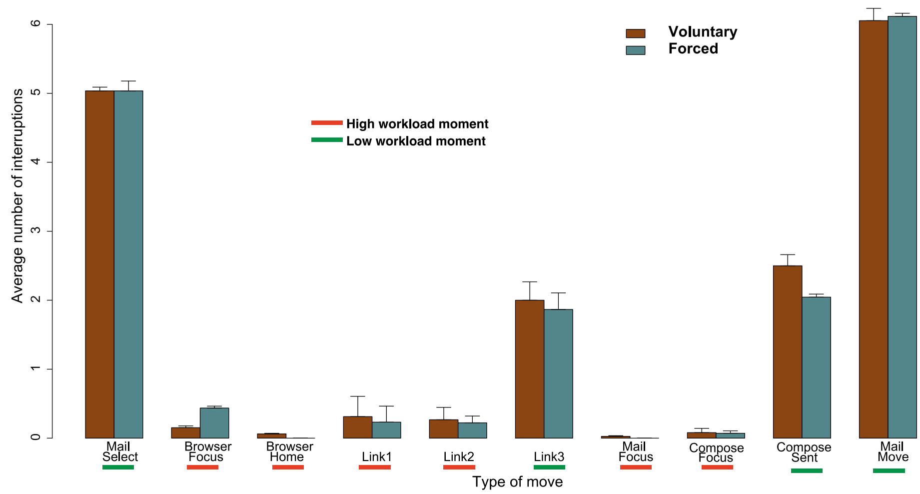


Fig. 4. Average number of interruptions per block for each step of the email task of Experiment 1.

2.3. Discussion

The main goal of Experiment 1 was to contrast self-interruptions (Voluntary condition) with external interruptions (Forced condition). As the external interruptions mirrored the self-interruptions and participants chose to switch mainly on low-workload moments (Fig. 4), the external interruptions also occurred mainly on low-workload moments.

The analysis revealed that the external interruptions were less disruptive than the self-interruptions: participants needed almost 1.5 s less to complete an email in the Forced condition than in the Voluntary condition. However, resumption lag analysis revealed no difference between the Forced and Voluntary conditions. Thus, the cause for the slowing in the Voluntary condition was not that they needed more time to resume the main task after the interruption. Most interruption-effect studies use resumption lag as a measure of disruptiveness (e.g. Monk et al. 2004, 2008), and it is therefore surprising that we did not find a significant difference in resumption lag, even though the time to complete the main task increased. However, if our hypothesis that self-interruption has a decision-time cost is right, the extra cost is incurred *before* the interrupted instead of after it. In order to find evidence for these pre-interruption costs, we conducted the same experiment (without the delay manipulation), but now also measured pupil dilation.

3. Experiment 2

3.1. Pupil dilation

Pupil dilation has been used in cognitive science at least since the 1960s (see Beatty & Lucero-Wagoner, 2000, for a review). The pupil is known to react to a number of cognitive processes. Numerous studies have shown that pupil dilation increases when cognitive workload increases. To give just two examples, when the number of digits to remember increased, pupil dilation also increased in the studies of Beatty (1982), Peavler (1974). The pupil also dilated more as the mathematics participants had to solve became harder in the study of Kahneman, Tursk, Shapiro, and Crider (1969). Beatty and Lucero-Wagoner (2000) mention more similar studies, which lead us to the conclusion that pupil dilation can be used as a continuous measure of cognitive workload.

Pupil dilation has also been used in interruption studies. Iqbal, Adamczyk, Zheng, and Bailey (2005) used pupil dilation to find the high- and low-workload moments in two tasks (route planning and document editing). Results showed that pupil dilation was indeed higher mid-subtask (which is a high-workload moment) than between subtasks (which is a low-workload moment) (also see Iqbal, Zheng, & Bailey, 2004). Iqbal and Bailey (2005) used the same tasks as Iqbal et al. (2005). The points where pupil dilation decreased were categorized as best interruption moments and the points were pupil dilation decreased as worst. Participant performance indeed was worse when they were interrupted on the worst interruption moments. Katidioti et al. (2014) used pupil dilation to compare self-interruption with external interruption. Results showed that there was a greater increase in the pupil dilation before a self-interruption than in the external interruption. The conclusion of this study was that the decision to self-interrupt is reflected in the pupil. In Experiment 2 of the current study we will use pupil dilation to investigate at what moment in time the difference between external and self-interruptions arises.

3.2. Method

3.2.1. Design

The tasks used in this experiment were the same as in Experiment 1.

3.2.2. Conditions

The Delay manipulation was not used in Experiment 2, since it did not prompt participants to switch on high-workload moments. There were three kinds of blocks – Control, Voluntary and Forced (as explained in Experiment 1). Control and Voluntary blocks were presented in random order, but Forced blocks always followed Voluntary blocks, therefore each Forced block was mirroring a different Voluntary block (which was not always the case in Experiment 1, where two Forced blocks could mirror the same Voluntary block). The experiment ended after nine blocks, three of each condition.

3.2.3. Apparatus and setup

Participants were tested individually in a small windowless room. They were seated at a desk with a 20 inch LCD monitor with screen resolution of 1600 × 1200 pixels and screen density of 64 pixels/inch. Participants were asked to use a chin-rest during the experiment, in order to keep their head more stable and have more clear measurements. The eyetracker was an Eyelink 1000 from SR Research, positioned approximately 45 cm from the end of the desk. 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 while participants looked for 0.5 s at a fixation cross.

3.2.4. Participants

21 participants (15 female) with a mean age of 22.71 participated in this experiment. They all had normal or corrected-to-normal vision. They all gave informed consent and received monetary compensation of 10 euros for their participation.

3.2.5. Procedure

The experiment lasted about one hour. Before the experiment started, participants completed 3 emails of each condition as practice. They were instructed that the Voluntary blocks end after they completed 10 emails and answered 15 chat messages. Every move the participant made and every change in the experiment was being recorded in a text file. Furthermore, every 4 ms the *x* and *y* position of the participant's gaze and the dilation of the pupil were being recorded by the eyetracker in a text file.

3.2.6. Analysis

The same analysis as in Experiment 1 was also used for the behavioral part of Experiment 2. For the analysis of the eye-tracker data, eye blinks were removed from the results, starting 100 ms before the blink and finishing 100 ms after the blink, and replaced with a linear interpolation. Then, the data were downsampled to 100 Hz. The percentage change in the pupil dilation was calculated from baseline, which was defined by a very slow lowess filter, i.e. 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 2/3; for more details see Cleveland, 1981).

3.3. Results

3.3.1. Behavioral results

Fig. 5 shows that, as in Experiment 1, participants decided to self-interrupt mostly at low-workload moments. In the Voluntary condition, 96.14% of the interruptions occurred on low-workload moments, and only 3.86% self-interruptions happened on high-workload moments. In the Forced condition, due to the mirroring of interruption moments, these percentages were very similar (95.41% and 4.59% respectively). That difference was also significant ($t(20) = -19.17$, $p < 0.001$, $d = 7.31$). This indicates that the mirroring was successful and that external interruptions (Forced condition) also occurred mostly on low-workload moments, similar to in the Voluntary condition.

As in Experiment 1, results showed that participants were fastest to complete an email in the Control condition (20.59 s, SE = 0.57), followed by the Forced condition (21.78 s, SE = 0.72) and finally they were slowest in the Voluntary condition (22.71 s, SE = 0.78). An ANOVA affirmed that this difference is significant ($F(2,40) = 8.48$, $p < 0.001$, $\eta_p^2 = 0.3$) and follow-up t-tests revealed that all conditions differed significantly from one another ($ps < 0.05$), Control and Forced marginally ($p = 0.07$). The resumption lag was 1.49 s (SE = 0.12) in the Voluntary condition and 1.53 s (SE = 0.14) in the Forced. As in Experiment 1, this difference was not significant according to a *t*-test ($t(20) = -1.45$, $p = 0.16$, $d = 0.08$). Finally, there was no significant difference between the three conditions in the number of times participants had to turn back to the mail window in order to re-read the product information ($F(2,40) < 1$). To summarize: participants were on average approximately one second faster per email in the Forced condition than in the Voluntary condition.

3.3.2. Pupil dilation results

Fig. 6 shows the average percentage change in pupil dilation, time-locked at the moment of interruption, for both the Voluntary and the Forced condition. There was a clear difference before the interruption: the large peak in the dilation signal occurs earlier relative to the interruption moment in the Voluntary condition. In order to quantify this difference, we compared the time from when pupil dilation reaches its maximum for each participant until the moment of the interruption. That time was on average 4.47 s (SE = 0.24) in the Voluntary condition and 3.66 s (SE = 0.22) in the Forced condition. A *t*-test revealed that this difference is significant ($t(20) = -4.17$, $p < 0.001$, $d = 0.76$).

To investigate whether this difference between voluntary and forced interruptions was specific to the moment within the task, we plotted the average percentage change in pupil dilation also time-locked at the moment participants clicked to open a new email (Fig. 7A) and the moment they drop an answered email in the Archive folder (Mail Move) (Fig. 7B). Here, we did not observe any differences between conditions.

Returning to Fig. 6, the two conditions differ again at about 800–1000 ms after the interruption. In this case, pupil dilation in the Forced condition seems to react slower than in the Voluntary condition. We compared the time from the interruption until the

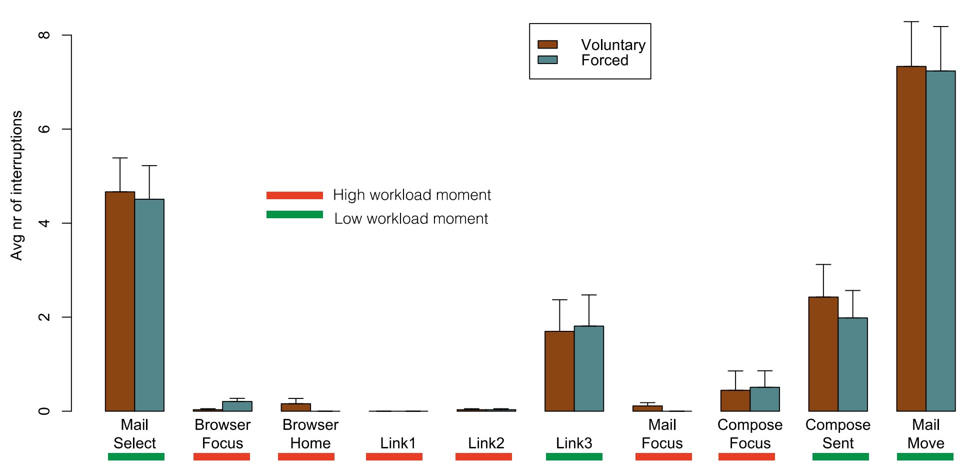


Fig. 5. Average number of interruptions per block for each step of the email task of Experiment 2.

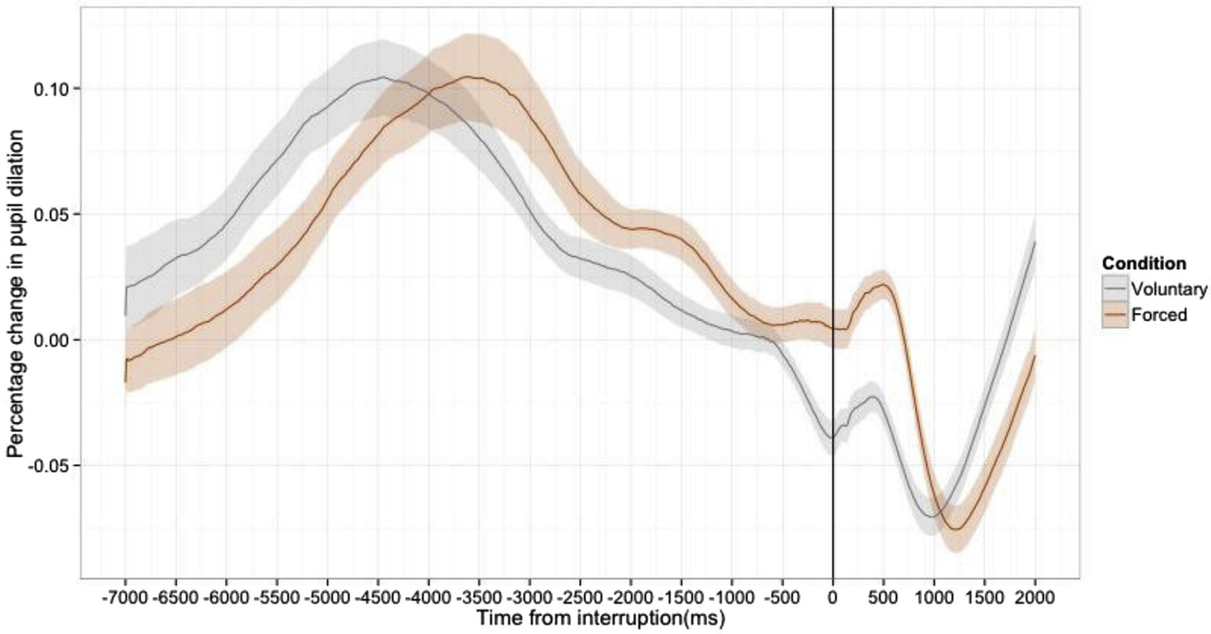
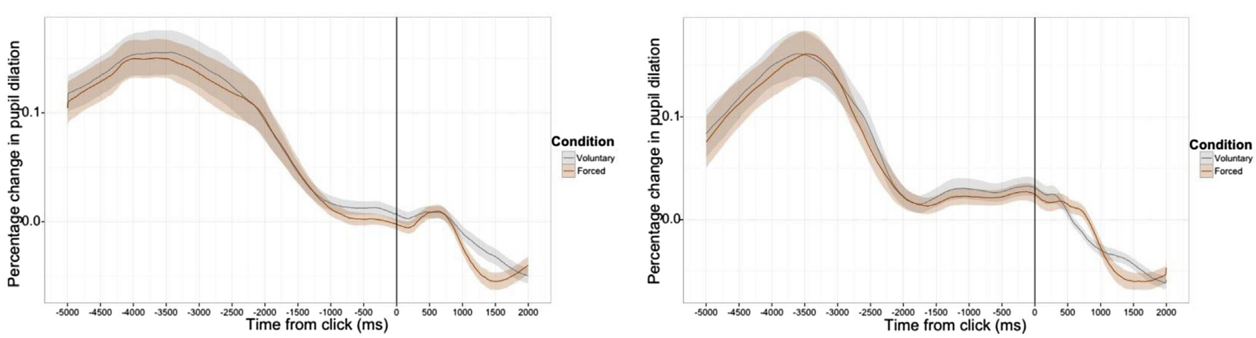


Fig. 6. Pupil dilation for the Voluntary and the Forced condition, time-locked to the moment of interruption. The vertical black line represents the moment of interruption.



A B

Fig. 7. Pupil dilation for the Voluntary and the Forced condition, time-locked on the moment of mail select (A) and mail move (B). The black lines represent the moment that participants clicked the mouse for these actions.

minimum point in each condition for each subject. The Voluntary condition reaches its lowest peak 9.93 (SE = 0.27) sec after the interruption and the Forced 1.13 (SE = 0.85) sec. This difference is not significant ($t(20) = -1.72$, $p = 0.1$, $d = 0.46$).

3.4. Discussion

The behavioral results of Experiment 2 replicated those of Experiment 1. Self-interruptions (Voluntary condition) were more disruptive than external interruptions (Forced condition). Furthermore, as in Experiment 1, the resumption lag was not significantly different between the two conditions. In order to find out what creates the difference between the two conditions – as the resumption lag was similar – we looked at the changes in pupil dilation around the moment of interruption.

As Fig. 7A and B indicate, the two conditions are synchronized throughout the experiment. However, if we time-lock the percentage change in pupil dilation on the moment of interruption (Fig. 6), there is approximately a 1-s phase difference between the two conditions before the interruption. In the Voluntary condition participants actively decide to switch, whereas in the Forced condition they do not make a decision themselves, but the decision is made by the software. We hypothesize that the phase difference in pupil dilation is caused by this decision: between the previous major cognitive event (the peak in the pupil dilation) and the interruption itself, the participants decide to switch in the voluntary condition, which costs about 1 s. This decision to switch is responsible for the Voluntary condition being more disruptive than the Forced condition: each self-interruption adds on average 0.93 s decision-time to completing the email. This is in line with the number of chats answered per email: participants completed on average 14.87 (SE = 1.27) emails and 16.6 (SE = 0.51) chat messages per block.

4. General discussion

In this paper we performed two experiments in order to compare self-interruptions and external interruptions. The goal for this study was to discover which kind of interruption is more disruptive and to complete the interruption timeline for self-

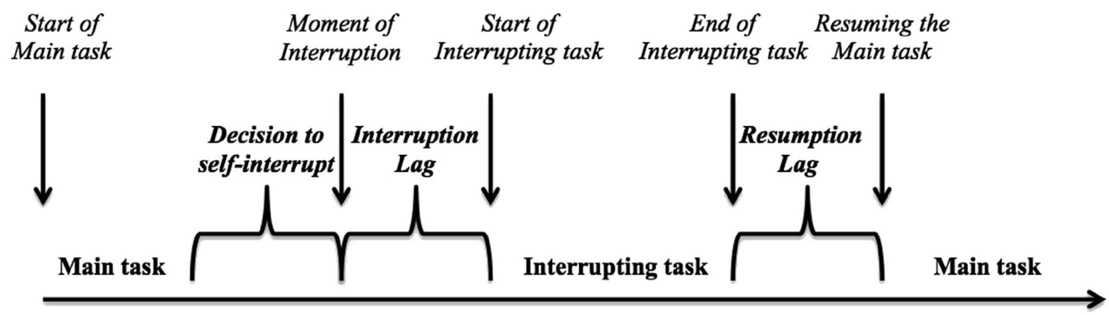


Fig. 8. Timeline of a self-interruption, informed by our pupil dilation results.

interruptions (Fig. 1). Katidioti et al. (2014) discovered that self-interruptions had a different pupil dilation reaction than external interruptions. That lead to the conclusion that there is a decision or preparation period before a self-interruption, which creates time costs that are not present in external interruptions. That difference could account for self-interruptions being more disruptive than external interruptions.

In both experiments of the current paper, participants were faster to complete the main task when they were externally interrupted than when they chose themselves when to self-interrupt. Interestingly enough, there was no difference in the resumption lag (time needed to resume the main task after the interruption) between the two kinds of interruptions – even though that is often taken as a measure of the disruptiveness of interruptions. Therefore, the reason for this difference in performance is not that people found it harder to resume the main task after a self-interruption than after an external interruption.

The pupil dilation measurements in Experiment 2 showed a phase difference before the interruption between self-interruption and external interruption, when pupil dilation was time-locked to the moment of interruption (Fig. 6). From this, we concluded that the preparation for self-interruption adds decision time to the task, which does not exist for external interruptions. These results suggest that the decision to self-interrupt was responsible for the self-interruption being more disruptive, as it creates extra time costs not present in external interruption. Therefore, we can support the idea that the self-interruption timeline needs the addition of "decision time" before the interruption, as it can be seen in Fig. 8. The validity of this timeline could be verified with more experiments using different methods (such as EEG) and with cognitive modeling (see Borst et al., 2015; Taatgen et al. 2015 for examples of cognitive models of interruptions and distraction).

One important detail to note is that the external interruptions in both experiments occurred mainly at low-workload moments (Figs. 4 and 5). High-workload moment interruptions are known to be more disruptive than low-workload ones (e.g. Iqbal & Bailey, 2005, 2006; Katidioti & Taatgen, 2014; Monk et al. 2004). Furthermore, people are known to self-interrupt on low-workload moments (e.g. Katidioti & Taatgen, 2014; Salvucci & Bogunovich, 2010). Therefore it would not be fair to say that all external interruptions are less disruptive than self-interruptions. As Katidioti et al. (2014) showed, the difference between the two kinds of interruptions existed only when the external interruptions occurred on low-workload moments, since in their first experiment the external interruptions occurred on high-workload moments and in that situation there was no significant difference between the two kinds of interruptions as participants self-interrupted at low-workload moments. However, we still do not know if there is a difference between high-workload external interruptions and high-workload self-interruptions, since people tend to self-interrupt mainly on low-workload moments.

4.1. Practical applications

The results of this study suggest that being interrupted by an external source on low-workload moments is less disruptive than deciding when to self-interrupt. There are many practical applications that can arise from this result for a number of different working environments. Environments where the user needs to deal with different tasks, such as client service with email and live chat options, air traffic control or piloting a plane, could be made more efficient if the decision to self-interrupt is replaced by a carefully timed external interruption. Instead of giving the users the choice to self-interrupt, the low-workload moments of their tasks can be identified (either with task-analysis or psychometric methods) and self-interruptions can be turned into external interruptions on those low-workload moments.

The results can also be generalized to a normal working environment, where tasks are not as restricted as the ones mentioned above. Office workers or students are known to self-interrupt constantly (e.g. Gonzalez & Mark, 2004; Rosen et al. 2013) and the results of our study could benefit them. For example an app that users can program to externally interrupt them on low-workload moments can replace the endless checking of emails or social media. Our next step is to create an interruption management system that interrupts users when pupil dilation indicates that it is a low-workload moment and compare these interruptions with random interruptions. Based on the current results, we should also compare interruptions managed by the system to self-interruptions – which might be the slowest kind of interruption of all.

Acknowledgments

This research was funded by ERC-StG grant 283597 awarded to Niels Taatgen. We thank Dario Salvucci for making the code for the experiment available to us.

References

Adler, R. F., & Benbunan-Fich, R. (2013). Self-interruptions in discretionary multi-tasking. *Computers in Human Behavior, 29*(4), 1441–1449.
Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211.
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.
Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*, 276–292.
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.

Blanchard, A. L., & Henle, C. A. (2008). Correlates of different forms of cyberloafing: The role of norms and external locus of control. *Computers in Human Behavior, 24*(3), 1067–1084.

Borst, J. P., Taatgen, N. A., & van Rijn, H. (2015). What makes interruptions disruptive? A process-model account of the effects of the problem state bottleneck on task interruption and resumption. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '15)*.

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*(2), 95–107.

Cades, D. M., Boehm-Davis, D. A., Trafton, G. J., & Monk, C. A. (2007). Does the difficulty of an interruption affect our ability to resume?. In *Proceedings of the human factors and ergonomics society annual meeting 2007* (pp. 234–238).

Cleveland, W. S. (1981). Lowess: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician, 35*(1), 54.

Czerwinski, M., Cutrell, E., & Horvitz, E. (2000). *Instant messaging: Effects of relevance and timing. BCS HCI '00* (Vol. 2, pp. 71–76).

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '04)* (pp. 175–182).

Dindar, M., & Akbulut, Y. (2016). Effects of multitasking on retention and topic interest. *Learning and Instruction, 41*, 94–105.

Gonzalez, V., & Mark, G. (2004). "Constant, constant, multi-tasking Craziness": Managing multiple working spheres. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '04)* (pp. 113–120).

Gould, S. J. J., Brumby, D. P., & Cox, A. L. (2013). What does it mean for an interruption to be relevant? An investigation of relevance as a memory effect. In *Proceedings of the human factors and ergonomics society annual meeting 2013* (pp. 149–153).

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.

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 *Proceedings of the SIGCHI conference on human factors in computing systems (CHI'05)* (pp. 311–320).

Iqbal, S. T., & Bailey, B. P. (2005). Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '05)* (pp. 1489–1492).

Iqbal, S. T., & Bailey, B. P. (2006). Leveraging characteristics of task structure to predict the cost of interruption. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '06)* (pp. 741–750).

Iqbal, S. T., & Horvitz, E. (2007). Disruption and recovery of computing tasks. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '07)* (pp. 677–686).

Iqbal, S. T., Zheng, X. S., & Bailey, B. P. (2004). Task- evoked pupillary response to mental workload in human-computer interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '04)* (pp. 1477–1480).

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*(1), 164–167.

Katidioti, I., Borst, J. P., & Taatgen, N. A. (2014). What happens when we switch tasks: Pupil dilation in multitasking. *Journal of Experimental Psychology: Applied, 20*(6), 380–396.

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*(4), 728–736.

Lim, V. K. G., & Teo, T. S. H. (2005). Prevalence, perceived seriousness, justification and regulation of cyberloafing in Singapore: An exploratory study. *Information & Management, 42*(8), 1081–1093.

Mark, G., Gonzalez, V., & Harris, J. (2005). No task left behind? Examining the nature of fragmented work. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '05)* (pp. 321–330).

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

Monk, C. A., Boehm-Davis, D. A., & Trafton, J. G. (2004). Recovering from interruptions: Implications for driver distraction research. *Human Factors, 46*(4), 650–663.

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*(4), 299–313.

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.

Peavler, W. S. (1974). Pupil size, information overload and performance differences. *Psychophysiology, 11*(5), 559–566.

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*(3), 948–958.

Salvucci, D. D., & Bogunovich, P. (2010). Multitasking and monotasking: The effects of mental workload on deferred task interruptions. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '10)* (pp. 85–88).

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

Salvucci, D. D., Taatgen, N. A., & Borst, J. P. (2009). Toward a unified theory of the multitasking continuum: From concurrent performance to task switching, interruption, and resumption. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '09)* (pp. 1819–1828).

Taatgen, N. A., Katidioti, I., Borst, J. P., & Van Vugt, M. K. (2015). A model of distraction using new architectural mechanisms to manage multiple goals. In *Proceedings ICCM 2015. Groningen, The Netherlands*.

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.

Wagner, D. T., Barnes, C. M., Lim, V. K. G., & Ferris, D. L. (2012). Lost sleep and cyberloafing: Evidence from the laboratory and a daylight saving time quasi experiment. *Journal of Applied Psychology, 97*(5), 1068–1076.